



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB

(b)(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	12

(A)(1) Reviewer Comments:

Smarter Balance Consortium (SBAC) application presents an ambitious and comprehensive assessment reform agenda that covers the 8 selection criteria outlined in the Notice Inviting Applications (NIA). It includes a clear plan with operational components aligned to its overarching Theory of Action (TOA). Areas requiring attention and further refinements are identified in the applicable sections throughout this review. The plan's best strengths are embedded in the development and implementation of the summative and interim assessment components. The plan could benefit from further clarity and more explicit articulation of plan elements (goals, deliverables, time frames, etc.) which are aligned to the TOA in the areas building teacher capacity and use of assessment data as a means to improve student outcomes.

The application narrative and additional descriptive information included in Appendices A 1 through 4 provide detailed responses to subsections i through v under the area of Structure and Operations. SBAC is comprised of 31 member states- 17 of which are Governing States and 14 have signed on as Advisory Members. The application does not address:

- broad based representation in the makeup of the SBAC governance and advisory structures
- outreach and inclusion of local level district and school representatives in the advisory and governance structures
- decision making process for reaching maximum consensus among participants
- how the work carried out with members involved in other consortia (if funded) will intersect
- the need for written procedures and meeting protocols, such as ByLaws, rules and procedural guidelines

The State of Washington will be legally responsible for the use of grant funds and for ensuring that the Consortium carries out the project in accordance with federal requirements. The Consortium's plan for managing funds received under this grant category will be governed by:

- the laws and rules of the State of Washington, as the Lead Procurement State
- guidelines for grant management associated with the American Recovery and Reinvestment Act of 2009 (ARRA)

Additional fiscal management procedures include a quarterly reporting system already established, WA's accounting practices and adherence to the state's comprehensive contracting rules.

MOUs submitted by each of the SBAC states adhere to a standard format. Except for the allowable variance in membership roles (i.e. Lead Procurement, Governing and Advisory), all state MOUs conform to the same terms and conditions. There is not a completed MOU included in the application for SD. Fifteen of the MOUs submitted by the states identify a complex range of existing barriers to the effective implementation of the proposed assessment system and provide timelines for addressing them.

Many of the barriers identified represent common concerns across states (e.g. adoption and potential conflicts related to the Common Core Standards, adequate budget appropriations and assurances of financial sustainability, local district capacity and readiness for the new assessments, adequate technology infrastructure, higher education consensus on high school assessments for use as non-remedial course placement, etc.)

Information submitted by the applicant satisfactorily addresses procurement and Consortium members' commitment to the described procurement process. The Lead Procurement State is WA. In this role WA has the authority and responsibilities assigned to the Governing States and is entitled to the state's negotiated indirect rate for federal grants. The Lead Procurement State also has responsibility for:

- overseeing the management of funds, in collaboration with the Steering Committee and Executive Committee
- overseeing all procurement on behalf of the consortium

The procurement process will be guided by the laws and rules of the State of WA. Evidence of individual state commitment to the procurement process is documented by the signatures provided in each of MOUs.

Recommendations

SBAC should:

Consortium Governance

- Establish guidelines for size, balance and stakeholder representation in the Policy Advisory Committee and Working Groups. Membership in these 2 groups should extend beyond the self-nomination process described in Appendix A1-3.
- Provide explicit guidance should be available for contractor's use in the selection processes for participant groups related to test development, scoring and achievement standard setting. Outreach efforts should be made to recruit needed talent and expertise and to ensure diversity in all groups and committees that are representative of Consortium membership.
- Include local district and school level personnel (the ultimate consumers) in the SBAC governance level (i.e. a membership slot on the Consortium's Steering Committee and Working Groups)
- Create operational structures for day to day management of individual components and overall project

Project Management

- Develop a detailed Project Management Plan which delineates more specific tasks under each of the major plan components along with associated potential risks

Decision Making

- Provide clarity regarding how consensus will be used as a decision making protocol and more detailed specification of areas assigned to the various governance structured entities (e.g. oversight, broad picture, advisory)
- Reconsider decision rule of simple majority vote following failure in order to achieve maximum consensus among Consortium members.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	3
<p>(A)(2) Reviewer Comments:</p> <p>SBAC's Theory of Action (TOA) is grounded in a set of fundamental beliefs about the intersection of assessments, instruction and student achievement. Central to its approach are innovative notions regarding how high quality assessments if used effectively can leverage stronger student learning. SBAC's theory along with the 7 undergirding principles outlined in the application are tightly aligned. There are a number of core themes embedded in the 7 principles that are derived from knowledge and best practices in the areas of assessment, teaching and learning. These core themes include:</p> <ul style="list-style-type: none"> • Integral involvement of teachers • Integration of system components including standards, curriculum, assessment, instruction and professional development • Evidence based student performance on challenging and rigorous tasks • Continuous improvement • Reporting based on multiple measures and reciprocal accountability • Adherence to professional standards • Transparent and inclusive leadership and governance <p>The key attributes and concepts of the SBAC TOA are closely aligned with the RtT Assessment selection criteria. The SBAC proposal sets forth far reaching goals. The TOA acknowledges in its proposal that in order to reach its stated goals there are other elements which fall outside of the SBAC direct scope of work that will require reform and coordination. Examples cited included accountability systems and pre-service/in-service professional development but there are others that warrant consideration if the new assessments are to result in positive impact for all students, e.g. opportunity to learn standards, engagement of collective bargain units, equity based resource distribution, and integration of social/emotional supports with the core academic program. Attention to these policy areas are deemed important to the ultimate achievement of what is called for in Selection Criteria A c and d-- a coherent educational system capable of serving the needs of all students. The SBAC proposal is strong in many of its core components. Two plan areas, however, are substantially underdeveloped: 1) integration of standards curriculum, instruction and professional development and 2) use of assessments for continuous improvement.</p> <p>Recommendations</p> <p>SBAC should:</p> <ul style="list-style-type: none"> • Give attention to the development of ongoing feedback mechanisms to gauge the level of broad-based support that is being generated in support of the assessment system • Balance the merits associated with the SBAC proposal with ongoing research on potential student and institutional consequences. The Consortium should identify early on where there might be challenges or barriers to the new assessment system and develop strategies to address them as implementation moves forward. 		

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	45

(A)(3) Reviewer Comments:

Overall the SBAC application lays out a comprehensive an assessment design process which is aligned to the Consortium’s overarching Theory of Action and responsive to the selection criteria in this section. The SBAC assessment system will be comprised of 3 major components in Mathematics and English Language Arts:

- a computer adaptive summative assessment given at the end of grades 3 -8 and 11
- interim/benchmark assessments built around learning progressions and administered at various points during the year
- research supported formative tools, resources, materials and training used at the classroom level

The SBAC is committed to investigating the feasibility of developing a “through course” option summative assessment based on content clusters that could be administered throughout the year. For states that may wish to measure student growth prior to grade 11, a summative adaptive test will be available in grades 9 and 10. Information available from each component will be used in a manner consistent with its design and purposes.

The SBAC assessment system will include computer adaptive selected-response items, technology-enhanced constructed-response items, and extended constructed-response items, as well as, standardized performance events at each grade level (3-8 and high school). The use of performance events will be central to system’s ability to measure student knowledge and skills against the full range of the college and career ready standards. The performance components will reflect more ambitious events that can measure aspects of student performance reflected in the Common Core Standards but have traditionally been difficult to measure on standardized assessments, including skills such as the use of relevant evidence and technology, thoughtful critique, and adaptive reasoning. The SBAC system will be built using a variety of innovative technology applications based on a computer adaptive model. This approach will provide a unique opportunity to create a large-scale assessment system that provides maximally accurate achievement results for each student. The adaptive assessments will be highly sensitive to the unique status of the learner and will sample content above and below grade level, as needed, to ensure the accurate assessment of individual student’s progress toward meeting the expectation for college and career readiness. End of year summative scores will provide for a common measure on which a stable measure of achievement and growth can be calculated. Computer adaptive testing has also been proven to be effective in measuring student growth over time.

Determining the extent to which the student achievement results from the SBAC summative assessment component will serve as a valid measure for whether students are on track or ready for college will involve a complex set of activities. Major tasks will include: development of test blueprints closely aligned to the Common Core Standards in depth and breadth, independent alignment studies to ensure all items and events fully assess the intended content, technical analyses based on selected samplings along with external validity studies to measure whether students who achieve mastery at a grade level do indeed achieve that grade level at the next grade. By far the most complex and significant step will be the setting/finalizing of achievement standards following full field testing (spring 2014-15). Prior to or during standard setting, a set of achievement descriptors for each grade level 3-8 and 11 will be developed. SBAC application does not address in substantial detail how the performance descriptors will be developed and sequenced nor who will be involved. It also does not indicate how test specifications and performance descriptors will provide a clear picture of performance expectations within each content strand. The application does speak to the importance of including in design features assurances that will protect against an unreasonable hierarchy among the knowledge and skills areas, e.g. requirement for higher order thinking only tied to high level content. SBAC will coordinate standard setting with other consortia funded under the RttT Assessment

competition, however, it is not clear how this will happen. If grant funds are awarded, much work remains to iron out the details on this essential feature taking into consideration validated psychometric procedures, timing and high stakes consequences. The level of effort and resources that may be required for quality standard setting and validation procedures are beyond what currently appears in the application and budget.

The framework for ensuring meaningful access for all students to the SBAC assessment system will be guided by the Access by Design Model included in Appendix A 3-2. Utilizing this framework, accessibility features will be addressed and incorporated purposefully from the beginning throughout the development process. According to the principles outlined, few specific accommodations would be needed according to the Access by Design Model. Without more operational details regarding how the Model's concepts would be implemented, it is difficult to assess whether this selection criteria has been adequately addressed. The application does not describe exemplary programs that are currently using the proposed Model's concepts to substantiate the merits of its proposed approach. The budget narrative expresses the need for funding of an Accommodations Study at an unspecified amount. Other strategies, activities or contract embedded tasks required to carry out the intent of this RtT Assessment area should be provided. In A 4 Assessment Development the issue of accommodations is also addressed, however, even the 2 sections together, do not advance an adequate operational plan nor substantiate the need for assessment supports in this area.

One of the essential features of the Smarter Balanced Assessment System is its timely release of assessment results. Student level data will be accessible throughout the year. The computer-administered Summative and Interim/Benchmark assessments will be able to generate immediate results for the selected response and technology enhanced items. In the application, the summative component is scheduled to occur during the 12th week before end of the instructional calendar. Teacher scoring will be required for select dimensions of the performance events. Once operational, a 2 week window for turnaround will likely be required. In the area of formative assessments, the expectation is that there would be a means to collect evidence immediately during the course of instruction. SBAC has committed to support the development of formative assessment tools and related professional development. This task has been delegated to the states for implementation. It is not clear what role the Consortium will play in guiding, monitoring or coordinating the states activities. Work in this area is also included under Selection Criterion A 6 Professional Capacity and Outreach.

The summative assessments in mathematics and English language arts will provide information about student progress toward college- and career-readiness in grades 3-8 and 11. SBAC will build a vertical scale and define vertically and horizontally articulated achievement standards in both subject areas. A description of the procedure that will be used to develop the vertical scale can be found in section (A)(5). Scale scores used for achievement or growth metric will be part of the summative component while scale scores and cluster-level data will be part of the Interim/Benchmark component. Only cluster level achievement and growth data will comprise the Formative component.

The manner in which the SBAC is being conceptualized and designed will allow use of the data produced for a variety of purposes and which take into account the multiple possibilities defined in this selection criteria, e.g. effectiveness, accountability and instructional improvement. As noted in the SBAC application, the full range of options for using data produced by the summative component especially in the area of accountability, will need to await further congressional action (i.e. reauthorization of ESEA) and USED guidance.

The 3 major SBAC assessment components serve different primary purposes within the system, thus, the frequency and time of administration vary by component:

- The adaptive summative component will be administered within the last 12 weeks of the instructional calendar. Students would be allowed two testing opportunities.
- The I/B assessments allow for flexible administration at the discretion of the State, district, school, or teacher.
- Frequency and timing of teacher administered formative processes would occur throughout the instructional year and as part of daily instruction.

The Summary Tables for (A)(3) outline the numbers and types of items and the distribution of items proposed for each assessment component. SBAC is committed to ensuring development of varied items that reflect important knowledge and that will elicit complex student demonstrations of knowledge or skill application—moving beyond a minimum basic skills test. Additionally, effective technology applications will facilitate expansion not only the nature of the content that can be presented but also the knowledge, skills, and processes that can be assessed.

SBAC is committed to the use of a computer adaptive summative assessment with performance events delivered via computer. The Consortium will provide a paper-and-pencil option for a limited time (three years) to support States where required student access to computers for the test administration window remain a barrier. The paper option will be offered as an accommodated form and will include selected-response items, constructed-response items, and performance events.

The assessment system will make use of automated computer-based scoring of selected-response and technology-enhanced items. These methodologies are already well developed and up and running within several SABC member states. AI scoring will be employed for the constructed-response items and performance events. The AI scoring can be used immediately upon the completion of the administration. For those select performance events that are found to tap student performances not effectively scored through AI technology, teachers will be involved as scorers and employ similar read-behind methodologies to ensure scorer accuracy. A two-week turnaround of teacher scores is projected during the operational administration. Teachers will play a greater role in scoring of Interim/Benchmark and formative assessments.

The SBAC system will provide several types of reports as described in Summary Table A(3). Providing such a rich array of accessible data will go a long way in creating credibility, public access and usability of the new assessments. The application does not raise any significant warnings about the potential misuse of assessments (e.g. as a single criterion for high stakes decisions, remedial placement, or grade retention).

Recommendations

SBAC should:

- Provide operational details for how principles of the Access by Design Model will be carried out including attention to accommodations
- Provide rationale and details for the performance descriptors process including collaboration with other similar entities, timeframe and who will be involved
- Clarify how collaboration will extend to standard setting--level of effort, resource needs and validation
- Provide assurance that item development does not imply an unreasonable hierarchy among knowledge and skill areas
- Develop guidelines on expectations for appropriate test use consistent with research based standards

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	25

(A)(4) Reviewer Comments:

It will be the charge of the Assessment Design Working Group to ensure that steps in the process of development for each assessment component are transparent and lead to measures that are valid. Key features of SBAC approach to item development will involve: (1) clear specification of progressions of learning expectations, coordinated across summative, I/B, and formative assessment tools at increasingly

more finely grained levels of detail; (2) test blueprints with clear rules for sampling the full domain of those learning expectations; (3) learning-based item and performance event design templates, including scoring rubrics that establish clear targets for teaching and learning; support coordinated sets of summative, interim, and formative tools; and provide a substantive basis for test comparability from year to year to complement psychometric indices; and (4) the potential for innovation and efficiency through technology-enhanced items and automated scoring. The application elaborates on the strategies and activities associated with each of the development phases. There is a high level of complexity associated with the work and tasks required to build an effective system capable of serving well all students. While the basic approaches and technical design features have precedence in prior experience and psychometric best practices, the new assessment system will require substantially more effort to build broad based understanding, acceptance and use across a diverse group of stakeholders. There does not appear to be sufficient breadth in participation given the multi state focus of this effort or adequacy of personnel assigned to the work. Of particular concern is the linking work between standards and assessments. This of necessity will need to involve knowledgeable content specialists, as well as, teachers of high end and low end students. Caution should be exercised in the hypothesizing of learning progressions involving content knowledge and skills to ensure various learning styles are fully accommodated.

The SBAC application provides minimal detail with respect to how it will fully incorporate SWD and ELL students except to indicate that its approach will be consistent with research findings and best practices pertaining to these groups. The application references a hopeful collaboration with an Enhanced Assessment Grant focused on accommodation policies and which is supported by 23 SBAC member states. The goals and proposed activities outlined in the EAG grant (Appendix A 4-3) are closely aligned to the principles of the Balanced Assessment system. Information pertaining to the EAG grant status is not provided nor is there a contingency plan included in the case grant funding is not awarded. Should the EAG proposal receive funding it would serve as a strong resource in building SBAC capacity and work in this area. This section of the application on Assessment System Development, Access by Design Model was promoted as the guiding framework for promoting and ensuring access of all students to SBAC's system. It is not clear how this framework ties to the A4 b selection criterion.

SBAC is committed to implement a reporting plan that affords an appropriate balance between standard and customized reports but also serves the various purposes of the reporting system. These specific purposes include :

- managing an integrated assessment and accountability system for use by states
- providing "early warning" information to monitor curriculum and support instruction
- making timely and informed improvements in curriculum
- supporting professional development
- providing meaningful achievement information to all stakeholders, including IHEs and providing comparable information about student achievement-for accountability at the local, State, and Federal levels

The reporting plan will include several key and innovative design features such as: a common electronic platform, a set of data analysis and report-generating tools that will allow for the development of customized reports that display data through a variety of tables and graphic formats; an array of trustworthy information about student achievement and growth, reporting results that are technically sound and consistent with the technical parameters and limitations of the data; multiple reporting interfaces and reports differentiated by audience. The reporting system will provide both standard paper-based reports and more technologically advanced web-based data analysis tools. Information on allowable accommodations and instructional supports will be accessible to teachers and test administrators as they prepare for instruction or assessment. District and school administrators and teachers will be trained in the interpretations. Overall, the approach presented in the application is descriptive as a concept but lacks specificity in terms of operational details, prioritization of work tasks, needed personnel or consultants.

Recommendations

SBAC should:

- Reassess the numbers, types of personnel and strategies for involvement in each development step.
- Establish tighter alignment between the budget and the program plan in such areas as personnel/consultants, tasks, activities and assumptions for resource allocation.
- Review budgetary allocation at the level of \$50,000 to support the wide ranging reports that are expected to be developed for the summative and I/B components. There is also merit in coordinating the reporting work plan with the outreach efforts in Section A 6.
- Increase participation in test development steps to be more broadly representative of SBAC member states
- Provide substantiation to support referenced activities for the linking work to standards tied to the learning progressions

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	20

(A)(5) Reviewer Comments:

The SBAC research and evaluation plan will be overseen and guided by the Research and Evaluation Working Group in collaboration with the Technical Advisory Committee. The plan is grounded in the principles adopted by the Joint Committee on Standards for Educational Evaluation. It will follow a “reasoning for evidence” based approach in collecting and evaluating documentation to support claims of validity, ensuring that different types of evidence are collected on an ongoing basis during all phase of system implementation. The key components of the evaluation plan will be the summative assessments and the optional interim/benchmark assessments. The primary purposes will be to ensure the assessment system has integrity, high technical quality and that the measurement components are fair and that the results are used for valid purposes. The Consortium also has an ambitious research agenda that will allow the systematic examination of a series of important empirical questions and provide contributions in important areas to support continuous improvement. As stated in the proposal, qualitative data will be collected about items and performance events for teachers and students. The proposal does not address whether the intended effects on institutions is being achieved.

The research and evaluation plan while strong in what it has set forth does not appear to be fully aligned with the SBAC Theory of Action (TOA). The Consortium’s overall TOA is firmly driven by a set of balanced assessment concepts which include technology supported tools, innovative assessments and classroom based mechanisms that work coherently to support teaching and learning. The plan does not address the area of Formative: Tools, Processes and Practices. There is no mention of how effectiveness of this system component will be evaluated or a rationale for excluding it e.g. complexity, cost factors, specific implementation barriers. Failure to address the classroom feature as part of the research and evaluation plan sends a wrong signal about importance. Further, in order to judge the effectiveness of the overall system, data is needed for all components.

Recommendations

SBAC should:

- Establish performance outcomes exist in all plan segments along with appropriate measurement strategies as part of a comprehensive evaluation plan
- Revise budgets and project management plans to include design and implementation of a more comprehensive evaluation strategy covering all components with resulting data used through out the grant period for continuous improvement

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	3

(A)(6) Reviewer Comments:

The application describes a number of major initiatives and activities that will be undertaken at both the state and local levels in this area. The essential features of the SBAC plan include:

- Teacher involvement in development, review and scoring of assessment items and performance events included in 2 components: Summative and Interim/Benchmark
- Providing models for local development of Formative Tools/Processes.
- Development and dissemination of standards and assessment related resources
- Establishment of a system portal including and educators dashboard and other on line tools for use by teachers, students and parents

The plan is structured around collaboration with professional development networks in each participating state that will in turn help schools and districts develop resources and training to support teachers, teacher leaders and administrators at the local level. The plan has interesting ideas, but does not provide sufficient detail regarding implementation. It talks about goals but it is not clear how the outlined approach is aligned to the three stated goals. Overall the plan does not articulate measurable outcomes or the need for several of the activities proposed has not been substantiated. For example, the plan makes several assumptions about work that will be undertaken related to the Common Core Standards, such as collection “unpacking” efforts across the 30 states combined with web based discussions, development of curriculum frameworks, support for curriculum alignment to the learning progressions. The plan does not acknowledge or incorporate the many initiatives and efforts that are currently addressing these same elements, e.g. Race to the Top plan components, existing web based resources, foundation funded projects directed to rolling out the Common Core Standards and developing teacher capacity, federally supported activities being generated out of regional Labs and national centers, etc. While the SBAC staff should be knowledgeable of major national and state related standards initiatives, duplicate efforts are not needed. Rather SBAC should seek partnerships with other organizations particularly in the area of Common Core Standards implementation while concurrently proceeding to focus on essential topics which it is uniquely capable of addressing, e.g. assessment literacy, formative assessment guidance for instruction, common frameworks for assessment, etc.

The applicant’s outreach plan seems overly ambitious given the wide range and complex activities included. It does not take into consideration that assignment of needed resources or staff effort that would be needed to accomplish successfully and at a high level of quality all of the included elements. The plan presented acknowledges the importance of clear and timely communication in order for the assessment system to be effective and contains interesting ideas targeted at a broad base of stakeholders that include tools, protocols and tools for public consumption. Implementation will rely largely on existing State, regional and federal communication mechanisms along with the private sector to disseminate information. The plan appears to be generally strong at a conceptual level. Further development will occur in collaboration with the Steering Committee and Working Groups focused on this area. While it is expected that additional details will evolve once the project is funded, there are important considerations to be addressed at the outset. For example, the plan talks about outreach efforts to diverse stakeholders, however, it does not address differentiated strategies nor target populations within stakeholder groups (e.g. non-English speaking parents will require materials in

multiple languages, outreach activities to grassroots and remote audiences where there are technology challenges, distinctions between urban and rural populations.) Outreach strategies should extend beyond traditional communication to include networks to national and regional organizations representing targeted constituent groups, as well as, faith based and community organizations. The plan as currently conceptualized views communication as primarily a one way strategy “getting info out and disseminated.” There do not appear to be mechanisms for “getting information back.”, e.g. how effectiveness of implementation efforts will be measured and feedback used for continuous improvement of communication activities. The plan also does not address how building of continuing support for the assessment effort will be generated—a specific expectation of this plan component. In addition to evaluating the effectiveness of the communications effort, the plan could benefit from deeper consideration of how overtime it will be determined that various stakeholders understand and are embracing the new assessment system. Techniques that might be utilized include focus groups and surveys. The plan’s serious inclusion of such a focus in the beginning will help in identifying where among stakeholders there are potential barriers and challenges that need to be addressed. Overall this plan element should be ratcheted up and the budget for this component reexamined to incorporate a more aggressive set of communication and outreach activities.

The plan identifies a combined Professional Capacity and Outreach Work Group to oversee the work under this component. Aside from the fact that the application format includes these 2 criteria to be addressed under the A 6 plan component, there is no rationale provided for a Work Group covering both areas. Consideration should be given to the separation of this group into 2 foci—one for the professional capacity in implementing and use of the system and another to concentrate on communications and outreach targeted to a broad group of stakeholders. Use in a more general sense would need to be included under the second communications purview (e.g. interpretation of assessment results) while attention to deeper applications to instructional practice by teachers and administrators would be the focus under professional capacity building. It is also recommended that formation of the Work Group focused on outreach should extend beyond the self nomination procedures described in Appendix A1-3 to ensure diverse membership representing the various stakeholders along with skills drawn from non-educators (e.g. communications, public relations and engagement experts.)

Recommendations

SBAC should:

- Develop strategies for Outreach to generate feedback to plan development, identify potential barriers and gauge public support
- Develop operational/project management plan for Professional Capacity to include:
 - Clear statement of goals and objectives and alignment of implementation strategies and activities to each goal statement
 - Specific deliverables and time frames
 - Per cent of project management personnel who will lead the professional capacity activities and required qualifications
 - Data collection processes to determine status of state capacity along with availability of existing resources
 - Assignment of personnel to specific task responsibilities
 - Performance benchmarks and measures of effectiveness
 - Continuous improvement strategies (e.g. lessons learned, monitoring for effectiveness data collection)
 - Resource needs for each plan component
- Establish guidelines for teacher involvement in assessment development and scoring that will ensure broad participation, e.g. ethnic/racial and gender diversity, urban and rural balance, elementary, high school and collegiate levels, public and public charter schools, teacher expertise in AP and remedial instruction, and skills/experience in SPED and ELL.
- Align budget with plan component priorities

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	7
(A)(7) Reviewer Comments:		
<p>Efficient and innovative use of technology is the cornerstone of the Consortium’s assessment model. SBAC asserts that the expanded use of the proposed technology applications will increase engagement in assessment, lead to more timely results, lead to expanded use of assessment data by a broader group of stakeholders and offer more efficiencies and enhancements for professional development. Technology applications will be used in various ways including: assessment development and implementation, data management and public accessibility and delivery of professional development.</p> <p>A key feature of the assessment system plan will be the system portal. The portal will serve as a central location for stakeholders such as administrators, educators, parents/guardians and students to access assessment results as well as to classroom materials. The portal will serve as the hub for the educator resource and educator dashboard. Teachers will be able to move seamlessly from student profile to linked materials that support targeted instructional change. The assessment system software will be developed using a combination of existing and newly developed open-source software and proprietary software. SBAC will create an innovative and flexible on line system and the flexibility to implement State specific approved variations and drawing on best practices in on line assessments.</p> <p>Cost effective reuse of the SBAC technology platforms will be addressed through a process that prioritizes the use of open-source and interoperable standards, alternative scoring models, performance event sharing, hosting options, infrastructure guidelines and administrative efficiency. Included as part of the system development will be code sharing and feedback from an unrestricted user base. At the completion of the development proces, steps will be taken to create a public license defining the resulting product as a free and open-source software application.</p> <p>As part of the initial assessment development process, SBAC will collect “Lessons Learned” from other states. Under the auspices of the Technology Task Force and working with states already engaged in on line assessments, a list of risks will be identified and published along with risk mitigation strategies and solutions to serve as guidelines for technology development. The response to this criterion is addressed at a highly conceptual and procedural level.</p> <p>Recommendations</p> <p>SBAC should:</p> <ul style="list-style-type: none"> • Identify preliminary technology implementation challenges illustrative of the kind of risks it is likely to face in this area and suggest up front program design/solutions implications, as well as, potential needed fiscal resources • Indicate in budget module narrative/budget table where resources have been specifically factored in to address this criterion 		

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	20
(A)(8) Reviewer Comments:		
<p>The SBAC’s approach to project management includes: 1) extension of the contract timeframe and utilizing the expertise of its current Grant Project Manager (WestEd) to service and manage the Consortium through</p>		

the start-up and project transition phase and 2) implementation of a comprehensive and competitive procurement process for the purpose of determining a permanent Partner Project Management Partner. The procurement process is being led by the WA State Department of Education as a consortium governing state and designated lead responsible for contract procurements. Since the final partner will not be selected until mid-August with work scheduled to begin October 1, this review is based on an evaluation of the materials and methods used to engage the Project Grant Manager, the current competencies of WestEd as both Grant Manager and Interim Manager and the RFP content and procurement process.

The application satisfactorily addresses the specified elements that have been included in the WA RFP issued for procurement of a Project Partner Manager. WestEd also conforms to the requirements and exemplifies strong credentials as evidenced by its mission, date of founding, experience and key personnel assigned to this project. The curriculum vitas presented described significant knowledge and skills in the areas of assessment management and development, technology and subject matter content. While the WA procurement RFP recognizes the need for specific management expertise evidenced by the requirement of 4 FTE Project Management certified personnel, it is not clear that this level of talent exists within the WestEd staff currently assigned to the Consortium. Significant project management expertise will be needed even on an interim basis to launch the comprehensive SBAC work plan. Additionally, allocation of only 50% dedicated staff time of the Interim project manager does not seem responsive to the level of start up effort that will be required.

The proposed work plan and timeline addresses primarily the 3 assessment development components. It includes a reasonable flow of activities and benchmarks which understandably will be further refined once the plan is funded, contractual engagements are secured and project management applications are implemented by the transition and permanent Project Manager. Specific budgetary resources in the amount of \$8.125 million are allocated to plan components focused on building professional capacity referenced in section A 6 professional capacity and outreach and collaboration with higher education. While this work is largely embedded in contracts expected to be negotiated with external organizations, an important aspect of the overall project, they should be included within the overall work plan and articulated in terms of major milestones, associated tasks start/end dates and responsible entities assigned. More specific commentary related to the implementation of these components is included in A 6.

The applicant speaks generally about risk mitigation strategies in describing its plan to work with states but gives little attention to this area except in its reference to possible computer adaptive failure. Under Section A 1, 15 SBAC members have identified barriers or risks covering a complex range of issues that would require either state legislative or policy action. MOU's submitted by these states, each includes action steps and timelines for removing the existing barriers, however, the application does not make clear what specific efforts will be undertaken by the Consortium to help states address these issues. Steps that will be undertaken by SBAC in this regard, should be incorporated into the project schedule development process and Project Work Plan/Timeline (Summary Table for A (8)(b)).

The SBAC plan consists of 7 Level 1 budget modules (Governance, Assessment Design, System Design, Research and Evaluation, Professional Capacity and Outreach, Technology and Higher Education Engagement) and one Level 2 Budget Module to develop 5 language translations, including sign language, of the Consortium's math summative and interim assessments. The overall expenditure projection totals \$149,987,819. Given the recommendations for plan revisions and further development, it is difficult to assess the adequacy of the budget. Consideration of the proposed changes, if acceptable, would need to be addressed both in plan re-design and possibly revised resource allocations. To this end, greater attention should be given to alignment of plan components, in particular, work plan objectives, implementation activities and budget modules. Specific budget concerns are described in more detail in the budget section.

The plan projects costs for the summative and benchmark assessments in the amounts of \$19.81 and \$7.50 per pupil respectively. The plan does not provide a specific time frame for maintaining these cost estimates beyond 2015 nor project potential inflationary factors over time. Consortium participants are expected to reallocate existing assessment resource allocations to pay for the new SBAC assessments. According to documentation provided in Appendix A 8-7, the costs of the new assessments should be easily sustained by the majority of the Consortium members. Increased costs over time would also need to be absorbed by Consortium members individually. Six states reported current assessment expenditures substantially below

the project SBAC pricing. Differentials range across the low of \$3.81 per pupil (\$3.4 million overall) in GA to a high of \$12.81 per pupil (\$3.9 million overall) in UT. The plan suggests that these projected cost differentials could potentially eliminate these member states from SBAC due to affordability, but that the remaining participants would be sufficient to meet the minimum state number required for RttT Assessment grant funding. SBAC has pledged to pursue cost reductions as well as to devise a cost allocation strategy that will allow all states to participate at the level of its current expenditure for summative assessments, however, the application lacks details regarding how such a cost methodology will be achieved nor how the system will be upgraded, managed and maintained over the long-term.

Recommendations

SBAC should:

- Reassess and increase allocation of staff time dedicated to role of Interim Project Management
- Adjust budget to align with plan revisions
- Provide more detailed plan for expenditure projections including future upgrade and ongoing system maintenance requirements and a financing strategy for the assessments over the long term
- Identify potential barriers/risks to program implementation and a plan to address them collaboratively with SBAC members
- Incorporate major milestones for contractual work into overall Project Management plan
- Develop long term plan for governing and managing the assessment system

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	15

Competitive Reviewer Comments:

SBAC has received Letters of Intent (LOIs) from 162 IHE/IHE systems across 30 states. There is no letter included for VT. There are several IHE/IHE systems representing 11 states that have pledge involvement in SBAC as well as another assessment consortium. The application does not address this cross over and it is not clear how higher education representatives will be distributed across the many advisory structures, design committees and working groups. For the most part, the LOIs follow a standardized format. All state collaborating IHEs or IHE systems have documented in their respective LOIs:

- a) A commitment to participate in the design and development of the consortium’s final high school summative assessments in mathematics and English Language Arts in order to ensure that assessments measure college readiness
- b) A commitment to implement policies, once the final high school summative assessments are implemented, that exempt from remedial courses and place into credit-bearing college courses any student who meets the consortium-adopted achievement standard for each assessment and any other placement requirement established by the IHE or IHE system. In the MO LOI an addendum is included along with the standard format components which allows the state to withdraw from the agreement if it is determined that the final assessments “fall short of institutional expectations.”
- c) Applicable signatures from the State’s higher education executive officer (if the State has one) and the president or head of each participating IHE or IHE system.

Strengths worth noting in the SBAC application include:

- support from private institutions as well as private IHEs in 4 of the 30 participating states-- MO, WI, NJ and AL
- overall number and diversity in types of institutions, e.g. 2 year colleges, entire systems along with individual colleges/universities, and institutions serving traditionally under-served populations
- exceptional strength of state commitments (90-100% of direct matriculating students in participating IHEs) as evidence by 9 states (CT, UT, ID, WI, OR, KS, DE, GA and NH)
- substantial impact that will result based on SBAC's partnership with higher education—over 700,000 students and a projected 74% participation rate (noted as a conservative estimate) across the total consortium (This number needs to be validated in accordance with NIA definitions and requirements)

There were no letters of support committing VT higher education institutions indicating support for the collaboration. Clarification is needed to assure that the data sources presented in the application summary chart meet the required NIA definition of “direct matriculation students” (students who entered college as a freshman within 2 years of high school graduation). The Summary Table includes information for all of the 30 participating states drawn either from state data or NCES IPEDS. The Summary Table cites IPEDS as the data source used for 28 states in calculating the total direct matriculation students except as in the cases of CO and OR, where the submissions provided by the individual State IHE/IHE System were higher than IPEDS. As described in the application narrative, IPEDS data are “based on a number count that includes first time degree holders and therefore, most likely higher than the direct matriculation number.” Establishing this correct base number that aligns with the NIA definition is important to validate an accurate percentage of the total number of direct matriculation students that will be impacted both within each state and across the Consortium.

Recommendation

SBAC should confirm that submitted data conforms to NIA definition of direct matriculation students.

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes

Absolute Reviewer Comments:

SBAC has created a strong proposal which describes a leadership, governance, participatory and organizational structure for designing and implementing a comprehensive assessment system and monitoring the grant. This ambitious educational reform agenda addresses each of the four priorities specified under the Absolute Priority along with the associated 10 sub-criterion. The plan's strategies and initiatives, if successfully implemented, hold great promise to increase student achievement and increase success in higher education by more direct matriculating students. The commitment from participating SEAs is strong. Overall, the SBAC is comprised of 31 total participating states with 17 serving as governing members and the remaining 14 functioning as Advisory. Further strength of the plan is evidenced by the Letters of Intent submitted a broad based group of 162 colleges and universities which have pledged to support the Consortium's work as well as the organizations that have offered to serve on the Policy Advisory Committee. The budget presents an overall and project level fiscal plan for allocating RttT

Assessment resources. This is undoubtedly an ambitious and bold commitment. With some important refinements including increased attention to equity, ongoing monitoring for effectiveness, broad based participation and identification of risk and barriers early in the process, SBAC should be well positioned to accomplish the goals of the RttT assessment competition.

Grand Total	220	150
-------------	-----	-----

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

Given the sizeable scope of work involving diverse participants, it should be acknowledged at the outset devising a budget is a complex undertaking. SBAC has requested an overall total of \$159,982,543 to implement its proposed balanced assessment system. The application consists of 8 separate Level 1 budget modules for a total of \$149,987,819. Level 1 modules conform with the format outlined in the NIA with varying levels of specificity by plan component. Budgets are presented in summary format and individual break outs by budget categories (e.g. personnel, fringe benefits, travel, equipment, etc.) as well as, specific modules (e.g. governance, assessment design, system design, etc.) which also include summary tables broken down by specific budget categories followed by a standard format addressing specific work plan tasks, contractual services. While the current expenditure proposal does not exceed the prescribed RttT Assessment fiscal parameters, it is very close to the \$150 million cap. The applicant gives minimal attention to how they plan to leverage other Federal, State, or philanthropic funds toward the design, development, implementation, and evaluation of the proposed Comprehensive Assessment System.

The Consortium will need to give serious consideration to how it will go about selected plan revisions which may necessitate a certain level of resource redirection including expansions in some categories and required reductions in other areas. Review commentary is included in several plan components throughout the application along with accompanying budget implications. Additional areas to be addressed or clarified are highlighted below:

- Clarification of the resource allocations for convenings of the Policy Advisory Committee
- Rationale for proposed varied pay rates for teachers and higher education faculty serving on work groups/advisory committees and consultants
- Clarification regarding the professional development curriculum support tasks included in Plan Components # 3 (Systems Design/\$50,000) and #5 (Professional Capacity/\$5,125,000)
- A more detailed scope of work for the professional capacity building component to justify the level and basis for the proposed \$5.1 Million in expenditures and the \$450,000 in planned expenditures in non SCASS fees. Plan should include clear objectives, per cent of personnel assigned from core staff, implementation strategies, deliverables (types of tools and supports), timelines and measures of effectiveness.
- Expansion of Outreach Component to incorporate or redirect resources targeted to community engagement (e.g. surveys and focus groups) and align with Research and Evaluation Component.
- Inclusions of technology applications to support convening and information dissemination, e.g. webinars, Skype, video conferencing, etc.
- Clarification of the proposed Higher Education Advisory Panel/Consultants and related travel/meeting costs budgeted @ \$1.3 Million. Details described under the description of services for the Advisory panel to participate in Working Groups is inconsistent with

Appendix A 1-3 wherein it states that the primary communication for Working Groups will be virtual meetings.

Level 2 Budgets

Name: Assessment Design - Translations

The SBAC Level 2 Module is requesting a total of \$9.6 million to support the translation of mathematics summative and interim/benchmarks assessments. The Level 2 Budget module argues for 5 separate language translations including sign language. Estimated expenses also include costs for reporting capability and ancillary developments @ \$125, 000 which are not specified in operational terms. While undoubtedly federal law and many state and local policies are demanding increased participation requirements for these special populations in all education activities, both as a means of establishing the needs and progress of individual students and for purposes of system accountability, the application does not provide sufficient justification to validate need. The application does not include a deliberative policy level procedure under the auspices of the Consortium governance structure to examine:

- the pros and cons of testing students with limited English proficiency in the student's native language
- the rationale and "beliefs" embedded in such a testing approach
- under what circumstance, if any, might native language translation be considered as an appropriate adaptation in large-scale testing

The application is absent a strong documented research base to support the proposed services and does not outline a detailed operational plan specifically aligned with the plan components of the larger assessment system. Direction for this development effort will fall under the responsibility of existing SBAC personnel, however, the application does not appear to have among its current staff complement nor described in its Project Management RFP specific staff capabilities in this area. It is anticipated that 5 different vendors will be engaged to perform the translations but the proposal does not give criteria for selection or required credentials. Lessons learned from prior considerations of suggests that there are several major challenges:

- Establishment of effective procedures for identifying and screening such students, so they can appropriately be included in appropriate assessment programs.
- Maintaining comparable test validity between native language assessments, those that are taken by the general population and regular assessments with accommodations to students with language challenges utilizing regular assessments

The Consortium's research and Evaluation Work Group will include as part of its work plan the development and implementation of comparability studies to ensure equal validity across the various math assessments. It is not clear where the required personnel and fiscal resources have been included that would make this possible. It is also unclear how the proposed native language assessments would be integrated into the overall assessment system. Given the inadequate resource allocation to several of the core budget components (i.e. professional capacity and outreach and evaluation) additional funding options above the \$150 million cap would be better spent in these critical areas.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b)
(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	10
(A)(1) Reviewer Comments:		
<p>(a) The proposed governance structure is top heavy. The Executive Committee and Steering Committee appear to have similar responsibilities. One of them could be eliminated or their responsibilities more delineated.</p> <p>(b) (i) (ii) The proposal clearly defines a state's role and differentiates the rights and responsibilities of those roles.</p> <p>(b)(v) Although the organization chart shows a number of places where there is technical input, it is not clear where the major technical issues – psychometric ones - reside. The psychometric properties of these assessments and their interpretations are at the heart of this endeavor. And what is being proposed pushes the psychometric limits of what we know and can do with either confidence or precision. The design of an assessment is obviously much more than its psychometrics; the implementation, however, will depend heavily on highly technical psychometric and statistical methods. The proposal should be more explicit about psychometric and statistical work and how it is represented in the governance schema.</p> <p>(d) The document does not describe how financial disputes between the lead state and other states might be resolved.</p> <p>Recommendation: This structure is top heavy. Preferably would be something like that of the National Assessment Governing Board (NAGB) where there is a representative of each major component sitting on the board. NAGB, for instance, has persons with policy perspectives, technical expertise, and reporting responsibilities. Persons related to what this proposal calls working groups are also there. In addition to those "functional" positions, there are representatives from varied interest groups – governors, state legislatures, teachers, etc. It is essential to have a forum where the inevitable "big" problems can be addressed from a variety of well-informed perspectives. .</p>		

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	2
(A)(2) Reviewer Comments:		
<p>(a) The proposal sets extraordinarily high standards for an assessment system. It is based on principles with various degrees of research evidence to support them. The proposal should distinguish between those activities where there is substantial knowledge and those which are less well investigated. The structure of the summative assessment as planned is extremely ambitious. For example, powerful performance tasks are not easy to construct and placing them on computers makes the work even more demanding.</p>		

There is little if any research evidence supporting the efficacy of benchmark testing. The theory should include ways to change emphases in the development phase when problems are encountered. It might be the case that one would first work on the construction of the summative piece and be willing, if necessary, to forgo the benchmark business.

(d) The proposal should reflect what has been learned about assessments that have as their dual purposes accountability and instruction. Twenty years of experience suggests that accountability will be emphasized to the detriment of instruction. How to minimize the influence of teaching to the test and test taking procedures should be a part of a Theory of Action.

Recommendation: Make allowances for making changes as the project unfolds. Plan B's are necessary since these huge enterprises inevitably must change directions. Formative notions, in fact, came from projects that did not do so well because they did not provide a way to change in the middle of the process. There should be some mechanism akin to formative evaluation put in place.

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	42

(A)(3) Reviewer Comments:

(a) The proposal reflects recent thinking and developments in the field. The mix of assessment components, item types, report and report types, using computer adapted testing, the emphasis on growth measures, all reflect current knowledge.

(b) (i)(ii) As written, everything in the proposal is deemed equally important and, most important and troublesome, equally doable. Yet more is known about writing items than knowing how test scores might be related to college-readiness, however defined. Measuring well a construct at one time point is easier than how one might detect how the construct changes over time. Growth models assume uni-dimensionality; there is not as much known and fewer well-established ways to discern multi-dimensionality and to model its possibilities. Setting achievement levels continue to be problematic. There is more empirical evidence for the usefulness of formative measures than for either benchmark measures or measures built on assumptions about learning progressions. There is a broader consensus about how to measure status but not how to model growth, especially at the classroom or school level. All of this is to say that competently completing each of the tasks is difficult but for some of the tasks doing them well will be exceedingly difficult and time consuming. This is related to the Theory of Action and a willingness to change course as the assessment is built.

Even if a state opts only for the summative test there is a lot of testing. There will be more work for teachers. The emphasis on mathematics and language subtracts from other important content areas. States may opt out when they see the degree of difficulty, the complexity and the size of the summative assessment, let alone the full package.

(c) (ii) The proposal includes ways to provide the required results as a component of the various evaluation requirements. Assessments are stronger, however, when they are used for a single purpose. Additional information is needed to fulfill the evaluation requirements. For instance, teacher effectiveness is broader than changes in scores in mathematics and language.

(c) (iv) Creating good items is the foundation of a good assessment. The proposal gave some very good exemplars. One example of a good idea that needs to be tweaked to conform to evidence-centered design is the technology enhanced item asking kids to construct a boxplot (X-17). That is a part of the standards and the item uses technology in creative ways to ascertain whether or not the student can create the box plot. As presently construed the data are organized by students' first names meaning that a typical respondent must first order the data before creating the boxplot. This means that if the respondent incorrectly creates the boxplot one does not know whether it is because of bad ordering or not knowing how

to construct the plot. Evidence-centered design suggests that the data should be ordered so an inference can be drawn about whether the student can construct a boxplot.

There is no obvious reason to order the data by first names in alphabetic order. The issue, of course, is that the form of the data display is likely to affect both the difficulty of the item and interpretation of results - both crucial aspects of evidence-centered design. For clearer inferences about a student's skills, it is important to separate ordering of the data from constructing the boxplot.

(c)(v)(vi) Computer adapted testing is a mode and method of testing that has as its strength the quick turnaround of results.

(c)(vii) Reporting is a crucial part of the assessment process. The proposal suggests producing a substantial number of reports for a variety of audiences. Past experience suggests that reporting has valued quantity over quality. Districts, schools, teachers, school councils, and especially parents, have been inundated with literally reams of virtual and real paper. Examples of the form of the reports especially for audiences outside of the school setting would have been desirable.

The Colorado displays should include error bands for the prior test scores and some indication that growth in achievement is not that smooth (i.e., the vertical scaling may not be that good). The graphic is large yet there are only a few data points displayed with lots of ink to cover a "page."

Experts disagree about how many grade levels can be legitimately scaled vertically.. This is because a vertical scale assumes uni-dimensionality of the attribute across, in this case, all assessed grade levels.

Recommendation: The Oregon system referred to in the proposal seems thorough. It should be evaluated to find out who uses the system, for what purposes, and how much training is needed to use it effectively.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	30

(A)(4) Reviewer Comments:

(a) This is a strong section of the proposal.

A crucial feature of the work is the development of the framework and test and item specifications. Although those products inform the test construction process, they can also be used, in perhaps less complex forms, to inform the various publics about the nature of the assessment.

Criticisms of earlier attempts at learning hierarchies apply to the learning progression business. These are very good ways to think that organizing instruction - teaching ; they may be less powerful in terms of defining learning. Students learn in a variety of ways in multiple contexts. There is little, if any, published research, especially in mathematics and English language, of the efficacy of learning progression as a basis for test construction. This is important to note if the construction phase of the assessment is to be heavily influenced by learning progressions.

(c) One wonders if the proposal is a bit too sanguine about the possibilities of AI scoring. It is honest, however, in pointing out that much of its success in scoring results of prompts requiring complex responses depends on further development of AI technology. The use of teachers in the process is desirable. There are those who say it is the best kind of professional development.

(e) Computer adapted testing is a strong way to get good measures of student proficiency at the extremes of the scales. The proposal puts in place ways to gather evidence about how well it works across settings and populations.

Recommendation: In the training, there should be a heavy emphasis on uncertainty in the scores and on the AERA/APA/NCME standards about not using one test score as a basis for making important educational decisions.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	20

(A)(5) Reviewer Comments:

(a) A large and complicated assessment presents equally large and complicated problems of proper test score interpretations. The proposal is correct to emphasize the "validation of the cognitive model ... for each component ... across all student performance levels.

Of major importance is the definition of the constructs, the setting of achievement standards, and defining and defending a growth measure. The proposed activities include ways to gather data about these issues.

The computer adaptive testing assumes a uni-dimensional scale across grade levels which, also assuming a sufficiently large and well-defined item pool, allows in theory comparably precise measurement across the performance distributions. There are further assumptions about the distributions of achievement across grade levels and across states. Each of these assumptions will be investigated according to the research and evaluation component of the proposal.

There is no agreed upon method for setting proficiency standards/achievement levels (the NAEP achievement levels are still considered experimental) yet the proposal advocates a number of standards at each grade level. The proposal does, however, include ways to investigate the definitions, constructions and interpretations of the proficiency standards.

What is growth and how should it be measured is question not completely answered by the proposal. In several places in the proposal, the internal consistency of measures are mentioned as indicators of reliable score interpretations. For growth measure, statistics that reflect the stability of the scores and the properties of the score differences are more appropriate.

The proposal makes the sound recommendation that states be able to use a variety of growth models. Whatever growth model or models are adopted, the contractor should be in a position both to offer guidance and provide evidence of the efficacy of the endeavors.

The proposal rightly includes a section about the comparability of scores across jurisdictions.

There is a sentence in the proposal that suggests looking at classification accuracy when it comes to interpreting scores emerging from the proficiency or achievement standards. That is a proper suggestion and it is assumed that such analyses will be conducted each time the cut-scores are used as a basis for interpreting the results.

Recommendation: In determining whether and how much there is, it would be desirable to conduct studies of a student's opportunity to learn the material on the test. It is unclear how the curriculum of a state and which students are exposed to what parts of it may be related to the achievement outcomes. Opportunity to learn, or lack of it, may be a crucial component of score comparability.

Recommendation: Evaluation designs should attempt to look at all effects, not just the intended ones.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	6
<p>(A)(6) Reviewer Comments:</p> <p>(a) The proposal describes a range of activities that could be used to enhance knowledge of and participation in the assessment system. Teachers in particular will benefit as active participants in the production of the assessment.</p> <p>In presenting the “educator dashboard” the proposal misses the single most important tool for helping persons understand the assessment – released items. Parents, school board members and other non-professional persons often change their opinions about assessment when they have an opportunity to see what is being measured or actually take a test. For example, most persons will find 8th grade mathematics items fairly difficult and will walk away from the activity with a better understanding of what is being assessed and often a change in perspective in terms of what students are expected to know and do.</p> <p>(b) While the proposal presents a variety of outreach activities, there are additional audiences that could be addressed. The proposal mentions state policy makers but does not specifically target legislators and their staffs. Members of the medial might also be included. How and when these activities will take place is not presented in detail.</p> <p>Recommendation: In addition to the legislature it is essential that the state's Department of Education personnel be trained in all aspects of the assessment. Often those representatives pontificate with only partial knowledge.</p> <p>Recommendation: The proposal should remind us how the professional development activities are to be evaluated. Professional development activities are thought to be most powerful when they change not only what persons know but what they do. How will we know if the activities are effective?</p>		

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	5
<p>(A)(7) Reviewer Comments:</p> <p>(a) Technology is used in virtually every aspect of this proposal. In fact, there is incredible dependency on technology. In some cases – allowing paper and pencil measures for three years for example – there are contingency plans. (That raises comparability issues.) In others – computer adapted assessing and AI scoring – there are no contingency plans. There is no contingency plan if AI technology does not develop in the assumed manner.</p> <p>(b) States have had a great deal of difficulty building data bases across grade levels and through levels of schooling for example. Glitches in the software to equate tests or improper use of the software has led to companies being fired. Almost always there are items that are mis-scored. The proposal deals with some but not all of these issues.</p> <p>Recommendation: Each component should have a Plan B - what to do if things do not go according to plan.</p>		

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	8
<p>(A)(8) Reviewer Comments:</p> <p>(a) The majority of the work will be conducted by the Assessment Design Working Group about which very little, if anything is known. The Executive Committee is asked to make a number of very important decisions. The composition of the committee as described earlier in Section (A)(1) does not have the right mix to make such decisions. There should be broader representation on the committee, especially in regard to having technical expertise represented. A strength of this part of the proposal is that WestEd which has broad experience and acclaimed competence, is the choice for interim management of the project.</p> <p>(b) The timeline contains the crucial elements in what appears to be a logical progression. But, there remains a major question of whom will be doing the work under what circumstances.</p> <p>(c)(ii)(iii) The proposal appears to reasonably allocate resources to produce the desired system. It includes references to each major component specified in the criteria.</p> <p>Recommendation: If, as stated earlier, one level of the management hierarchy is eliminated, one should look at its implications for the budget.</p>		

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	15
<p>Competitive Reviewer Comments:</p> <p>The proposal appears to have at least 30 percent of students represented and therefore meets the requirement.</p> <p>(b) The memoranda are complete. However, the states varied greatly in their responses. Some states have just one letter from the state higher education authority. Others have letters from individual universities. It is difficult, therefore, to determine exactly the strength of commitment. The fact that 30+ states signed on is a good first step.</p> <p>(c) Each of the states have the required letter.</p>		

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes
<p>Absolute Reviewer Comments:</p> <p>The proposal is strongest and most responsive in the areas of assessment design and development. The latest assessment techniques are advocated.</p>		

(a) There is a question of whether one set of measurements can do both career and college readiness.

(a)(iii)The use of computer adapted testing, and the goal of having adequate measurement precision across the achievement domain, implies a large pool of very good items. How that pool will be articulated in terms of, say, learning progressions, and applied comparably to each student, is neither clear nor easy to do. The strong assumption is that the pool contains vertically equated measurements and that each testing component loads properly on the one dimension being measured. It is interesting that the notion of testlets was not introduced.

The properties of the resulting test score interpretations must be carefully investigated. One assumes the achievement constructs do not change over time. According to the specifications each component of the system – multiple-choice, technology enhanced, performance based – will be represented appropriately in the pool at all levels. Other assumptions are that items load comparably on one underlying scale and the same scale is similar across grade levels, and they “grow” at the same rate. These are some of the interpretations that must be justified.

The foundation for the proper use of test scores is, of course, a good test score. Implementing this proposal could provide those. But good measures of achievement are not necessarily all that is needed for a number of purposes mentioned in our evaluation schema. They cannot automatically be used for judging, say, teacher or school effectiveness. A much stronger argument must be made about appropriate models and additional data for such uses. The results from one test should not be used as the sole basis for making decisions about any important educational matter.

(a)(iv) The proposal finesses the growth specifications. And rightfully so! There is no one model that is universally acclaimed. The competing models make varied strong assumptions about the data and have, as one might expect serious weaknesses.

(b)(ii)The best predictor of success in college is successful performance in a rigorous secondary school course of study. High school grades and class rank are inevitably stronger predictors than test scores when colleges and universities do validity studies. The notion of a score reflecting college readiness should be described within the context of a student having had the opportunity to pursue a rigorous academic program.

(d) For each of the sub-criteria, (i) through (iv). It is necessary to be clear about what “used to inform” means. It should not mean used to determine.

Elementary school teachers, for example, teach content other than mathematics and language, so the accountability system should include such information. The problem is that states may not measure such outcomes.

Grand Total	220	138
-------------	-----	-----

Budgets

Level 2 Budgets
<p>Name: Assessment Design - Translations</p> <p>When compared to the amount in the budget for Research and Evaluation, this allocation, which is almost twice as much, seems too high. If this is an adequate budget for translations then the Research and Evaluation budget should be increased.</p>
Level 1 Budget
<p>Name: Level 1 Budget(s)</p>

Over 80% of the budget is devoted to Assessment Design and Technology components. That is putting the emphasis where it should be.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b)(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	15

(A)(1) Reviewer Comments:

SBAC's vision and goals for a "next-generation assessment system" are driven by clearly expressed and commendable values and beliefs about improving teaching and learning. The proposed system would include a rich array of item types to address the full range of the Common Core State Standards in ELA and mathematics through a balance of formative, interim, and summative components. One highlight of their plan is to utilize computer adaptive testing, as a key means of attaining adequate and accurate information about all students: low and high-achieving students, English learners, and students with disabilities.

A substantial number of States have committed to this Consortium, either as Governing States (17) or Advisory States (14). The roles and responsibilities of each category are clearly outlined in the proposal. SBAC also clearly identifies anticipated processes for decision-making, changing roles among member States, and adding/removing member States.

The proposal is also reasonably thorough in outlining the organizational structure. The various committees and working groups are described in the narrative and Appendix A1-2, and their interrelationships are illustrated in Appendix A1-1; however the proposal fails to make clear the nature and interconnection of task forces such as the Technology Implementation Task Force whose function is later described in (A)(7) or their decision-making processes. Without this information, there is some risk that decisions may be made without sufficient diversity in point of view or input from, and consideration of, different (and sometimes highly specialized) areas of expertise.

While the Consortium processes for determining Steering and Executive Committee/working group members and elected leaders is partially addressed in Appendix A1-3, there is no detail at all in the proposal regarding the selection of members of other committees (e.g., Technical Advisory Committee, Policy Advisory Committee). In addition, it is not clear who within each working group would provide leadership. It should be noted that the Consortium has secured commitments from an impressive array of organizations to serve on the Policy Advisory Committee.

The proposal sets forth the dates by which steps towards adoption of key policies and definitions will be initiated and completed; policies not outlined in their Summary Table for (A)(1)(b)(v): Consortium's Policy and Definition Time Line are to be developed and determined by the Executive and Steering Committees.

Substantial information is provided to support the role of the State of Washington as the Lead Procurement State. The signature of every member State's procurement officer was included as demonstration of their commitment to the Consortium's procurement process.

Each State agreed to identify existing barriers to implementing this assessment system. Many of the MOUs from member States included details on existing barriers, action(s) needed to remove those barriers, and the timeframe for doing so. In instances in which this information was absent (e.g., Missouri, Nevada, Wisconsin, Hawaii, Vermont, Kansas, Montana, West Virginia, Ohio, Colorado, New Hampshire, Pennsylvania, Oklahoma, New Jersey [note: missing this page from South Dakota]), SBAC could have avoided ambiguity by having member States enter N/A or "none," since without this one must assume that those States leaving the chart blank identified no barriers whatsoever.

Based on all of these observations, the response received a score in the low "high" range.

Suggestions to DOE:

It seems advisable to seek from SBAC a more detailed plan to flesh out ways and times when various committees and task forces will interact, as well as processes by which their decisions will be made. Particularly when dealing with highly specialized or technical areas, it is critical that experts inform discussions and decisions reached by the Executive and Steering Committees.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	3

(A)(2) Reviewer Comments:

The SBAC presents a coherent theory of action supported by a number of sound and clearly articulated principles (see Seven Principles Undergirding the Theory of Action). The credibility of this theory of action is subject to challenge, however; this is because while adherence to some principles is/will be easy to demonstrate (e.g., #7: Design and implementation strategies adhere to established professional standards), evidence of adherence to others may be less-well understood or less-easily defined (e.g., #3: Teachers are integrally involved).

Particularly given that this proposal includes numerous references to cutting-edge approaches to assessment design, delivery, and scoring (among which are anticipated outgrowths of ongoing efforts of various members of this consortium), SBAC's theory of action would have been strengthened by a more moderate stance. The stated intent to "radically reshape the education enterprise in participating states" does a disservice to states (consortium members and others) who are engaged at present in innovative assessment enterprise in support of improved teaching and learning, as does the reference to "the current 'drop from the sky' approaches to educational testing."

The SBAC has clearly made good use of member states' experiences in their presentation of the components of their theory of action, particularly as regards communication of SBAC policies and standards. However, this section of the theory of action includes claims--particularly regarding the impact of teacher-moderated scoring as professional development and the impact of technology--for which there is insufficient substantiation elsewhere in the proposal.

SBAC conveys in general terms the causal (and recursive) relationships between specific components and the desired outcomes of their comprehensive assessment system (supported by Appendix A2-1: Overview of the Theory of Action). This graphic representation is not always supported in the narrative however (for example, how consortium and state policy and practice in support of high expectations and increased learning opportunities feeds into the role of technology); furthermore, it appears to be missing at least one key component--the training of teachers required in order to allow them to effectively design and score assessments *using technology* (a new experience for virtually all teachers, even those who have considerable item and task writing experience).

For these reasons, the response was scored at the "medium" range.

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	38

(A)(3) Reviewer Comments:

The SBAC's assessment system proposal is innovative in a number of ways: in the proposed use of computer adaptive delivery, the consideration of a "through-course" option, and in the conceptualization of item types and delivery mode.

- The computer adaptive feature of this system has the capacity to increase accuracy of measurement for students along the full range of ability (more precision than currently available for those at the lowest and highest ends) while controlling the amount of time required for testing (a continuing issue for teachers, parents, students, and other stakeholders). A computer adaptive model is also capable of supporting the accurate measurement of individual students' growth of time.
- Particularly impressive is the SBAC's thinking about a "through-course" option, and its recognition of the desirability of flexibility within the constraints of comparability and fairness. The consortium's plans to implement a comprehensive research agenda to ensure technical adequacy and equity while promoting flexibility is commendable, as is the flexibility manifest in the option to offer an adaptive summative assessment in Grades 9 and/or 10 as well as Grade 11.
- The SBAC demonstrates both considerable familiarity with, and confidence in, innovative item types that capitalize upon technology; while one may take issue with the degree of optimism about emerging technologies (both as applied to the development of technology-enhanced items or scoring methodology), the consortium is to be commended for their commitment to continued research and development in this area (planning, as they are, to "invest heavily in prototyping and testing innovative items during the assessment design and development process": see Appendix A.3-A).

The strength of this proposal in describing and justifying the use of computer adaptive summative assessment is not as well sustained when the narrative turns to the second key component of the system. Adaptive I/B assessments are intended to inform ongoing instruction, and as such would be a valuable component. However, neither the proposal narrative nor supporting documents for this section expand upon the description of these items and performance events as ones that will provide "more finely grained information." Since the intent appears to be to provide an item/task bank (a "pool of items") to teachers and instructional leaders containing the same types of items as well as released summative items and events, it remains somewhat unclear what distinguishes I/B assessments from the summative assessment other than timing.

The third component (formative tools, processes and practices) is given the most cursory treatment, raising several concerns, foremost the lack of detail on ways that formative assessment design often differs intentionally from summative assessment; sometimes these items/events may include more scaffolding (e.g., constrained sources of information, breakdown into component steps, etc.). This type of resource may also include tools for student self-and peer assessment, a documented means of building understanding and helping students reach a given performance target. The proposal would have been enhanced by more clearly differentiating the three components.

The battery of proposed components is such that this assessment system should be able to address even those standards (e.g., speaking and listening, reasoning and critical thinking) achievement in which has been more challenging to measure, and provide accurate measures across a wider performance continuum than is currently reflected in state assessment systems. While the proposed system plans to attend to all the CCSS in English language arts and mathematics (the "full range" called for in the request for proposals), there is little in the proposal to suggest that balance and/or the intentional weighting of some standards more than others will be considered (other than general references to "blueprints" and test specifications).

The consortium has appropriate plans for research and alignment studies to inform the assessment blueprint as well as for standard setting to translate the CCSS into performance standards. While the consortium asserts that their proposed blueprint "provides sufficient data across the clusters of the CCSS to measure achievement, the proposal would have been strengthened by providing at least some suggestion of issues and implications of the scope of information on individual students required (in other words, some explanation of how the need to obtain sufficient performance data for each student on every standard may impact the time required for assessment purposes and/or the proposed number of items/events). There is no explanation or justification for the number of items of each type attached to any of the summary Tables for (A)(3) (ELA-Summative Reading; ELA Summative Writing, Listening and Speaking, and Language; Summative Mathematics; ELA-Interim/Benchmark Reading; ELA Interim/Benchmark Writing, Listening and Speaking, and Language; or Interim/Benchmark Mathematics.

The SBAC proposal contains a detailed and thoughtful plan for achieving accessibility for students with disabilities. Given the possibility that even under the best of circumstances, some items may not be accessible to all students, it would have been useful to know the consortium's thinking regarding exemption/substitution at the item level, the possible use of imputed performance based on partial measures, and other related questions. A more serious omission appears to be detail on how this system plans to address the needs and improve the measure of intended constructs for English learners.

Overall concerns with section (A) (3) of the proposal center primarily around feasibility. First and foremost, while consistent with the articulated theory of action, the detailed role to be played by technology at all phases of the assessment is not necessarily consistent with what the assessment community knows or even believes it realistically will be able to do during a window of time like the one covered by the RTT-Assessment Grant. Although it was good to see acknowledged the likely need for "teacher scoring" (which one might anticipate winding up being human scoring by teachers and/or others), there was no defense of, or explanation for, specifying a two-week window for anticipated turnaround of items requiring teacher scoring (p. 49). This may in fact be unrealistic. Also perhaps unrealistic are the claims regarding the scope of information available through the system's formative assessment. Particular if AI scoring is used for writing, it is questionable how specific or useful feedback may be, at least based on the most sophisticated and successful essay-scoring technologies available or in production at the present time.

The format and cueing of some of the examples of items in Appendix A.3-A do not lend themselves to AI scoring without very intensive efforts to build a bank of content elements from a very large sample of responses first scored by human raters, and there is no evidence in the section of the proposal to which these items are appended that such a plan is in place or even anticipated. Another feasibility issue related to technology is the need for access by teacher scorers to online databases and/or libraries (not a given at this point in time according to the recent NCES report on technology resources in America's schools).

It is worth noting here a concern that without careful item reviews that focus on precisely this issue, it is easy for technology-enhanced items to wind up being little or no more than a "fancy" computer-delivered version of what could be presented more easily as a computer-adaptive selected response item (something akin to "fake constructed response" items that use CR format where SR would be more efficient and just as effective). Just using "drag and drop" capability of a computer, for example, does not lead to a better, more meaningful, or higher-order item.

The SBAC's plan to provide a paper-and-pencil option for a limited time is both realistic and appropriate, particularly given a lingering "digital divide" in terms of computer access and functionality in school districts nationwide. Although more and more computers are available, significant discrepancies exist in capacity for course management and delivery, online access to databases, and access to online libraries, just to name a few areas of concern. This continuing gap will impact not only the delivery of the assessment itself, but the distributed scoring process and dissemination of professional development resources (which are described as using electronic technology as the platform and for the tools for professional development).

Quite striking was the absence of a rationale supporting the use of computers by students across all grade levels covered by this assessment to respond to ECRs, TE items and performance events (e.g., in the production of written texts or otherwise making extended use of keyboarding). Too little is known at present about the relationship between age and the ability to perform some of the more complex manipulations of screen images, databases, and online resources to assume that computer adaptive assessment for all grade level standards is advisable. Existing research supports the idea that students need to internalize various functions first so that computer use does not confound measurement of the intended construct or create greater difficulty.

Finally, in regard to the extent to which the assessment system design is consistent with the theory of action—many key elements of the SBAC theory of action are evident in the narrative on Assessment System Design. However, although SBAC points in this section of the narrative to the need to balance "the sometimes conflicting values" of flexibility, comparability, and technical adequacy, and refers to "the various values explicit in our Theory of Action," there appears to be no mention in SBAC's theory of action of flexibility, comparability, or technical adequacy. It is unclear whether SBAC's intent was to allude to the seven principles they have identified that are shared by assessment systems of high-achieving nations and States in the U.S. or to "components" of the Consortium's theory of action. Less casual substitution throughout of words like "principles," "values," and "components" might have helped to prevent any confusion.

In light of the various strengths and weaknesses in this section of the SBAC proposal, it was scored in the high "medium" range.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	23

(A)(4) Reviewer Comments:

The SBAC proposal clearly establishes as the starting point the specification of learning constructs to be measured and identifies the "hallmarks" of their approach. They are to be commended for their commitment to the transparency of the development process. There is no mention, however, of grounding in and documentation of prior research supporting steps and decisions related to their approach and strategy for assessment development, something that would certainly enhance transparency and contribute to stakeholder "buy-in" for the system.

The proposed starting point in the development of assessment frameworks built upon the CCSS is necessary but incomplete, as described in this proposal. From the outset, it is critical to engage not only key stakeholders and content specialists but scoring experts as well—those who can identify the ways that what students know and should be able to do related to each content standard are measurable and what formats/item types may be best suited to the acquisition of evidence of that learning. Past experience of numerous state assessment programs provides examples of the problems and pitfalls of moving forward in too linear a way, without examination up-front of the implications each standard has for the development of evaluative criteria (whether applied by live raters or via AI). More recursiveness in the process is advisable.

As noted in comments on section (3)(A), it will be important that attention to the "full range of the CCSS" address not only scope of measures, but also their distribution and weight in the assessment.

It is not clear from the narrative for this section whether SBAC envisions some items/performance events eliciting multiple measures—that is, measures of more than one objective for a single or multiple standards. Research and practice both support the feasibility of doing so, but with a critical caveat that the opportunity to measure something is not the same thing as an occasion when it will with certainty be measured. The brief paragraph dealing with the scoring of performance events with "a multiple-content-domain rubric" was insufficiently clear about whether this would involve one multi-dimensional rubric yielding scores that contribute to different scales or the application of multiple rubrics to a given performance.

SBAC has clearly given much thought to the use of design templates for generating items and performance events, and has highlighted the benefits of this approach. Experience with such templates has demonstrated that they certainly offer many benefits; however, drawbacks are possible as well, including a distortion of the construct being measured and over-simplification of evaluative criteria that may lead to misrepresentation of some students' knowledge and skills. Of particular concern, given the optimistic claims made for templates as a resource for teachers "to generate rich classroom assignments and curriculum-embedded assessment tasks" is the documented capacity for item and task templates to inappropriately narrow instruction. One example is the use of templates for writing prompts—something that is a requirement for computerized scoring of writing—leading to highly formulaic essays and a reductive view of writing for teaching and learning.

Strengths of SBAC's item development plan include the "repurposing" of items from existing State item banks (making sure they meet alignment and quality criteria) and the involvement of States in developing new items. It would have been helpful—and have enhanced the proposal—for some detail on anticipated number of items to be developed per year, to demonstrate the feasibility of this item development plan.

The item review process (quality control) outlined by SBAC is well-conceptualized to address the need for valid, reliable, and fair measures of student learning.

SBAC has demonstrated considerable thought about the development of common understandings, guidelines/policies and practices related to EL and SWDs. This is evidenced by Appendix A4-3 (Goals for Enhanced Assessment Grant for Accommodations).

It is in regard to SBAC's plan for scoring that some of the deepest reservations or concerns arise. The Consortium asserts that they "will build on the existing methods and capacities for automated scoring"; however, these very methods call for much more up-front labor than is acknowledged. Specifically, programming of scoring software for scoring extended constructed responses and performance events (whether an analytical or natural language approach is used) calls for a considerable body of pre-scored responses to the item or performance event to be eventually assigned scores by computer (a training set from which to identify predictive features). The scope of field testing may not be sufficient to support provision of the number of examples needed, particularly since programming calls for a large corpus of successful/high scoring responses. It is also worth noting that given plans for a certain percentage of teacher-assigned second scores, all training materials must be developed, subject to field testing, and used in the identification of exemplars for programming purposes well before teacher scoring is introduced for professional development or other purposes.

This is much less of an issue for the technology-enhanced items that resemble, in scope and format, many prior-generation brief constructed response items that were typically scored by reference to "answer keys" and "answer cues" (a menu of likely effective/correct response elements). However, in both instances (extended constructed responses/performance events and technology-enhanced items) care will need to be taken to build in a "kick-out" feature whereby unusual responses (which may still have merit) are handed off to a human rater in a timely and efficient manner.

Plans for the monitoring of scoring are vague, with no indication of who will conduct such monitoring, in what timeframe, and how (although it's fair to make the assumption that this too would involve technology).

The treatment of quality control measures was thoughtful and appropriate, although insufficient. It would have been helpful to have more detail in this regard about how the proposed item management system will be used to document previously generated items contributed by member States to the item bank. In addition, while some information was provided about agreement rates expected during training, there was no indication of the percent of exact and/or adjacent agreement in order for a given rater's scores to count or what provisions will be made for review and revision of scores given by weak raters.

In contrast, the SBAC's proposed plans for reporting are detailed and sound. It was helpful for them to include various examples of reports in the appendices.

Overall, the SBAC's plan for developing the proposed assessment system covers all aspects of test development, including some innovative and/or particularly well thought out ideas; some parts of the narrative remain too general, however, and the projected results of some proposed strategies seem overly-optimistic. In particular, the claims made for uses of technology—given the state of the art in technology-enhanced item development and AI scoring—seem unrealistic (overstated and overly-optimistic), raising doubts about the readiness and affordability of the proposed system, in the format described, by 2014-15.

It is relevant (and substantiates concerns expressed here and in other sections of this grant review) to note that of the 17 out of 31 member States that identified existing State barriers in their MOU included in the proposal, five explicitly identified existing and/or anticipated technology issues as a barrier, five identify budgetary barriers (which tie into cost-effectiveness), and two identify contractual barriers that relate to expectations regarding teacher involvement in various phases of development, scoring, and outreach.

For these reasons, and based on issues/concerns raised in the review of this section, it scored in the high "medium" range.

Suggestions for DOE:

The Department of Education may wish to look into yet another approach to supplement SBAC's item development plans. While only a few states (as noted in the proposal) have experience developing technology-enhanced items, far more have a history of developing brief constructed response items for reading and mathematics—the sorts of "activities" that characterized New Standards tasks, for example. It

is not difficult to convert these into technology-enhanced items (and in fact, many of the sample items submitted by SBAC looked remarkably like the earlier items, but with “drop and drag” or electronic drawing tools replacing the pencil or pen as response tool). This strategy would improve the cost-effectiveness and timeliness of item development given the window of time from award to implementation of the proposed assessment system.

Although it will impact cost and timing, the Department may wish to consider having the Consortium implement a “co-reader” function initially (one AI score and one human score); this is much more expensive, but may be a useful—even necessary—transition to learn more about AI scoring for at least the most complex performance events.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	22

(A)(5) Reviewer Comments:

The Consortium’s Research and Evaluation Plan is generally strong and detailed, and is consistent with their theory of action. The proposal plan to ensure that the assessments are valid, reliable, and fair for the purposes for which they are intended is well grounded in research (for which at least some key references are provided). Particular strengths of this section of the proposal are:

- the detailing of many and varied collaborations within and among key groups (the Consortium’s Research and Evaluation Working Group, the Technical Advisory Committee [TAC], and the Steering Committee
- acknowledgment of tradeoffs (from a psychometric perspective) that may be desirable in support of key concerns, especially that of flexibility; the argument against adoption of a single growth model
- the recognition of the need for—and description of the function of—professional development providers to ensure accurate and widespread dissemination and comprehension of comparability guidelines
- the planned testing of assumptions related to differences between adaptive I/B testing and adaptive summative testing.

Another strength of SBAC’s research and evaluation plan is their intention to report and disseminate findings, an action that will certainly contribute to the goal of transparency.

SBAC’s plan for evaluating the assessment system was clearly presented; especially thoughtful are some of the long-term steps outlined (e.g., consideration of unanticipated and unintended consequences—both positive and negative—that emerge over time) and specific evaluation questions that are likely to be addressed at various stages (beginning, transition to full implementation, and in the long term).

While SBAC’s research agenda is ambitious and forward-thinking, there appears to be very limited attention to research to address the many issues and implications related to the many uses of technology in the proposed assessment system (with only examination of characteristics of innovative technology-based performance tasks or events and reliability of AI scoring included under areas of research). Particularly given that SBAC acknowledges AI scoring as an “emerging” technology and recognizes that development of technology-enhanced items and performance events are still in a nascent stage compared to paper-and-pencil assessment formats, it behooves SBAC to identify relevant research questions and strategies for addressing them. As a related point, the statement that “the scoring process will capitalize on emerging technology as feasible” in a footnote about monitoring the reliability of AI scoring seems to run counter to the plans identified in the summary tables in an earlier section (A) (3) for 100% of all ECRs to be AI-scored and performance events to be scored using a combination of AI and teacher scoring. If SBAC does not have full confidence in AI scoring at this point in time, more detail about related research and contingency plans is warranted.

A few additional concerns:

- Interviews are identified at one point (under Validity and Fairness in the section on summative assessment) as a source of qualitative data about items and performance events, but there was no mention of "think aloud" protocols, a valuable data source; however, at later points in this section there are references to "cognitive interviews." More clarity about intended qualitative data sources would be helpful.
- Recognizing the need for tables to be streamlined, some of the entries in Table A5-1 are so general as to not be helpful (e.g., the intended focus and goals of activities such as "review contingency plans" and "observe test administrations" are not clear).
- It was also surprising to see no reference to the digital divide under the topic of Fairness, since there very well may still be access issues within and across school populations.
- Use of technology also brings with it a whole host of test security issues, mentioned only in terms of proposing a research and evaluation process that will support "implementation of quality control and test security measures" for a variety of purposes.

Given that issues and concerns center on only one aspect of the assessment system (albeit a critical one), this section was scored at the low end of the "high" range.

Suggestions for DOE:

Regardless of whether SBAC's research agenda is expanded to deal with the impact of technology in all aspects of the assessment—test development, scoring, reporting, professional development, etc.—this is something that the Department of Education should explore further.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	9

(A)(6) Reviewer Comments:

The response to this section appears fairly general; SBAC acknowledges the enormous importance of building professional capacity of teachers and administrators and support among all stakeholders. It is not sufficiently clear, however, how the means outlined to do this will accomplish intended ends. Specifically: while it is commendable—and indeed may be essential for a whole host of reasons—to involve teachers in test development (including creating and reviewing items and performance events) and scoring, experience with, and research on the impact of teacher involvement on subsequent teaching and student learning is complex and not necessarily as intended. Sometimes teachers become reductive in their thinking about item/task demands, the embodiment of standards in given items, and the application of scoring criteria, for example. This is not to say that involvement of State educators is not highly valuable and should be encouraged; however, such experiences as test development and scoring alone may have negative consequences without carefully designed support materials to help teachers "translate" what they have experienced into applications suitable to the classroom.

A comprehensive communication network is another feature of the applicant's plan for professional capacity and outreach—again, commendable, but not without some potential problems at this point in time, when the most recent report from NCEs on school-based technology resources indicates that there remain a considerable number of schools and even whole districts as yet unable to depend on technology-supported professional development and dissemination of information (e.g., through webinars or videoconferences). Particularly for this reason, it was good to see a host of other tools and resources suggested by SBAC, among them recommended readings, group discussions, sharing of best practices, a reminder that while more efficient and perhaps more effective, not all networking need be digital.

It is important to note that the section on the use of technology to support professional capacity is quite exciting and forward-looking. It is likely that all of what SBAC proposes is doable, if not immediately, then

over time. However, it is not clear that this plan is realistic from a practical standpoint without addressing the time it takes to build such an online system, populate it, and maintain it both in terms of the technology and in terms of content integrity. As the assessment system itself must meet the needs of a diverse student population, including EL and SWD, it would seem that a web-based resource needs to as well; this is something it would have been helpful to hear more about.

In considering coverage of issues and topics in any outreach and communications plan, it will be important to include information about AI and to be proactive. At present, most stakeholder groups identified (teachers, parents, the public at large, for example) probably have little or no understanding of what it means for a product or performance to be "scored" by a computer (assigned what amounts to a prediction of a human rater's judgment using syntactic elements, analysis of rhetorical structure, and topical analysis, etc.), or what the attributes of an item are that make it amenable to online delivery and scoring. Particularly in light of strong positions at both ends—both for and against computerized scoring of extended constructed response items and performance events—care must be exercised to present information on AI objectively, neither over-selling the technology nor ignoring those (the Conference on College Composition and Communication, for example) to whom computer-based scoring is anathema. These technologies have great promise; however, the best advice would be "proceed deliberately."

With the reservations noted above, the plan is generally feasible and consistent with the applicant's theory of action, and thus received a score in the high "medium" range.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	6

(A)(7) Reviewer Comments:

Use of technology informs virtually every proposed aspect of the system proposed by SBAC. They identify many exciting and forward-looking ways of integrating technology in the development, administration, scoring, and reporting of results of all components of their system. Proposed uses of technology for data analysis and access are realistic and reachable within the lifetime of this project.

Misgivings arise, however, based on a number of factors:

- While development of new technologies that will support the proposal is certainly ongoing, reports from a wide array of projects (district-wide and state assessments and/or exploratory studies, along with NAEP's various forays into use of computer-based assessment, for example) suggest that there are still many unknowns that may very well impact when, and if, all of the ways in which SBAC proposes to utilize technology are realistic and/or cost-effective.
- At this point in time, there is still a digital divide between and among American schools and school systems. The proposal lacks sufficient evidence of the capacity to develop, operationalize, and maintain the system portal.
- Technology is already in place to support item/task development (one use of technology proposed by SBAC) and has been demonstrated to enhance this aspect of assessment. Although many very useful models are proprietary at this point in time (e.g., those used by commercial test development companies), it is realistic to foresee online development and tracking of items and tasks; however, no detail is provided by SBAC to demonstrate that the Michigan online item-authoring and banking system, mentioned in (A)(8) as available for use as a cost-saving measure really is well-suited for, and can handle, the wide range--and more complex types--of items envisioned by SBAC. Furthermore, it is not clear how already developed items and performance events would be folded into an electronic item pool (whether, for example, previously created items will be merely "warehoused" or supported by the same or similar kinds of information to make development, review and refinement transparent). A related concern is that since electronic tracking systems retain evidence of virtually every change or comment, without a carefully designed hierarchy of users, the history of items/events can be obscured and their integrity compromised.

- While technology has the capacity to reduce the human workload at each phase of the assessment (test development, administration, data analysis and reporting, etc.), the project proposal only partially addresses the human "cost" of implementing a technology-centered system—managing, monitoring, trouble-shooting, are only a few of the activities that require labor for which there must be some further account.
- Many aspects of the electronic scoring proposed make good sense—among them, standard machine scoring of SR items, automated scoring of CR items, and the management of distributed scoring by teacher-raters. However, SBAC appears to underestimate all that is required to develop and fine-tune software for scoring performance events and extended constructed responses.
- The proposal does an excellent job of identifying possible and plausible directions in which technology may one day take assessment of student learning. However, some statements (e.g., "As advancements are made with automated scoring solutions during the life of this project, they will be integrated into the system") send up "red flags." It is insufficient to indicate that where technology solutions do not yet exist, humans will pick up the slack. Without detailing when, how, and who would do so, the proposal remains incomplete.
- Finally, there is much merit to establishing a Technology Implementation Task Force, as SBAC proposes to do, and also to collect and evaluate "lessons learned" from Consortium States currently administering (or even exploring the option to administer) online assessments. However, there is a significant gap between the kinds of items currently characterizing online assessment and those envisioned by SBAC. More than an examination of past and current experience is needed to identify risks and problems.

Based on all of these observations, the response to this section was scored in the "medium" range.

Suggestions to DOE:

A much wider net needs to be cast to determine "lessons learned" than that proposed by SBAC (who plan to gather this information from Consortium States), including but certainly not limited to the the uses of technology--in pilot projects, studies, pilot and operational administrations--by and for NAEP. It is likely that at present, some technical details of scoring engines are proprietary, but may be made available to the Department of Education.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	18
<p>(A)(8) Reviewer Comments:</p> <p>Given that SBAC decided to work with an Interim Project Management Partner and to conduct an open, competitive procedure for selecting a permanent Project Management Partner, the decision to issue an RFP prior to notification of the grant award and to allow for short-term extension of services by the Interim Project Management Partner makes sense as a way to find the best possible candidate without any "down time."</p> <p>The reputation of the Interim Project Management Partner is well-known and conforms to SBAC's description of the team as one that has "deep, current assessment knowledge and leadership experience in guiding states on the development, implementation, and evaluation of comprehensive, state-of-the-art assessment systems. Most of the personnel identified have stellar credentials, and the rest are at least satisfactory. Of some concern, should the Interim Project Management Partner be selected as the permanent Project Management Partner, is their lack of management experience with projects of this vast scope.</p> <p>According to the applicant's narrative, "the primary risk management strategy will be to create comprehensive work plans as soon as possible to ensure that sufficient time and resources are allocated to complete the work." This sounds reasonable, but for the fact that the issues and implications associated with the many intended uses of technology within the proposed system make it quite difficult to form accurate estimates of time and resources needed to accomplish their stated goals.</p>		

This section included some details that appear to run counter to information contained elsewhere in the proposal or are not otherwise reflected in contexts where they might be expected. Among these details are the following:

- Within the narrative is a statement that the Project Management Partner will work with participating States to "identify implementation barriers, risks, and possible solutions or mitigation strategies." It is not clear what relation this information has or will have to the information on barriers contained in the MOUs.
- Contingency planning, mentioned here in the context of allowance in the budget for a paper form in the event of adaptive computer system failure, appears no where else in the proposal, as one might expect it should. Furthermore, the Project Work Plan and Time Line (Summary Table for [A][8][b]) identifies windows for dependent tasks that are so compressed that it seems unlikely that contingency plans can be developed to address any unanticipated findings. For example, the timeframe for identifying technology available from vendors, developing and piloting an item authoring system based on technology available for delivering and scoring items, and then initiating item development (which must include formulating evaluative criteria regardless of the scoring methodology ultimately used) seems exceedingly tight.
- The Time Line of Major Milestones indicates that item development will be completed during 2011; given the fact that many item types are dependent upon emerging technology, this seems unrealistic and contrary to information presented elsewhere in the proposal.
- SBAC proposes using a Michigan-developed web-based item authoring and banking system for development of new items and archiving of existing ones, and assert that using an existing State system (and enhancing it for Consortium use) is a "significant cost-savings approach." It is not clear that any system currently exists that goes beyond item-authoring computer-scored items other than SR items; if that is truly the case, then it's unclear how much saving will result from using the Michigan system given the need for computer adaptive SRS, ECRS, and TEs, along with computer delivered and computer scored performance events
- The dependence upon teachers for a great many tasks (e.g., item development, scoring, outreach) calls into question the sustainability of this system over time. Experience has shown that among teachers who demonstrate interest in participating in assessment-related activities, the motivation to do so is often to "get an edge" over other teachers by gaining greater familiarity with standards and the ways they are embodied in test items, as well as gaining understanding about evaluative criteria. Over time, this motivation decreases, and fewer teachers are inclined to sign on, even when remunerated for their efforts. Given that the proposal does not always make clear that payment will be made to teachers for their involvement in particular activities, and also given the issues that some of the member States already have with both funding and contracts/collective bargaining agreements, the confidence expressed by the Consortium in this aspect of their budget is open to question.

These various concerns/inconsistencies detract from full confidence that the applicant's project management plan will result in implementation of the proposed assessment system on time, within budget, and in a manner that is financially sustainable over time. Therefore, the response was scored in the "medium" range.

Suggestion to DOE:

Given that this plan calls for the involvement of many vendor partners, it is advisable--perhaps necessary--to bring on board a partner who can serve as a quality assurance contractor to ensure that all parties are working well together, sharing resources in a non-proprietary way and avoiding redundancy.

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score

Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	18
<p>Competitive Reviewer Comments:</p> <p>The SBAC has obtained an impressive range of support from IHE's and IHE leadership organizations (State Higher Education Executive Officers, National Association of System Heads, American Council on Education). The degree of enthusiasm and commitment is illustrated as well by the fact that the Consortium has secured letters of commitment from over a dozen private IHE's among four Consortium States.</p> <p>With a requirement of at least 30 percent of direct matriculation students in public IHEs in member States, strength of commitment is considerable, with the total number in SBAC States accounting for 74% of the total number of direct matriculation students across all States in that Consortium.</p> <p>A small number of IHEs (see Missouri entries) qualified their support somewhat by noting that the LOI is a non-binding agreement, and that the decision whether to remain in the Consortium and/or use the assessments for college placement purposes remains up to the institution. It would have been helpful for the SBAC to address these infrequent but still meaningful reservations in the narrative for this section.</p>		

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes
<p>Absolute Reviewer Comments:</p> <p>Based on the terms of the Absolute Priority, which focus only on what will be measured, how, and towards what ends, without any mention of feasibility or cost-effectiveness, it appears that the applicant has generally demonstrated the capacity to develop and implement an assessment system that meets those terms. Its weakness lies primarily in the lack of recursiveness of processes for moving forward in a timely yet deliberate way that allows for alternatives based on what is learned along the way. Particularly given the tremendous degree of dependence on emerging technologies, this "next generation assessment" may need to be more iterative, with web-based applications and use of AI scoring expanding only over time and in a way that capitalizes on the reduction in costs that typically accompanies the shift from prototype to mass-market delivery. Were it a question of confidence that the applicant can and will develop the assessment system described herein, in the timeline and within the budget outlined, the verdict would be "no" instead.</p> <p><u>Suggestions to DOE:</u></p> <p>Should the decision be made, after deliberation, to award a grant to SBAC, it might be wise to establish some initial targets and proceed only if/when more evidence accrued that their plan is feasible and cost-effective; this would require their providing much more information about how they plan to engage and retain the support of teachers for the many roles they've been assigned in this plan, and how States anticipate sustaining the assessment system beyond the funding period.</p>		

Grand Total	220	152
--------------------	------------	------------

Budgets

Level 1 Budget		
-----------------------	--	--

Name: Level 1 Budget(s)

The governance budget appears realistic and appropriate for all categories.

Without more detail on the extent to which the Consortium wishes to pursue new and innovative technology-enhanced items, it is difficult to determine the feasibility of producing specifications and training materials within the proposed budget. Care must be taken that there is not a "fall back" to more conventional item types for which specifications and training may be far more easily prepared. Under this category and others, no training stipends are budgeted for professional development; this raises a "red flag" since in many school districts, teachers and other school and district personnel either receive a stipend for any training or professional development that is not part of the regular PD agenda, or money is allocated for substitutes to provide release time for teachers. It is questionable whether good will and curiosity alone will bring to these and other training activities the numbers needed to ensure staffing for the activities which training is leading towards. Furthermore, there is no detail to suggest if and at what rate teachers will be remunerated for item development.

The anticipated rates (hourly and daily) for consultants seems somewhat high; similarly challenging and important tasks requiring consultants with the same level of expertise are known to cost 10-15% less. Particularly given the amount of effort expected from school system and SEA personnel as part of their regular duties, it may be reasonable to expect outside consultants to accept somewhat more modest fees given the significance of the work they are called upon to do.

It is difficult to question anticipated costs for the assessments themselves, since the formulas/model used to determine those costs is based on proprietary information. Caution should be exercised, however, since the assessments will be charting at least some new territory in terms of design, delivery, and scoring for which there are relatively few precedents (and the details of which are themselves often proprietary).

The projected per-student operational cost of the summative and interim/benchmark assessments is difficult to confirm because while there certainly are many cost-saving features to the SBAC plan, development costs are not as clear as suggested. Scoring costs, in particular, may wind up far exceeding the figures given, based on the complexity of the sample performance events and even some of the ECRS included in the proposal. On one hand, AI scoring costs are shaped by the degree of similarity between/among items/events (to limit the need for more "retraining" of the scoring software), which will impact construct validity; on the other hand, where the richness and complexity of some performance events and even ECRs winds up requiring human raters, unanticipated costs can quickly mount up.

Projected fall-out rates for items and performance events are difficult to evaluate without knowing what arrangements will be made for remuneration. Typically, where writers are paid for each approved item, rates go up considerably, and are less promising under other conditions. The precise role of vendors in item development (aside from developing and conducting training) is not clear, making evaluation of this aspect of the budget difficult.

The plans for a paper "back-up" version of the assessment may have additional fiscal impact not addressed in this budget. For example, one might assume that studies will need to be done to ensure comparability of forms. The paper-and-pencil analogues to many of the technology-enhanced items are likely to require scoring by human raters or, at the very least, scanning of booklet-based responses in preparation for electronic scoring (which has its own associated costs).

It is difficult to ascertain the feasibility of developing AI scoring for the amount projected without a clearer idea of the scope and complexity of performance tasks and ECRs. The number of responses required to "train" the scoring system on each item/event seems low, based on the complexity and variety of performance task types. Again, since information used to calculate cost of live scoring is propriety, there is no assurance that they will remain within the proposed budget.

Level 2 Budgets

Name:

Name: Assessment Design - Translations

The rationale for translation is very general and does not support the request for nearly ten million dollars. Since some surveying of Consortium states has evidently already been done (resulting in the statement that 92% of ELs would be covered by offering Spanish and four other languages), it would have been appropriate to identify the most likely other languages identified.

This section contains contradictory information regarding how many languages would be addressed, since the **Associated Work Plan** indicates the intent to develop "five language translations, including sign language (e.g., a total of five), while priority mentions "Spanish and up to four other languages." The last paragraph under **Priority** clarifies that indeed, Spanish, three other languages, and a system of sign language to be determined are the focus.

The proposal does not make a sufficient case for why a translation into sign is needed or advisable for mathematics assessment, since hearing-impaired students for the most part will be able to read text and interact with graphic materials (whether hard copy or online).

Suggestion to DOE:

It seems advisable to obtain more information about the sign translation (the need for it and how it would increase the pool of tested students).



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b) (6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	9

(A)(1) Reviewer Comments:

Overall

SBAC's vision and MOU are strong insofar as the Consortium's proposed balanced assessment system is concerned but not as strong in regard to prioritizing and assuring innovation, in and the usefulness of, the system for accountability purposes and developing common mechanisms for those purposes. Its governance plan is solid in regard to making policy decisions but weaker in regard to facilitating operational success. The timelines and plans for the Consortium's making of key policy decisions and especially for States to acknowledge and remove barriers to participation in the new assessment system are not set forth with specificity, leaving a risk that key actions to set policy (e.g., common achievement standards) and remove barriers will occur too late to facilitate full implementation by some or most states by 2014-15. The approach to managing funds and procurement -- using the State of Washington's normal procedures -- is solid and can be improved in particular respects. Overall, the Consortium's proposed structure warrants a moderate degree of confidence that it will lead to the successful design, development and implementation of an assessment system that meets the goals of the Notice Inviting Applications (NIA).

Vision, goals, role, deliverables:

Evaluation: SBAC's vision, goals, role and deliverables encompass an innovative, balanced assessment system built around the Common Core State Standards (CCSS), in service of rigorous student learning objectives, and taking substantial advantage of advancements in the technology of assessments (including computer adaptive tests, computer-enhanced constructed response items, artificial intelligence scoring, and the use of performance events). SBAC's vision and goals, however, devotes virtually no attention to the NIA's aim of fostering the use of growth measures for accountability purposes, developing and implementing common measures of whether elementary and middle school children are on track to graduate prepare for college and careers, and assuring that achievement and growth data are useful and effective for evaluating schools, principals and teachers. Overall, SBAC intends to build and implement an impressive assessment system but has weak goals in regard to developing common uses of the system for accountability-related purposes. While making a strong case as to the need for formative tools, it does not clearly demonstrate the value of the particular package of formative materials it contemplates.

Explanations:

* SBAC's overarching goal is the one contemplated by the NIA: to "radically reshape the education enterprise in participating States in order to improve student outcomes"

* In its vision and application generally, SBAC focuses heavily on achieving a reasonable balance of summative, interim, benchmark, and formative assessments, all fully aligned with each other and with rigorous college- and career-ready standards. Balance is sought, as well, in what learning outcomes are measured (deep disciplinary knowledge as well as higher-order thinking skills), the types of mechanisms used to measure that learning (selected response, conventional constructed response, computer-enhanced

constructed response, and performance events), and the relative roles of vendors and technology on the one hand, and educators on the other hand, in item development and scoring. In these respects, the balance sought is a positive feature of the system, which can contribute to demonstrably improved student outcomes at a reasonable cost in time and money.

* The assessment system aims to make good use of advances in the methodology and technology of assessments to accomplish this vision. Computer-adaptive methods enable the system to test effectively off-grade and to speed test administration, scoring and reporting and lower the operational costs (albeit with initial increases in development and infrastructure costs) of summative and other forms of assessment. Also positive is the effort to use other advances in assessment methodology, including in item types and scoring.

* SBAC's principle of "flexible responsibility" is a source of some uncertainty. The Consortium clearly intends to provide a common assessment system and commits member States to use the system to administer and score the summative assessments that are part of that system. It also evidently intends to create a common standard for determining whether students at the point of graduation are college/career ready (CCR) and whether students in grades 3-8 and 10 are "on grade." There is no requirement that member States use the interim/benchmark or formative assessment features of the system -- a type of flexibility that makes good sense. Where the confusion comes in is in regard to whether member States agree to create (A) common growth measures; (B) common cut scores for determining whether elementary and middle school students are on track to graduate CCR; or (C) common measures of school, principal and teacher effectiveness. In the latter regards, the proposal's discussion of what the reports and what the system can do suggests that at least features (A) and (B) will be present. Indeed, the Reports section says directly that reports will provide comparable data by States, districts, schools and teachers (classes) on student growth and on the proportion of students who are on track to graduate CCR. And other sections discuss a timeline for setting common achievement standards. But in the Research and Evaluation section, the proposal states that the Consortium's intention is only to "be prepared to set standards for grades 3-8, using 'on-track to college- and career-readiness' as the vertical articulation criterion," and to create a common growth measure, but not actually to do so unless a single model is required for Title I accountability purposes or is "required by the funder" (i.e., by ED). Indeed, in the Research and Evaluation section, the proposal expresses doubts about the feasibility of growth models and about whether comparability among States is possible, and it does not clearly commit member States to development of their own accurate growth measures or to a mechanism through which the Consortium will determine or compare the accuracy of those measures. The Consortium thus evidently leaves it to each State to decide whether or how to use the system for purposes of reporting on growth and "on track to graduate CCR." The Consortium does promise that its research staff will try to determine whether comparability is possible and conduct research on growth measures. But the consequences for a state of developing measures that are not accurate or are not comparable to those used in the other Consortium states are not spelled out in the application or the MOU. Nor is there any stated mechanism for benchmarking each State's methodology against others, to create a "race to the top" that favors approaches that better contribute to improved student outcomes. The Consortium thus risks having a multitude of different and incompatible measures, some weaker than others, of sufficient growth and of measures of whether students are on track to graduate college- and career-ready.

* The content of SBAC's formative and professional development component is only vaguely stated. Evidently, actual formative assessments are not contemplated. Instead, materials supporting the development and use of such assessments and some training are contemplated. The content and value of such seemingly generic materials is not clear.

* SBAC's vision is compatible with its theory of action, both in terms of its ambitions in regard to the creation of a well-balanced, technologically advanced assessment system and in terms of the weakness of its goals as to common achievement, growth and accountability measures, and the lack of specificity in regard to the system's formative mechanisms and uses.

Organizational structure and the roles, rights and responsibilities of member states

Evaluation: The structure of the Consortium and the roles, rights and responsibilities of member States are clear. The extent of the responsibilities of the Consortium and paucity of responsibilities of member States, particularly with regard to the Work Groups and the efforts needed to complete the work create some risk that the Consortium will not successfully design, develop, and deploy the assessment system on time.

Explanation:

* SBAC provides a clear structure of differentiated state participation that follows the requirements of the NIA. Over half of the 31 States (17) have agreed to be Governing States, with a good geographic mix and a good mix of relatively more rural and urban states. Washington (WA) is clearly identified as having three roles – lead, governing, procurement State.

* Advisory States have no right to any ongoing role in decision making, although either the Steering Committee or Executive Committee may call for a vote of the full membership on issues that one of the committees identifies. The absence of any formal or regular occasions for advisory States to participate in decision making or informational meetings passes up a potentially valuable opportunity for building buy-in among these States.

* There is no statement of when or how often the Steering and Executive Committees are to meet, or whether members of the Executive Committee are expected to serve in that role full-time. The timeline provided in section (A)(8) refers to "SBAC Planning and Management Meetings", which are to occur twice annually. The "responsible agency" is identified as the Executive Committee. If this refers to a meeting of the full membership, it is responsive to the concern in the bullet just above. If, on the other hand, it is contemplated that the Executive Committee will meet only twice a year, that is not likely sufficient to provide the necessary oversight and decision making capacity.

* The responsibilities placed on Governing and Advisory States in the application itself, Apps A1-1 to 3 and in the MOU are thin, apart from responsibilities specifically required by the NIA. Governing States are subject to only two actual concrete requirements beyond those specified in the NIA: (1) contribute one representative to the Steering Committee who is either the State's Chief State School Officer or that person's designee, and who has prior experience in design or implementation of curriculum or assessments and agrees to be a "liaison between total state membership and working groups" (a duty that is never defined) and (2) provide at least one person to serve on one of the Consortium's seven Work Groups. There are other statements of governing States' responsibilities -- e.g., "have an active role in policy-making," "approve ... the Executive Committee Members," "participate in the final decision-making of changes in governance and specific design elements," "be willing to participate in the decision making process and, if a Governing State, final decision," "contribut[e] significantly to policy, logistical, and implementation discussions that are necessary to fully operationalize the assessment system," "participat[e] significantly in Consortium-wide activities" and be "fully committed to the application and ... support its implementation," but these requirements are never defined and do not clearly go beyond participation on the Steering Committee and the contribution of one member to at least one Work Group. Because there are 17 Governing States and seven Work Groups, it is possible that most Work Groups will have only two or three State-contributed members. It is not clear what the governance of the Work Groups will be, who will lead each group and be responsible for its deliverables, and what power, if any, that leader will have to assure that the contributions of the other members from other States are sufficient to enable the group to get its work done. Also, although the proposal assigns significant roles to actors within the States who are not on a Work Group (e.g., developing and reviewing test items) and implicitly will require other support from such individuals (e.g., in understanding the format State data systems), the application does not say who has the authority to assure that this crucial work gets done. In a project of this size and complexity, uncertainty as to the holders of such responsibilities can undermine timely and successful design, development and deployment. It is also unclear how the inevitably overlapping responsibilities and inconsistent decisions of different Work Groups will be resolved.

Decision making methods, process; operational protocols

Evaluation: The decision making and membership methods, process and protocols are well specified. But in the interest of increasing consensus, these approaches create a risk of decisional delay.

Explanation:

- * The membership (entrance and exit) and voting rules for the Consortium and the membership and voting rules for its Steering and Executive Committee are clear in both the application and the MOU.
- * Limiting the Executive Committee to a relatively small number of members elected by the Steering Committee on a rolling basis increases the decision making capacity of the former committee, while providing a capacity for representation of additional States in successive years. Assuring the ongoing representation of the Lead/Procurement State and having a representative from higher education are sensible, although it remains to be seen whether a single individual can represent the interests of the wide array of higher educational institutions that have agreed to be subject to the Consortium.
- * In the interest of providing checks and balances on decision making, the Consortium has opted for rules that could delay decision making by the Executive Committee. These include having two chairs with no method identified for resolving disputes between them, having an even number of members (increasing the chance of 4-4 deadlocks, in the event that consensus fails), and the consensus preference, which for the Executive Committee is a requirement of the support of 75% of the members. In regard to the latter two of these concerns, the application and appendices state that majority rules, subject to a requirement of reconsideration of decisions that don't meet the super-majority requirements of the consensus policy. Reconsideration may be frequent, given the 75% consensus rule; and deadlocks in the event of disagreements between the co-chairs or 4-4 votes of the membership are a real possibility. This concern is enhanced by the operational, as well as policymaking duties assigned to the Executive Committee (e.g., day-to-day operational issues escalated by the Project Management Partner (PMP)).
- * Overall, there is a lack of clarity about day-to-day operational decision making. Good management practice suggests that escalation to the Executive Committee should be a last resort, but the application appears to treat that escalation process as the primary way to assure that obstacles encountered in the day-to-day work are resolved. There is mention in the budget of a Consortium Executive Director, but the role and authority of this individual are not defined, he or she is not referenced on the Consortium organizational chart in the MOU, and he or she has no assigned staff. Neither this individual nor the PMP itself is given any specific authority in regard to the Work Groups or in regard to work occurring in the States. Although there are suggestions that a Policy Coordinator, Content Advisor, and Technical Advisory and Policy Advisory Committees can escalate issues to the Executive Committee, these individuals and groups seem to be focused on policy and expert advice, not operations, so they are not well-placed to play an operational management or oversight role. All of this suggests considerable uncertainty in governance with regard to day-to-day operational issues. Inevitably, in a project of this size, there will be disagreements, different assumptions, and dependencies for each work group and State created by the timelines and decisions of other groups and States. Resolving these kinds of issues requires someone, or a body, with the authority acknowledged by the member States and their contributed personnel to make day-to-day operational decisions. How this will occur within the SBAC Consortium is unclear.
- * There is overlap in the responsibilities of the Steering and Executive Committees, e.g., in regard to which committee can escalate issues for consideration by the full membership of the Consortium. And there is some lack of clarity in regard to the role and authority of each of those committees and the Lead Procurement State in regard to overseeing the expenditure of funds. (Both committees and the Lead Procurement State are given "oversight" authority in regard to the expenditure of funds "in collaboration with" each other.) Leaving some flexibility for these roles and relationships to evolve is sensible, but they should be monitored early on to see if that evolution is occurring smoothly, and whether additional clarity is required.

Plan and timeline for setting key policies

Evaluation: The proposal provides a summary timeline that encompasses the setting of the policies and definitions that are explicitly listed as examples in the selection criteria for (A)(1)(b)(v). No information about the plan or process for achieving each milestone is provided in this section apart from Summary Table (A)(1)(b)(v). Evaluation of this timeline is hampered by discrepancies between it and the timeline included in section (A)(8)(b) – different milestones, different start dates, and in some cases different end dates. Because all of the start dates are the same in the (A)(1)(b)(v) table, it is difficult to determine whether there

is any phasing of activity. Subject to these constraints on evaluating this aspect of the proposal, the progression of the decisions is reasonable, except for a concern about (1) the setting of test administration policy after the testing of items begins (making it unclear that the testing will be conducted under the relevant policies), and (2) the possible simultaneity, late date and incompleteness of the decisions as to common performance level descriptors and common achievement standards.

Explanation:

* The (A)(1)(b)(v) timeline lists the start date for "Common set of performance level descriptors" and "common set of achievement standards" as October 1, 2010 and the end date as August 31, 2014. The (A)(8)(b) timeline sets the start date for both decisions as August 31, 2014 and the end date as the same date, August 31, 2014. The description of this activity in the text of the proposal, suggests that the process for setting these descriptors and standards is to commence on August 1, 2014 with the convocation of representative stakeholders to make proposals as of August 31, 2014. However, according to the (A)(8)(b) timeline, the Total State Membership won't be asked to approve them until September 1, 2015. Although it is not clear, the proposal narrative also suggests in places that the exercise occurring from August 1, 2014 to September 1, 2015 will be focused on common standards for determining whether students set to graduate have attained a "college- and career-ready" (CCR) status, but will not include common standards or cut points for whether students in earlier grades are on track to graduate or common growth measures or clearly require member States to develop their own such measures – matters the Consortium evidently intends to leave to the decision of individual States. This timeline raises concerns:

(1) The setting of common performance level descriptors and common achievement standards seems to be back-loaded to the very end of the relevant time period, concluding at a point (September 1, 2015), which may be beyond the timeframe provided for by the NIA for common achievement standards (the 2014-15 school year). There are indications in some parts of the proposal narrative that these efforts will begin earlier but the timelines are not clear on this point.

(2) Common performance level descriptors are a building block for common achievement standards, so it would make sense to begin developing the former before the latter.

(3) The NIA's definition of common achievement standards encompasses common standards and cut points for grade 3-8 students being on track to graduate CCR, but it appears that the Consortium may intend to limit its common achievement standards to actual graduation CCR.

(4) The Consortium evidently does not contemplate common growth measures or a requirement that States develop growth measures that are comparable to each other or provide a mechanism for determining the comparability and accuracy of growth measures that are developed.

Managing Funds

Evaluation and Explanation: The specified process, based on Washington State's rules and procedures, seems sound subject to a concern already mentioned – that the Steering and Executive Committees and the Lead Procurement State have overlapping and undefined responsibilities in regard to managing funds.

Consistency with MOU

Evaluation: In general there is strong congruence between the MOU and the undertakings in the Consortium's application. The MOU does not very clearly acknowledge the Consortium's vision and theory of action as presented in the application or mention the goal of improved student outcomes. The MOU reframes the "balance" goal as one of "balanc[ing] concerns for innovative assessment with the need for a fiscally sustainable system that is feasible to implement."

Explanation:

* In signing the MOU, each state has "agree[d] to be bound by the statements and assurances made in the application." Governing states are "fully committed to the application and will support its implementation."

Removal of barriers

Evaluation and Explanation: The Consortium leaves barrier-removal to each member State. In the MOU, each State commits to "identify existing barriers . . . by noting the barrier and the plan to remove the barrier." The MOU invites, but does not require, States to note on a template provided in the MOU the barriers each anticipates and the plan and timeline for removing the barriers. Most States (16 of 31, including 9 of 17 Governing States) did not fill out the template. The barrier removal process thus remains unknown for some a significant number of States, creating substantial risks in regard to effective implementation.

Procurement process

Evaluation and Explanation: Using WA's usual procurement process appears to be a sound approach, to which the other States have agreed. Although WA law allows both upfront distribution and pay-out based on deliverables, WA practice is to use only the latter. Given the magnitude of the effort required of vendors by this proposals, the speed with which they must complete their work, and the gravity of the dependencies from other vendors (e.g., the assessment system depends for success on the technology system and vice versa), and from the States (which are assigned significant operational tasks), finding qualified vendors that are willing to develop the contemplated components within the tight budgets that are proposed may be difficult without some upfront distribution.

Recommendations:

* Encourage the Consortium either (1) to adopt common growth measures and achievement standards/cut scores that apply to whether students, prior to their terminal year, are on track to graduate, or (2) to adopt a system for comparing and benchmarking different standards and measures adopted by different States to assure that such measures are developed and are accurate and comparable.

* Obtain more clarity as to (1) the membership, rights, responsibilities, and leadership of the Work Groups (including whether there are sufficient resources on the Work Groups to get the work done), (2) the responsibilities of the member States or of others for assuring that work assigned to, or required of, actors within each State will be completed in a timely fashion, and (3) the mechanism for coordinating overlapping activities of the Work Groups. This clarity could be provided up-front, or by carefully monitoring the operational activity of the Consortium early on.

* Obtain more clarity about the responsibility for, and authority to make, day-to-day operational decisions. This probably requires additional administrative structure between the Executive Committee and Work Groups. For example a coordinating committee might be established that is led by an executive director, has representatives from the Work Groups, and has responsibility and authority, acknowledged by the member States, for assuring that work assigned to the Work Groups and the States gets done at quality and on time.

* Set an early milestone (and perhaps later milestones at increasing intervals) to review whether the consensus-oriented policies (e.g., dual Executive Committee chairs, an even number of members of that committee, the consensus preference, and the overlap in responsibilities between the Steering and Executive Committee and the Lead Procurement State) are working well or are delaying action and require reconsideration.

* Clarify the time line for starting, phasing and finalizing performance-level descriptors and common achievement standards.

* Clarify and solidify the Consortium's intentions in regard to setting a common growth measure and common cut points and achievement standards for whether students in grades 3-8 are on track to graduate or, alternatively, creating clear benchmarking mechanisms through which growth measures and cut points and standards set by different States can be compared and incentives created for States with weaker standards to adopt stronger ones.

* Create a process through which member States inform the Consortium of their barrier-removal plans, share effective practices with each other, and provide updates on those plans.

* If state law permits, consider retaining the flexibility to use upfront payments to mitigate risk to vendors, while using incentive payments for on-time, at-quality performance, or penalties for late or low-quality performance, in order to mitigate the risk to WA and the grant from upfront payments.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	2
<p>(A)(2) Reviewer Comments:</p> <p>Evaluation and Explanation:</p> <p>SBAC's theory of action consists of a further explication of its impressive vision, described in the comments to (A)(1), of a coherent, balanced and aligned system of rigorous standards and assessments that seeks to get the most out of each of several different types of assessments, assessment item types, and scoring practices, while also taking full advantage of technology. There is a well-developed theory of how the assessment system will work to generate useful summative and formative information.</p> <p>What is lacking is a clear statement of how the Consortium expects the new system to change and enhance the ways in which educators work in schools to improve student outcomes. SBAC places almost no emphasis on the accountability aspects of the system it is developing as a mechanism to drive those kinds of improvements. The Consortium gives some attention to the value of aligning the assessment system with curriculum, professional development and a variety of formative practices in schools, but doesn't provide a very convincing discussion of what formative processes are envisioned and how those processes are expected to work to improve instruction and student outcomes.</p> <p>There are other respects, as well, in which the theory of action could promote more effective use of assessment results to improve student achievement and college- and career-readiness. For example, the theory of action places little emphasis on efforts to determine whether the initial balance of the various types and methods of assessments, items and scoring methodology could be improved over time by increasing the uses of measures that are relatively more predictive of favorable student outcomes, and by measuring the comparative return on investment on the different types of assessments and item types, given their varying costs and outcomes. Steps in this direction are, however, contemplated in the Consortium's research and evaluation plan.</p> <p>Also, the theory of action recognizes the need to ensure that all students have access to the equipment and technology needed for all of aspects of the SBAC assessment system, but its proposal takes no concerted steps to determine whether the equipment and infrastructure available to schools in the Consortium States are up to that task and to help those States that are further behind in this regard to develop strategies for remedying the situation.</p>		

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	39
<p>(A)(3) Reviewer Comments:</p> <p><u>Assessment System</u></p> <p>Consistently with its theory of action, SBAC's assessment system is highly innovative in its:</p>		

- * adoption of state-of the art learning standards in ELA and math;

- * extension of computer-adaptive assessment techniques to summative (once annually) uses as well as interim predictions of student attainment and progress and targeting of student needs in particular content areas (both of which would be flexibly available throughout the year), allowing it to test students across the full range of achievement status, including well above and below grade level and to provide students, educators and parents with nearly instantaneous results for computer-scored items (and two-week turn-around for teacher-scored items);

- * rich mix of assessment items and methods and use of newly developing test item types such as computer-enhanced constructed-response items (the example of which provided in the proposal are encouraging) and performance events to measure effectively both deep disciplinary knowledge as well as hard-to-measure higher-order skills and to provide a mechanism for assessing ELA and math skills in the context of science or social studies topics; use of artificial intelligence technology and mixes of teacher and electronic analysis for scoring purposes; and use of technology to permit teachers to construct and share new test items and other kinds of innovative practices including in regard to curriculum and formative practices;

- * flexibility granted member States to create their own mixtures of no-stakes predictive, targeted, and formative assessment, all aligned to each other and to CCSS standards, while holding firm to a single system of high-quality summative assessment;

- * commitment to research whether summative assessments can be distributed across the school year without hampering the rigor of the assessments and the comparability of results; and

- *the integrative Access by Design methodology and careful up-front research the Consortium proposes to enable it to design and build assessments and delivery techniques that are accessible from the outset to special needs populations, diminishing the need for disruptive and costly accommodations. (The discussion in Appendix A3-2 of the intent to use Access by Design principles is impressive, consistent with the generally high quality of the proposal's treatment of accommodations for special populations.)

SBAC's planned research will support ongoing innovation by revealing the most effective and efficient mixtures of item types and scoring techniques that it aims to "balance," such as the mix of computer-enhanced constructed response items and performance events, and of teacher and electronic scoring.

In the technology area, the SBAC system does not push the envelope as much, but aspires (although it is not currently adequately budgeted or described) to take advantage of the best existing components for:

- * delivering summative, interim and predictive assessments;

- * reporting, graphically displaying, revealing standard-error bands for, and aggregating and disaggregating assessment results at various levels of detail (using both fixed and customizable features; providing data at the standard, strand, and item level; comparing States, districts, schools, and teachers; disaggregating by key groups, and focusing on multiple audiences, including parents and teachers as well as administrators);

- * enabling teachers to contribute to and customize assessment content for interim, predictive and targeted purposes; and

- * facilitating teacher collaboration and social networking around assessment and possibly curriculum and formative practices.

Innovation often, of course, trades off against feasibility, but where SBAC pushed the envelope the most -- for example, in the development of computer-enhanced items and performance events and in the creation of composite scores out of of computer-adaptive and performance event outcomes -- it sensibly intends to proceed with some caution, using its balance of components to give it flexibility to tip one way or the other if the new mechanisms turn out to be more difficult or easier to implement than predicted. The modest use of performance events and of a mixture of electronic (AI) and teacher scoring is an example of this admirable approach to innovation. In these regards, SBAC is effectively accomplishing its goal of balancing innovation and feasibility.

One major concern that arises, however, is in regard to the new system's capacity to measure student achievement for especially high- and low-performing students. Computer-adaptive tests (CAT) create the possibility of doing so but require a very large number of items to be able to (1) identify all students in all grades who are performing well above- or below-grade, (2) determine the approximate performance level of each such student, and (3) then assess the student at that level comparably to how other students are measured at their actual grade level. Although CAT allow students in one grade to be measured using items that typically are appropriate for students at other ages, this isn't possible with many items. For example, a student who is in seventh grade and is socially adapted to that grade may not be willing to seriously address items prepared for a third grader, even if that is the cognitive level at which the seventh grader is performing. (Likewise, a third-grader performing above grade level may not have the social context, even if he or she has the cognitive capacity, to answer an item prepared for children in older grades.) Also, items have to be created for students functioning below the third grade and above the tenth grade, even though those are the beginning and ending grades in which summative assessments are conducted. SBAC does not provide any information in its proposal that specifically addresses the number of items needed to measure very high- and low- performing students in all grades (including in the proportion at which such students appear in the student bodies of the respective member States) and that demonstrates that the requisite number of items will be provided.

There also is a feasibility question triggered by SBAC's fundamental commitment to computer-adaptive technology and to using the same platform for all summative and no-stakes assessment except in rare cases where a particular student's special needs require paper and pencil administration. In this instance, the Consortium is not proposing a balance -- here of computer and paper-and-pencil administration -- and instead is opting almost entirely for the former. The proposal, however, does not clearly document, nor provide a process for determining, the feasibility of an entirely computerized system in its various member States, given the demands such a system places on equipment and infrastructure. Steps to develop more clarity on this point are recommended.

Another major concern is whether artificial intelligence (AI) scoring can bear the weight placed on it by the proposal. There are insufficient steps taken to determine early on in the development process whether AI scoring can accomplish all that is hoped for it and, if not, to make back-up arrangements.

An additional concern has to do with reporting. The general statements in regard to the goals for this aspect of the system are strong but very little detail is provided, and the examples given in the appendix are not consistently impressive. They do not, that is, provide information in a simple, intelligible, efficient fashion.

Accountability

SBAC is not as disposed to innovate in regard to developing and implementing common accountability mechanisms that use the results of summative assessment (1) to create strong incentives to improve student outcomes, while (2) providing educators with the capacity to do so through examination of the diagnostic information embedded in student summative outcomes. For example, there is a limited evidence of innovation in regard to:

- * providing educators with ways of measuring, and incentives to measure, whether each child is moving forward sufficiently each year towards the target the system has set (in this case graduation college- and career ready); and

- * providing students, parents, educators and the public with transparent and comparable mechanisms for determining how effectively each State, district, school and (ultimately) educator has accomplished that goal across the range of students for whom they are responsible.

For the reasons discussed in regard to the Consortium's vision (see (A)(1) comments), the Consortium evidently does not intend to set common growth measures or common measures of whether students are on track to graduate college- and career-ready. Nor, alternatively, does SBAC identify steps it will take to assure the accuracy of the growth measures individual States adopt or assure the measures' effectiveness in using assessment outcomes to drive improvements in student learning. This could be done, for example, by benchmarking and comparing member States' different standards for assessing student growth, "on track" status, and other evaluative criteria in order to create a "race to the top" among States based, for

example, on which of their competing methods are associated with higher student achievement. The Consortium does intend to generate data, and to conduct research on these topics that can be used for these purposes, if required for Title I accountability purposes.

Unlike in regard to assessment design, therefore, where SBAC is committed to starting with the best available techniques and technologies and carefully innovating, in the area of using assessment results for accountability purposes, the Consortium has adopted a more static approach that forgoes innovation, implementation and continuous improvement in favor of a laboratory-focused search for a more perfect model.

A related concern is the underdeveloped process that is described for devising performance descriptors and common achievement standards.

At the level of detail, a couple of concerns arise in regard to test security or, at least, the the public appearance that security is being maintained. Giving States access to the secure item pool for developing high school cluster/domain assessments or EOCs and allowing students to retake the summative assessments a second time each year both raise test security concerns that deserve attention.

Recommendation:

Integrate the Access by Design approach into the Consortium's theory of action.

Encourage additional innovation, commitment and monitoring in regard to growth measures, measures of whether students are on track to graduate college- and career ready, and measures of school, principal and teacher effectiveness.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	20

(A)(4) Reviewer Comments:

Subject to concerns expressed elsewhere about governance and management, the substance of SBAC's plans for developing assessment items and accommodations provides a reasonable degree of confidence that the assessment system will be ready for wide-scale and timely administration. There is less clarity and less of a basis for confidence in regard to assessment scoring, technology/reporting, quality control and field-testing. The proposal does not include an explicit plan or process or adequate focus on obtaining ongoing feedback to facilitate continuous improvement.

The approach to designing summative and interim assessment items is thoughtful and convincing in substance, particularly the deliberate way the system is built up from assessment frameworks tied to progressions of learning, to test blueprints, to learning-based item templates, to the items themselves. This goes especially for the concept of reusable event design templates or shells for selected- and constructed-response items and performance events, which are designed so that, if teachers "teach to the template," the result will be authentic learning. The role of performance events is also important from the perspective of measuring higher-order skills, although the proposal may undersell the extent to which the computer-enhanced constructed-response items it intends to develop can accomplish the same or similar goals less expensively. Also positive are: the effort to identify and make use of items already in existence in the States (contributing a significant minority of the overall items); the combination of inputs from experts and higher-education representatives gathered centrally and teachers working regionally and in States at the design, writing, and review stages; and the back-up review of one quarter of the items by an independent contractor. However, because this overall development approach, by design, has many moving parts across the 17 Governing States, the governance and management tasks are demanding to say the least, contributing to the concerns expressed in the (A)(1) and (A)(8) comments.

As set forth in Appendix A4-3, the accommodations development process is impressive.

The scoring development process also seems generally reasonable in substance. Particularly praiseworthy is the Consortium's commitment that teachers will not be allowed to score the work of students from their own States, which will lend credibility to outcomes. Scoring is perhaps the area in which the Consortium is pushing the envelope the most, given (1) the need to "train" the automated scoring system for purpose of scoring and providing partial credit for technology-enhanced items in ways that are novel to the K-12 domain; (2) the similarly (or even more) emergent nature of the Artificial Intelligence methods to be used, in part, in scoring performance events and verbal responses, and (3) the uncertainty of the proposed mixture of teacher and electronic scoring for constructed response items and performance events. As a result, the management risk here especially high. The scoring process is not slated to begin until field testing has been completed, which evidently means sometime after spring 2014. (There is some confusion in the proposal as to exactly when field testing occurs, and whether the commencement of scoring after field testing means after the initial field testing of items in spring 2013 or full field testing in spring 2014). The proposal does not make use of the pilots discussed in the field testing section as a way to push up the time when scoring development efforts can begin to be piloted. A weakness of the proposal, therefore, is how late in the process efforts begin to determine whether AI scoring can accomplish all that the Consortium wants it to accomplish and, if not, to develop a back-up approach.

Consistent with other parts of the proposal, the discussion of the technology-related facets of the development process (in this case the reporting function) is not well-developed. Instead of saying how the reports will be developed, this section states again, with a bit more detail, what information the reports will provide, and to whom. The exemplars in Appendix A-4a are too divergent in their content and quality to provide much guidance, and overall the level of quality, simplicity and usability is not high. The narrative here, as elsewhere, focuses more heavily on achievement reporting than on growth reporting and temporizes a bit, compared to the descriptions in part (A)(3) of the proposal, on the extent to which comparable information will be provided across States ("Base reporting of summative assessment results will be common across States for comparison purposes, but schools, districts, and States will have [flexibility] . . .").

The description of the quality control processes is also scanty, especially for the technology system and reporting. Reference is made to "developing interconnected systems for each phase of development" and "a monitoring plan". The former depends on the quality of the management and coordination structure, about which concerns are expressed elsewhere. The latter is undefined.

The field testing process again raises management concerns given the involvement and need for coordination of all member States. Together with item development and scoring, the field testing process underlines the need to identify and empower project managers in each State and build them into the governance and management systems.

Except in the area of accommodations for special populations, there is little discussion in any part of the SBAC proposal about how the Consortium intends to leverage user comments, difficulties and complaints in regard to the operation of any aspect of the system as a basis for continuous improvement. Most or all of the user acceptance testing and much of the quality control will be conducted by third-party vendors, not by actual users. Full field testing comes so late in the process (Spring 2014) that it will mainly be useful for bug -fixing, not for redesign purposes. Although the Research and Evaluation unit will be conducting reports on cost and usefulness, there is no clear feedback loop identified for assuring that this information, as well, is used as a basis for continuous improvement.

Recommendations:

- * Identify and empower SBAC project managers in each State (or at least each Governing State) and build them into the governance and management system.
- * Undertake efforts to commence pilot or simulated scoring efforts at the earliest possible point, in order to flag difficulties and permit the timely and effective development of the backup plan that is suggested. It is noted that small-scale pilot testing and convenience samples will be conducted (see the field testing section). These resources might be used or expanded as a way to allow earlier piloting of scoring methods.

Because human scoring is a crucial foundation for knowing what to "train" the artificial intelligence (AI) scoring system to do, it is important that both human and AI scoring efforts and pilots begin as early as possible.

* Develop a plan to use simulations (e.g., screen shots), prototypes, pilots, capture of field-testing and help-desk feedback, and information in Research and Evaluation Reports as mechanisms for routine, user-oriented continuous improvement.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	21

(A)(5) Reviewer Comments:

The Research and Evaluation (R&E) team has been assigned what amount to both a research/evaluation role and a role in determining policy with regard to the viability and content of a potential growth model and of achievement standards for determining whether students are on track to graduate college and career ready (CCR) and whether it is possible to compare results among States on these measures. The proposal narrative in section (A)(5) lays out an ambitious scope of work on both counts -- so ambitious that it is unlikely that both sets of work can be accomplished within the \$5 million budgeted for it. One reason for this large amount of work is the Consortium's approach of treating the development of growth measures and "on-track" standards as research questions without operative outcomes -- unlike, for example, the development of performance-level descriptors and common achievement standards for determining whether students are college- and career-ready.

The R&E work cross cuts all of the other functions of the Consortium and requires data on every student, teacher, class, school, principal, district and State in the Consortium, some of which data may not exist at the State level and have to be collected from individual districts and schools. Table A5-2 lays out the data the R&E team will need to carry out the planned studies, but provides no plan for obtaining that data in a timely manner. Nor do the States commit in their MOUs or elsewhere in the proposal to obtain and provide that data in a timely fashion.

As ambitious as the research and evaluation plan is, it is missing studies to reveal whether students who appeared to be "on track to graduate college- and career-ready" (CCR) and students deemed to be CCR at the time of graduation actually did succeed in college or in other careers. The plan, that is, contains no full-scale commitment to study whether these students upon entering college were able to succeed in credit-bearing courses, finished their first year successfully, and graduated on time.

If (as is certainly feasible) SBAC responds to important concerns about (1) the excessive workload of the R&E group, (2) the management and data-wrangling issues just mentioned, (3) the absence of any standards for monitoring, comparing and benchmarking state approaches to growth and "on track" standards (assuming the Consortium does not develop common measures), (4) the Consortium's stated doubts in this section about whether it is possible to develop comparable measures of growth and "on track" across different States, and (5) the absence of studies of whether students deemed "on track" and CCR actually do succeed thereafter in college and other careers, the Consortium can achieve its aspiration to ensure that the assessments developed are valid, reliable, effective and fair for their intended purposes and for all student subgroups. If not, there will be aspects of the system that are not effective for their intended purposes.

There are a number of particularly commendable aspects of SBAC's R&E plan, including its intention to:

* conduct what amount to "return on investment" (ROI) studies of the various assessment types (e.g., technology-enhanced constructed response items vs. performance events);

- * determine whether the computer-adaptive assessments live up to their promise to expose each student to items that measure the full range of the CCSS standards in a manner tailored to the students' current achievement level and grade-level expectations;
- * determine whether computer-adaptive and performance event scores can be combined into reliable and comparable composite scores;
- * determine whether the Consortium's achievement standards are, in fact, predictive of college and career success;
- * explore the capacity to conduct through-course summative assessments while still maintaining reliability and comparability of results;
- * assure the reliability of automated and artificial intelligence scoring; and
- * determine whether the predictive component of the I/B assessments actually predict student progress towards post-secondary readiness.

Recommendation:

- * Remove the policy-making activities from the R&E team and assign to the relevant policy-making group, particularly with regard to growth measures and "on track" standards. This comment aligns with ones elsewhere in these comments focused on encouraging SBAC to come up with common growth measures on "on track" standards. This would also bring the R&E group's workload into better balance with its budget.
- * Add studies of whether students deemed to be "on track" to graduate CCR and those who did graduate CCR actually succeed in college and in other careers.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	4
(A)(6) Reviewer Comments:		
<p>The proposal only thinly describes its plan for supporting teachers and administrators in implementing the assessment system and for developing the professional capacity to use the assessments and results to inform and improve instruction. This calls into question the extent to which the Consortium's plan for implementing the proposed assessment system is feasible, cost-effective, and consistent with its theory of action.</p> <p>As the Consortium's theory of action notes, a key value of the summative, interim and formative data the new system will generate is the use of that information by educators to diagnose and cure instructional failure in regard to particular students, topics and classrooms. Without intensive, embedded, results-driven training for educators in this endeavor, much of the value of the system will go unrealized. The Consortium's proposal says relatively little, however, on the subject of training educators to use the data the new system will generate to improve instructions. The proposal points out that the Consortium cannot itself conduct professional development for all of the relevant educators across the many member States, and it says that the Consortium will "collaborate with existing professional development networks in each participating State to ensure that capacity building penetrates the State, regional, and local levels." No plan for, or even a general description of an approach to, doing this is provided, however. Although the budget narrative refers to something called the SCASS professional development model and a plan to "fund two SCASS groups at \$15,000 each for 15 States for two year," there is no mention or description of this plan anywhere else in the proposal, including in the instant section on professional capacity building. Nor does this approach appear to be comprehensive enough to live up to the Consortium's theory of action.</p>		

The proposal then states that the Consortium will develop a number of "materials and tools on which local professional development efforts can build." Very few of the materials described, however, have much to do with training educators in the use of the new system or in the use of assessment and other data to improve instruction. Instead, the materials focus on more tangential issues, and the materials seemingly will have to be developed at such a high level of generality, given different approaches among the States, that their value is unclear and their cost-effectiveness (especially compared to other uses of the same resources) is doubtful. Examples of this are "guidelines for effective formative assessment practices," "research supported exemplars of curricular/instructional materials to support teachers' professional growth," "recommended readings," "focused group discussions" on unstated topics, and "the development of formative tools, processes and practices." The limited apparent value of these materials suggests that some of the budget allocated for them could be reallocated to under-funded components of the SBAC plan, particularly the technology system.

Missing from the Consortium's plan is either (1) a common template that member States commit to use to conduct effective training in the use of assessment data to cure instructional failure (e.g., based on the inquiry team model developed in the writings of Mike Schmoker, such as *Results Now*), (2) a mechanism for transparently monitoring, comparing and benchmarking different professional development approaches used by member States, and (3) a clearinghouse and networking function within the Consortium's new technology system to facilitate cross-State collaboration in developing professional development and data training/inquiry team approaches.

In this regard, a minority of materials that are currently proposed do appear to have value that relates directly to the Consortium's theory of action and do support effective use of assessment results. Among these materials are:

- * descriptions of the components of the new assessment system;
- * materials to foster educator use of the functionality in the Consortium's technology platform through which educators share exemplary formative and other resources, rate them, and collaborate in their use;
- * explanations of the resources provided by the new assessment system for teachers to build and contribute their own assessment items;
- * common frameworks for conducting formative assessments and common templates for recording formative assessment results in a manner that can be absorbed into the Consortium data system; and
- * common methods for scoring the constructed response and performance event items in the new assessment system.

The strategy for informing the public and key stakeholders about the assessment system and for building support for the system from the public and those stakeholders is not effectively developed in the proposal. The transition to a system of this ambition will require substantial efforts to assure that tens of thousands of teachers and millions of parents understand, and appreciate the value of, the new system. The proposal does not present a plan capable of achieving that degree of understanding.

Recommendations:

- * Develop (1) a common template that member States commit to use to conduct effective training in the use of assessment data to cure instructional failure (e.g., based on the inquiry team model developed in the writings of Mike Schmoker, such as *Results Now*), (2) a mechanism for transparently monitoring, comparing and benchmarking different professional development approaches used by member States, and/or (3) a clearinghouse and networking function within the Consortium's new technology system to facilitate cross-State collaboration in developing professional development and data training/inquiry team approaches.
- * Reallocate funding from generic and tangential materials to underfunded areas of the SBAC proposal, e.g., the technology system.

* Provide more detail in regard to the Consortium's public outreach strategy for explaining to parents how resources devoted to the new assessment system will generate better student learning and outcomes.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	4

(A)(7) Reviewer Comments:

SBAC intends to build a new platform for developing, delivering, scoring, reporting on, contributing and accessing instructional and other resources relating to, and offering tools for educators to collaborate in building and using summative, interim, benchmark, and formative assessments. The proposal gives an impressive list of individual features that it is hoped the system will include. If those features were developed, the Consortium could use technology effectively to improve the quality, accessibility, cost-effectiveness, and efficiency of the proposed assessment system.

With only one exception, however, the proposal provides no account of the types of technology to be used (including whether the technology is existing and commercially-available or is being newly developed), how other States or organizations can re-use in a cost effective manner any technology platforms and technology components developed under this grant; or (as is noted elsewhere) how technology-related implementation or deployment barriers will be addressed (e.g., issues relating to local access to Internet-based assessments). The one exception is a Michigan-developed web-based item authoring and item banking system that represents less than 8% of the overall budget for the technology component.

Also omitted from the application is any discussion of the difficult and potentially expensive process through which States provide the data on students, teachers, classrooms, schools, and districts that the system will require in order to conform to the Consortium's theory of action. There is no where in the application or MOU that States commit to providing this data or to a process for obtaining the data.

Systems that provide a good bit, but not all, of the functionality the Consortium desires do exist, both commercially and on a custom-built basis, but the cost of implementing them is likely to be significantly higher than the \$27 million budgeted for this component. Because the proposal does not identify systems in place in Consortium States, or school districts within them, or other States or districts in the US, that provide a model for what is planned, or that the particular States or districts would be willing to share with the Consortium at reduced cost, there is no basis in the proposal to believe that there are sufficient resources devoted to this component.

Although much of the promise of the Consortium's theory of action is built around the use of computer-adaptive, artificial intelligence and other technology effectively to improve the quality, accessibility, cost-effectiveness, and efficiency of the proposed assessment system, the description given of the technology platform the Consortium intends to build around that computer-adaptive feature does not yet fully live up to this aspiration.

Recommendations:

* Identify the planned architecture of the new system, the high-level requirements for building that system, and a cost model demonstrating the system's feasibility within the projected budget.

* Consider whether costs could be contained by using or customizing a State's or LEA's existing system or by making common cause with another RTTT-A consortium, assuming this and another consortium are both funded. Unlike in many other areas of the NIA's goals, where the benefits of competitive development of different systems is large, developing competing systems may be an unnecessary expense. If, instead, efforts were undertaken to use an existing model, or build a single new system together with

another consortium or consortia (assuming this one and others are funded). There very likely would be significant economies of scale and network economies.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	7

(A)(8) Reviewer Comments:

Overall: Compared to the risks posed by the proposal, the level and quality of management support contemplated is weak. The interim Project Management Partner (PMP) does not appear to be qualified to manage the project, and the RFP for the permanent PMP does not provide sufficient confidence that the permanent PMP will have the necessary qualifications and experience. The project planning information and timelines provided are thin and incomplete in many respects, as is the budget documentation (except for the Assessment Development component), and as also is the information about States' plans for funding the system going forward. Although poorly documented, the technology component appears to be seriously under-funded, and the rationale provided for some of the expenditures in the professional-capacity component does not justify significant amounts of the budget devoted to that component.

Project Manager Partner

Evaluation: SBAC does not have a Project Manager Partner (PMP). It is implementing an RFP to select a PMP. SBAC has an Interim PMP, which served as its grant manager, and the consortium contemplates the possibility that its Interim PMP may (or may not) be selected as PMP. (SBAC has an option to continue its Interim PMP for three months, to provide a bridge to a new PMP.) SBAC provides the set of information about its Interim PMP that the selection criteria request in regard to the permanent PMP, and SBAC's RFP for a permanent PMP is appended to its application. A review of this information reveals that SBAC's Interim PMP has impressive substantive expertise across all but one of the components of the contemplated assessment system. The SBAC's Interim PMP, however, lacks the necessary management and implementation experience and capacity needed to provide reasonable confidence that the proposed assessment system will be implemented on time and within budget. The specifications in SBAC's RFP for a permanent PMP are imbalanced in this same regard, creating the possibility that the Consortium will end up with a PMP that likewise has the required substantive expertise, but lacks the necessary management and implementation experience and qualifications. The proposal to devote only part of each project manager's work day or week to this project is ill-advised.

Explanation:

* SBAC's proposed system is an impressive combination of roughly a dozen assessment, technological, measurement and evaluation, accommodation, professional development, and other components. Developing any one of the dozen or so components to a point ready for full implementation would be a substantial project, requiring top-notch management. But developing all of them to that point, doing so all at once as part of a single, integrated system, and doing so on an expedited time frame, requires superior management skills and experience, particularly given SBAC's stated intention to divide close to \$150 million worth of work among numerous vendors, seven major work groups, a number of advisory committees, dozens of consultants, and large teams of teachers in as many as 17 or even 31 States. Managing the work of all of these actors is a major task. Integrating all of their products into a single seamless system delivered on-time and on-budget is an even more demanding task. For this project to succeed, these tasks must be managed by individuals or an entity with a proven record of management, integration and implementation success in projects of something approaching this magnitude.

* A program of this ambition also requires substantive design expertise in the K-12 education domain. SBAC's Interim PMP has this latter, substantive, domain-specific expertise, across nearly all of the relevant

system components. The only relevant areas in which the partner lacks demonstrated experience and expertise are in the design of complex technology platforms and, possibly, the use of Artificial Intelligence for scoring purposes and the design of Performance Events. Compensating for this lack of substantive expertise on the technology side is a planned expenditure of \$2 million for a systems architecture and integrator contractor.

* The Interim PMP is not a management, systems integration, or implementation expert. Rather, this partner describes itself in materials in the application, as a "preeminent educational research, development, and service organization" and as having expertise in "Assessment and Accountability; Mathematics and Science; Curriculum and Instruction; Literacy; Culture, Diversity, a Secondary and Postsecondary Education; Evaluation; Special Education; Leadership and Teacher Professional Development; Healthy Kids, Schools, and Communities; Early Childhood; Web, Database, and Interactive Services; and Policy." Likewise, the five employees who managed the grant-development process and are featured in the proposal, have no professional training as project managers and at best only modest recent experience managing large time-pressured projects, none of which is at anything close to the scale of this project. For the most part, the entity and the five employees have worked as highly skilled test and item designers and creators, researchers, evaluators, policy advisors, and providers of technical assistance and data analysis.

* As is noted in other parts of these comments, SBAC's overall proposal, which the Interim PMP helped to develop, mirrors the partner's substantive strengths and operational inexperience.

* The requirements and qualifications set out in SBAC's RFP for a PMP seem to invite the same collection of strengths and weaknesses. SBAC states that it intends to use the same selection criteria for the PMP that it used in procuring the services of the current Interim PMP to support the development of its proposal. The RFP requires that the PMP have considerable substantive expertise ("Five (5) years of demonstrated knowledge and experience with education and the field of large scale assessment/test measurement") but requires less extensive and specific managerial, integrative and implementation experience, none of which needs to be at the scale of SBAC's proposed project ("Demonstrated project planning and project management skills on a large scale project that incorporates resources spread out over large distances").

* The application proposes that the relevant management personnel devote 50% (in several cases) and 25-30% (in other cases) of their time to this management effort. A project of this size demands a full-time management team. Individuals with competing demands on their time are unlikely to succeed.

* As a risk mitigation strategy, SBAC proposes to hire the PMP for only a year, with an option to extend the contract for three additional years. Changing management partners in the middle of a project like this is disruptive and creates significant risks of its own. This strategy is reasonable in extreme circumstances but is not a viable alternative to procuring a highly qualified management team.

Work plan and timeline

Evaluation: The Summary Table for (A)(8)(b) and supporting materials provide only modest information beyond a timeline of start and end dates. The information is inconsistent in places with the timeline in Summary Table (A)(1)(b)(v), omits some important milestones, and provides for important tasks to occur very late in, or even beyond, the timeframe contemplated by the NIA. There is insufficient information about the entities responsible for executing on the particular items. The risk-mitigation strategies are reasonable in themselves but inadequate to the overall level of risk.

Explanation:

* Exemplifying discrepancies between the two timetables, the (A)(1)(b)(v) timeline has the "common achievement standards" work stream commencing on October 1, 2010, and identifies August 31, 2014 as the "date to adopt." The (A)(8)(b) timeline lists August 31, 2014 as the start date, and lists September 1, 2015 as the date of final adoption by the States.

* Important milestones not covered by the (A)(8)(b) timeline include system design, development of a test security plan, development of common assessment administration procedures, development of common performance level descriptors, and full field testing of the IT system (beyond the initial testing of items).

Although the timeline includes the identification of a vendor for curriculum and formative materials, there is no timeline or plan for the actual creating of those materials or for this program component as a whole.

* Consistent with the consortium's de-emphasis of these points overall, there also are no milestones in the timeline for assuring that the following occur in a timely fashion: providing system capacity for evaluating schools, principals, and teachers based on student outcomes; creating a growth model or the raw materials for one, such as a vertical scale; creating common cut scores for determining if elementary and middle school students are "on track to graduate college and career ready."

* Too little time (one month) is left for analyzing the results of full field testing.

* There is an impressively comprehensive set of time frames and activities described in regard to the development of definitions of English Learner status and of accommodations for special needs students. Fully half of the timeline and plan is devoted to this one component, which encompasses no more than 2% of the overall budget. In contrast, only a few lines and almost no detail are devoted to describing the development process for the \$27 million Test Development, Delivery, Scoring and Reporting Application.

* For the vast majority of milestones and associated tasks on the timeline, the responsible entity is listed as "contractor" or "vendor." The individuals or entities within the Consortium that are responsible for managing each of those contractors and assuring that the work is completed are not indicated.

* For other tasks, e.g., all of the "distributed item writing," the responsible entity is the "Governing States." Again, there is no indication of who within the Consortium or the Governing States is responsible, and has the authority, to assure that this crucial work, distributed across 17 States, will be completed on time.

* SBAC recognizes that the many dependencies of each work stream on the others are a major risk to this project. SBAC proposes to mitigate this risk by using "the working group structure [to] distribute the work into manageable components so that State assessment experts as well as technical advisors will incrementally monitor progress and escalate issues to the Project Management Partner and the Executive Committee when appropriate." "[T]he Executive Committee will include technical and policy assessment experts who will prioritize the urgency of problems or concerns, based on their own experience and the advice of the Project Management Partner." Evidently the "State assessment experts" noted here are the contributed line members of the work groups, and the technical advisors are consultants that some, but not all, of the work groups will employ to assist them. Asking donated line members of each team and technical and policy-focused consultants working for them – both of whom are employed as experts, not managers, lack line authority, and will not have full operational information – to monitor themselves and others is not a proven management strategy. Nor is it clear that members of the PMP will have the authority to make the demands and changes necessary to assure work is coordinated among groups of employees contributed by different States and that the work is finished on time and at quality. Giving the Executive Committee responsibility to resolve every-day coordination and work-completion issues, rather than serving as a high-level decision maker and being an escalation point of last resort, also is not likely to be an effective management strategy.

* Another SBAC risk-management strategy is a plan to complete most key deliverables by the end of 2012-13, and to implement in 2014-15, leaving the 2013-14 year as a flex year -- what SBAC describes as "frontloading tasks in the project plan as far ahead as possible to ensure that there is sufficient time to move dates back as necessary to meet the operational administration requirement in 2014-15." In regard to the work of vendors, SBAC states that there will be no flexibility on quality and strict "not to exceed" limitations on cost, so "[t]ime . . . is the only variable in our project management approach that will have any flexibility, as long as the assessment is operational by 2014-15." This is a sensible strategy under the circumstances, but depends crucially on the capacity, questioned above, of the management structure to hold all participants in the process – vendors, consultants, members of the work groups, work teams in the various states – to the initial timeline, except in rare cases where slippage is unavoidable. Additionally, (1) this strategy is inconsistent with language in the MOU; (2) some of the proposed dates appear to be unrealistic; and (3) there are key milestones that, in fact, are scheduled very late in the process, providing no flexibility in the event difficulties are encountered;

(1) Item (c)(15) in the MOU says the Consortium will give States "Throughout the 2013-14 school year, access to an online test administration application, student constructed-response scoring application and secure test administration browser that can be used by the Total State membership to administer the assessment." If any slippage occurs in the initial 2012-13 finalization dates, the Consortium will not be able to follow through on this commitment.

(2) The year-to-year budget, the vast majority of which is made up of payments to vendors upon completion of deliverables, contemplates that \$73 million of the \$149 million budget will be paid out in 2013-14. This suggests that there is an opening expectation that deliverables slated for 2012-13 completion in the timeline will actually be completed a year later. This assumption could defeat the mitigation strategy.

(3) The intended completion date for item development is 12/20/12. Training for field testing of those items is to start on 2/1/13, and the field testing itself is to start on 3/1/13. The completion date for the assessment-delivery application is 3/1/13. This provides only 20 months (from the projected contract date of 4/1/11) to complete all item development and 23 months to complete the multi-faceted test-delivery and reporting application. Individually, both of these timelines is aggressive but may be workable. However, although it is hoped that the assessment-delivery application can be used in the item field testing, these dates leave no time for integrating the items into the assessment-delivery application. In other words, the two largest, separate work streams in the project, measured in dollars, converge on 3/1/13, with no time set aside to address problems that inevitably occur when two major systems, developed separately, are asked to work together.

(4) The System Design work stream contemplates the creation of detailed Statements of Work for each of six components of the SBAC plan, but it does not include an overall Statement of Work or Project Plan to set the full range of deadlines necessary for all of the work to occur on time and to assure the proper coordination and integration of the different work streams. Additionally, the contemplated Statements of Work do not encompass important deliverables, including EL and SWD accommodations; scoring; research and evaluation; creation of growth measures and/or a vertical scale, common performance level descriptors and achievement standards; and capacity of the system to permit comparable evaluation of States, schools, principals, and teachers.

* Another risk management strategy is for the Lead Procurement State to review the progress of the entire project "from a compliance perspective." Like back-up oversight from the Executive Committee, this is a useful failsafe, but not a substitute for the effective every-day management structures.

* Another major source of risk in this project is the very large number of contracts and RFPs, each one of which presents vendor-management demands, and the coordination of all of which will be a major task. The timeline indicates that in the six months between October 1, 2010 and April 1, 2011, SBAC intends to complete at least three major RFP processes at once for work budgeted at well over \$125 million. Magnifying the vendor-selection and management risk is SBAC's intention, in order to mitigate the risk of having to write so many assessment items at once, to use "separate contracts for key services, for example, item development for each of the content areas or for the technology-enabled items and performance events Additionally, more than one vendor will be selected to develop items in order to enhance competition once the development contract(s) have been awarded." Simultaneously managing not only work teams in as many as 17 or 31 States and seven work groups, but also what could be dozens of substantial development contracts is a monumental management task, requiring the Consortium to implement additional mitigation strategies.

Budget

Evaluation: SBAC clearly defines Level 1 and Level 2 budget modules and makes reasonable assumptions about fringe rates and indirect costs. Expenditures occurring through vendor contracts that have not yet been developed (approximately 98% of all expenditures) are, in general, harder to evaluate than other expenditures. The level of detail provided in the narratives for the different components of the Level 1 budget varies greatly. The Assessment Design component is impressively detailed, apart from some of the assumptions behind the proprietary cost model that was used. On the other hand, the Technology, Professional Capacity and Outreach, and Research and Evaluation components have very little detail, making effective evaluation difficult. The amount devoted to System Design may be modestly insufficient.

The amount devoted to Technology seems substantially insufficient. Evaluating the amount devoted to Assessment Design requires more information about assumptions underlying the model. The amount devoted to Professional Capacity and Outreach seems excessive relative to the usefulness of what is proposed and competing requirements.

Explanation:

**Project Management and Governance:* As is discussed in earlier comments, the attention the proposal devotes to managerial and integrative functions is insufficient relative to the attention given to the substantive aspects of the proposal. If, however, the \$2 million proposed for the technology integrator function is considered together with the \$10.4 million proposed for project management and governance, the total amount available for the managerial and integrative functions is sufficient or even more than sufficient. (This assumes, however, that the \$2 million slated for technology integration be made available, perhaps with some augmentation, for integration across the entire project, not just technology.) A modest shift of resources to permit a somewhat larger, dedicated project staff – currently only a single, insufficient Consortium Executive Director is contemplated without any explanation of that person's duties – is warranted. At the level of detail: (1) Consider whether the 90% of budgeted personnel costs designated for Washington personnel involved in the RFP and contract-compliance process could be diminished if, as is proposed above, (A) a single omnibus RFP, rather than three separate ones, is created and (more importantly) (B) a small number of general contractors emerges through that combined RFP process, thus limiting the number of contracts that Washington State's compliance personnel must supervise, invoices they must process, etc. (2) About \$391,000 is budgeted for 1470 hours of legal services, 1099 of those hours for outside counsel at \$350/hour. This seems high, both overall and in terms of the mix of inside and outside counsel. Consider whether legal services provided by GC offices in other states might provide savings here.

** Assessment Design:*

(1) The nearly \$100 million dollars allocated to this component seems high, even given the ambition of the Consortium's assessment design. Evidently, the largest portion of the nearly \$98 million (65%) of the budget devoted to this component is for teacher time to write, review and score assessment items. The portion of teacher time built into the assessment- writing and reviewing activities, and the cost of that time, is not disclosed, evidently because of the proprietary nature of the cost-model used. The proposal does, however, disclose the cost of teacher time for scoring purposes, namely, \$150/day. Overall, teacher time seems to account for something approaching \$70 million of the budget. Modestly decreasing the daily rate or the amount of teacher time required should be considered, as it would generate much-needed resources for other parts of the project that appear to be significantly underfunded. Another approach would be to use retired teachers or vendors for some of this work, again, at lower rates. The application makes stronger claims for the value of participation by in-service teachers than (1) is warranted by the research cited in the proposal and (2) is justified by the small proportion of each State's teaching corps that, in all likelihood, will actually be involved in the process.

(2) Uncertainty arises because the assumptions underlying the proprietary cost model for item-writing and reviewing are not fully disclosed. In addition to the cost of teacher item-writers, it would be useful to know why the cost of writing and reviewing interim and benchmark assessment items is as expensive (even from an item-review perspective) as summative items. In addition, because computer adaptive technology is not now in wide use for summative assessments, there is some doubt about the confidence with which these estimates were made.

(3) In costing out item writing, it is important to know not only the cost of writing and reviewing each item, but also the likely longevity of the items in the system. An item that can be used for multiple years is less expensive than an item that can only be used once, even if the initial cost of both is the same. Performance Event items can be used only once, whereas selected- and constructed-response items can be used for a longer period of time. There is some information provided on this issue, but a clear specification of the length of time over which summative items will be used, and the amortized cost of each would be helpful. The Consortium has prudently decided to create only a modest number of Performance Events (which cost three to five times more than other items to create, are substantially more costly to score, can be used for

only a single year, and require very substantial amounts of class time), but the relative cost of using Performance Events is not clearly laid out in the application. Given the impressive nature of the technology-enabled constructed response items given as examples in SBAC's application (they appear to be effective ways to assess both disciplinary knowledge and higher-order thinking skills), the relative cost and value of those items compared to Performance Events should be closely examined both initially and over the long haul based on full information.

(4) The application budgets \$2 million @ \$200/hour or \$1500/day for five sets of advisors and consultants on item content and performance level descriptors. Because these are budgeted separately, outside of the assessment vendor contract itself, it appears that Washington plans to issue an RFP for, or otherwise procure, these resources separately. The category of vendor the Consortium has used for its Interim PMP and has contracted for by way of its PMP RFP – and that also would be a member of the more managerially focused PMP team recommended below – should be able to provide these services as part of its overall contract. So should assessment experts in the States (or in their school districts). These resources might be obtained at a lower cost via one or the other of these two alternative sources.

(5) The \$280,000 allocated to training item writers and reviewers across the multiple States seems very low, given how important this training will be to the ultimate quality and consistency of the items and also to the development of Artificial Intelligence scoring systems and to "training" those systems to mimic human scoring in an appropriate way.

(6) Given concerns expressed in the "barriers" section to some of the MOUs about equipment and infrastructure costs associated with computer-adaptive assessments, SBAC may face substantial pressure to make the new assessments available, wholesale, on a pencil-and-paper basis. If so, the \$1 million SBAC allocates to paper and pencil testing in "rare" circumstances will be insufficient. SBAC does not appear to have surveyed (or plan to survey) member States in order to collect reliable information about how many schools in each State (representing how many students) currently have the computer equipment and bandwidth necessary to support the summative and interim use of computer adaptive technologies. The viability of the overall expenditure SBAC contemplates, or at least the balance of expenditures between technology-driven and paper-and-pencil applications of the new system, depends heavily on such information.

(7) Artificial intelligence (AI) scoring makes a substantial and important contribution to the viability of SBAC's system. But this technology is in its infancy when applied to K-12 assessment. The cost estimate here (\$19 million) is risky, therefore, because it is based on projections, not actual experience, and because (as the application says) "the estimates for the system training and per item fees [for AI scoring] are *lower than current prices* for AI systems and represent expectations, after discussions with vendors and results of the technology survey, of future prices and the impact that significant volumes of business will have on price." For an item with this large a price tag, additional information about the cost estimate is warranted.

(8) For similar reasons, the absence of a back-up plan in the event that AI scoring does not turn out to be as useful as hoped is a weakness of this proposal.

* *System Design*: The \$450,000 allocated to this system design – including \$50,000 for six Statements of Work – seems insufficient. Modest additional resources are required to create an omnibus Statement of Work encompassing the entire project and to add Statements of Work for parts of the project (listed above) that not currently covered by this component.

**Technology*:

(1) The detail provided here is insufficient. The application acknowledges that three major, separate systems must be built and integrated, yet the budget narrative and table devote only a single, undeveloped line item to each of those three large systems. It is not clear, for example, which one of the three systems includes the scoring mechanism. Nor is it clear whether and how much user-acceptance testing is contemplated.

(2) (b)(6) \$7.5 million for the data base, hub/CPU, reporting application; central archive data base application; and professional development delivery application (all of which are covered in a single line item with no detail) is likely to be insufficient.

(3) Probably the most expensive and difficult parts of bringing a system like this online is obtaining required data from each State (e.g., student, teacher and principal identifiers, student biographical information, student links to particular schools, classrooms and teachers, etc.). Currently, every State uses different data formats for this data. In many States, some of the data (e.g., teacher identifiers and links between teachers and students) are not collected centrally, which magnifies the problem because of the need to deal with different formats on a district-by-district basis. The important process by which SBAC will determine the format most compatible with its member States is not spelled out.

**Professional Capacity and Outreach:* Formative assessments and professional capacity to use them and other information to improve instruction are a central part of the Consortium's vision and theory of action and for good reason. The Consortium recognizes, however, that these aspects of the work must be customized to local conditions. As a result, SBAC does not intend to provide actual assessments, but instead proposes to provide resource materials at a cost of over \$5 million of the \$7.5 million for this component. The high level of generality that these materials may need to have in order to align to the various possible approaches to be taken by the different States and districts and schools within States creates a likelihood that the materials will be of limited use in fostering improved instruction. Additionally, \$900,000 is devoted to training support. The only explanation of the training funds is: "The Consortium will work through CCSSO via its SCASS model to support States in their training efforts around the new assessment system, uses of the system, and reason for its implementation. The Consortium will fund two SCASS groups at \$15,000 each for 15 States for two years." This training mechanism is not well-explained and appears to have limited utility to the project and States as a whole. Focusing this component exclusively on (1) designing common formats for reporting formative results (so those results can be seamlessly absorbed and reported along with summative and interim results by SBAC's new reporting application) and (2) an electronic knowledge management and distribution function for the formative assessment and training initiatives and innovations by educators in each of the various States would have a considerably higher utility and lower cost and free up some of the additional resources needed by the technology system.

** Research and Evaluation:* The one-page narrative here is an insufficient basis to judge this component.

** Quality control and user acceptance testing:* The amounts here seem on the low side based on experience quality assuring systems of this magnitude.

** Higher Education:* This component seems reasonable.

** Level-2 Budget:* Translations are an admirable goal, but the amount requested for them is not justified in the proposal, particularly given (1) that the translations proposed apply only to math and do not cover ELA (which is understandable but should reduce the cost significantly), (2) the failure to state whether the translated tests will simply be translations of tests developed in English for children proficient in English or will be developed from scratch in the relevant language other than English, and (3) the absence of any justification for the costs, which seem quite high and out of proportion to other budget components such as research and evaluation and technology. Additionally, there is no explanation for the list of foreign languages chosen (e.g., the number of students in member States who would take advantage of translations into these, as opposed to other, languages). Recommendations: Substantially reduce the amount of funding in regard to this component. Provide a more well-developed description of what approach will be used to develop tests in a language other than English. Consider funding translations only in a single language (probably Spanish) as a pilot to demonstrate feasibility.

Ongoing operation

Evaluation and Explanation: The ongoing, operational per-pupil expense for the summative assessments and for the package of summative and interim/benchmark assessments (where noted) is lower than the cost currently being borne by 25 of the 31 states in the consortium. This is good evidence that the operational costs are within reach of the States or are substantially lower than current costs. On the other

hand the basis for this estimate is not well-specified. In addition, to participate effectively in SBAC's computer adaptive system, States may have to incur significant additional costs beyond ongoing operational costs, for equipment and infrastructure. SBAC provides no information about what percentage of classrooms and schools in each State have the necessary equipment and infrastructure and what it will cost to provide the remaining schools with these necessary resources. Additional, unaccounted for costs to each State will be imposed by the need to conform their data formats to those in the new system. Some of the States identified these costs in the voluntary "barriers" table in the MOU, but most did not. Without more information of this sort, it is not possible to evaluate fully how each State will, or can, fund the new assessment system over time.

Recommendations:

* Combine the \$2 million set aside for a technology integrator with the \$10.4 million set aside for management and governance and procure a management team (or augment one already procured by adding a team) with strong experience managing, integrating and implementing a multi-faceted system of the proposed scale. If no single vendor with the combination of substantive K-12/assessment expertise and managerial/integration/implementation experience is available, the necessary team may be created by partnering a large-scale project management and systems integration firm with an entity that has substantive K-12 and assessment expertise. The leaders of this team should devote 100% of their time to this project.

* As a strategy for mitigating the risk caused by the division of work among several consortium work groups as well as teams in each of the member (or, at least, Governing, States), create a transparent hierarchy of responsible actors, e.g., by designating (1) an individual responsible for the work product of each work team and (2) an intermediate coordinating team tying the work groups together, which has its own responsible team leader (e.g., the Consortium project executive). Both the work group leader and the coordinating committee would need to have their authority recognized by the member States, could use the PMP as staff, and would report to the Executive Committee and escalate matters to that Committee only in the rare instances when the matters are not handled effectively through this every-day work structure.

* Insofar as work is occurring in member States (as apparently will be the case of the lion's share of item writing and reviewing), identify a project manager in each State who reports, and is responsible in some way, to either the leader of the relevant Consortium work group or to the coordinating committee or to some other individual responsible and explicitly authorized to take the steps necessary to get the work done on time.

* The budget narrative for the Governance refers to a "Consortium Executive Director [who] performs a director role to manage the production of deliverables consistent with the agreement with USED and the direction of the Executive Committee." The role, responsibility, and authority of this individual, e.g., with regard to the Executive Committee, PMP, Work Groups, consultants and advisors and work teams in the States are not stated, and the role is not indicated on the organizational chart in the application and MOU. With sufficient staff, this individual could effectively lead the coordinating group described above.

* Provide more clarity about the sufficiency of the staffing of the Work Groups, an issue noted in comments to (A)(1). Currently, it is possible that there will only be two or three members of each group provided by Governing States, perhaps only part time, and it is unclear whether the consultants the consortium plans to hire will provide sufficient work resources.

* To facilitate the Consortium's risk-management strategy of front-loading the work, develop, early on, an overarching master Statement of Work and/or Project Plan; add its creation to the timeline as a crucial milestone; and use it as the basis to determine the categories and content, guide the development, and assess the viability of the individual Statements of Work. In that process, create a new master timeline incorporating all key milestones and deliverables and consider whether realistic time frames have been developed, e.g., for integrating assessment items into the new assessment-delivery application and for coordinating final field testing and initial State use of the system. Procurement of a PMP with substantial systems integration experience and creation of an intermediate management structure, as recommended above, are also important to the success of this mitigation strategy.

* To mitigate the risk posed by so many vendors, consider combining the RFPs into a single procurement vehicle and inviting vendors to make proposals covering any subset, or all, of the proposed work. This would create the possibility that one or more large integrators would make proposals to perform collections of work cutting across what are currently envisioned as separate RFPs. To preserve the risk-management strategy for item writing, such integrators could be required to use multiple subcontractors to write different kinds of items. Doing this would enable SBAC to minimize the number of contractors it is directly managing, by dealing with a small number of general contractors, who, then, are responsible for managing various subcontractors. (If this strategy is undertaken, SBAC must enforce strong project- and vendor-management qualifications and requirements via its RFP and contracting process, to be sure the integrator/vendor has the capacity to manage many sub-contractors.) Together with other recommendations above, the goal would be to have strong integration capacity both on the vendor side (in the context of one or only a small number of general contractors responsible for most or all of the sub-contractors) and on the SBAC PMP side. Integrator-to-integrator interactions would go along way towards mitigating the risk from so many responsible actors.

* Establish a separate work group with responsibility for the various attributes of the system covered in the NIA that focus on accountability, including common performance level descriptors; common achievement standards, including cut scores for being "on track" as well as for graduation "college and career ready"; growth model/vertical scaling; and system capacity to evaluate schools, principals and teachers based on student outcomes. Currently this overwhelmingly policy-focused work is assigned to the Research and Evaluation group. It is preferable to separate policy-making and research/evaluation functions, so individuals engaged in the latter activities can remain "policy-neutral" and dedicate themselves to rigorous pursuit of the truth. Creating a separate work group for this set of activities will also assure a priority to aspects of the requirements of the NIA that currently are under-emphasized in the SBAC proposal.

* Consider decreasing the number of the daily rate paid for teachers doing item writing, review and scoring. Or consider using vendors for some of this work, at least as a pilot to permit comparisons in quality, cost and time required. These steps could generate much-needed resources for other parts of the project that appear to be underfunded.

* It is recommended, more generally, that ED confidentially examine the assumptions underlying the proprietary cost model for item-writing and reviewing. In addition to the cost of teacher item-writers, it is worth examining why the cost of writing and reviewing interim and benchmark assessment items is as expensive (even from an item-review perspective) as summative items. In addition, because computer adaptive technology is not now in wide use for summative assessments, there is some doubt about the confidence with which these estimates were made.

* Seek cost information on a per-item basis that takes into consideration the longevity of items.

* Survey member States to collect information about how many schools in each State (representing how many students) currently have the computer equipment and bandwidth necessary to support the summative and interim use of computer adaptive technologies.

* Survey member States to determine the data formats they currently use, and to estimate the cost to States of conforming their formats to one chosen by the SBAC or installing new formats. Such information is an important basis for considering the likely adoption of the SBAC system.

*Survey States to determine (1) the current capacity of the equipment and infrastructure in all of their schools to accommodate computer-adaptive assessment of all students and (2) the nature of their current data formats for the key data, and the projected cost of conforming to SBAC's format.

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score

Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	18
---	----	----

Competitive Reviewer Comments:

The Consortium has an impressive list of higher education institutions, ranging from junior and community colleges to flagship state universities to private institutions that have committed to exempting students from remedial requirements if they meet the Consortium's achievement standards and other placement criteria. According to the data provided, these institutions educate 74% of the direct matriculation students attending public institutions in the member States. Of the 31 member States, one had no participating institutions of higher education, and two others have institutions representing only 2% and 16% of the relevant students, respectively. All of the other States, individually, report that they educate well more than 30% of the relevant students. As such, on the data provided, the Consortium is eligible to receive up to 20 points.

All of the institutions signed a commitment to implement policies, once the final high school summative assessments are deployed, that exempt from remedial courses and place into credit-bearing college courses any student who meets the Consortium-adopted achievement standard (as defined in the NIA) for each assessment and any other placement requirement established by the IHE or IHE system. A number of institutions added additional language or letters noting the "non-binding" nature of the agreement, the institutions' ability to withdraw from the agreement under certain circumstances, or the institution's intention to use a student's college-ready graduation status under the new assessment system as "part of" their placement status. These detract modestly from the level of commitment displayed, but don't greatly differ from the undertakings that other States make.

Each of the institutions also commits to participate in the development of the Consortium's final high school summative assessments, and the Consortium has included a higher education representative on its Executive Committee and made numerous efforts throughout its proposal to involve IHE representatives, both as experts and as policy makers, in relevant decisions.

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		No

Absolute Reviewer Comments:

As a result of three areas of significant concern, SBAC does not satisfy the absolute priority.

In order to satisfy the absolute priority, SBAC must demonstrate in its application not only that it aspires or is willing to develop and implement an assessment system meeting certain specified requirements, but also that it "will develop and implement" an assessment system meeting those requirements. Determining whether a proposal "will" do so calls for a prediction about the capacity of the aspirations, plans, designs, and management structures actually set forth in the proposal to add up, in the end, to an assessment system with the required features. For the reasons noted below, SBAC has not provided sufficient confidence that its proposal will meet the specified requirements. In all of these cases, changes are fully possible -- and are recommended below and in other parts of these comments -- that could bring this Consortium into conformity with the absolute priority. But the proposal has not done so, yet.

First, the Consortium's proposal states in some places that it intends to produce student achievement data and student growth data that can be used to determine whether individual students are college- and career-ready (CCR) or on track to being CCR. But the Consortium has expressed an intention -- unless ED insists otherwise -- (A) not to provide a common measure of student growth and (B) not to measure student knowledge and skills against one aspect of the NIA's definition of a common set of college- and career-ready standards, namely, a common measure of whether students at a grade level other than the final high

school year in which summative assessments are given are on track to being college- and career-ready. Additionally, the Consortium does not provide a method for monitoring growth measures and "on track" standards that individual States develop to be sure that (1) they provide an accurate and comparable measure of student growth over a full academic year or course, and (2) the standards and cut points States set are sufficient to assure that students in the State are on track to graduate CCR. Finally, in its vision statement and theory of action, the Consortium ascribes little, if any, priority to the accountability aspects of its proposal, particularly when it comes to growth measures and measures of whether students are on track to graduate CCR. On the contrary, the proposal's Research and Evaluation section expresses doubt whether accurate growth measures of the sort the absolute priority calls for can be achieved and whether comparability among States in these regard can be achieved. Because the absolute priority calls for "new assessment systems that will . . . measure student knowledge and skills against a common set of college- and career-ready standards in mathematics and English language arts", and because the NIA defines common achievement standards to include measures of whether students are on track to graduate college- and career-ready (CCR), the absolute priority is not satisfied in this respect. The Consortium's proposal thus could end up allowing wide variation among States in the accuracy, rigor and comparability of their growth measures and of their measures of whether students are on track to graduate CCR. This in turn could deprive educators of the incentives, data and other support they need to assure that students make enough progress in their elementary and middle school years to graduate CCR and succeed in college and careers.

Second, as is developed in more detail in the comments to (A)(3), the Consortium does not demonstrate in its proposal that it intends to develop enough items for use in its computer-adaptive system to provide an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students. A large number of items is required in order for computer-adaptive technologies to satisfy this requirement, and the proposal includes no demonstration, or claim specific to the number of planned items, that the number of items is sufficient for this purpose.

Third, without the significant upgrades in project management that are discussed in the comments and recommendations to (A)(8), the proposal does not demonstrate that, despite the Consortium's good intentions, it "will" be able to develop and implement the planned system.

Recommendations

- * The Consortium should be asked to develop an assessment system that will measure student knowledge and skills against a common set of standards for whether students are on track to graduate CCR. This should include a clear commitment to the feasibility of developing accurate measures of student growth over a full academic year and to the development, either in common or individually, of growth measures that are demonstrated to be accurate either to the satisfaction of the Consortium or in a manner that is consistent with a standard set by the Consortium to which the member States agree.
- * The Consortium should demonstrate that the number of items in its summative item bank to provide an accurate measure of student achievement for high- and low-achieving students.
- * As is more fully set forth in the recommendations in (A)(8), the Consortium should develop a strategy for providing adequate project management and integration capacity for an undertaking of this magnitude and complexity.

Grand Total	220	124
-------------	-----	-----

Budgets

Level 1 Budget
Name: Level 1 Budget(s)

Relevant comments on the Level 1 budget are included in the comments to Selection Criterion (A)(8).

Level 2 Budgets

Name: Assessment Design - Translations

See comments on this budget component in Selection Criterion (A)(8).

Name:



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b)(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	12

(A)(1) Reviewer Comments:

- SBAC's vision is well-supported by its identification of key deliverables emphasizing a program of balanced assessment involving summative and formative components, where the goal is to improve both teaching (input) and learning (output) for ALL students.
- Assessments necessary for college and career readiness are substantially enhanced by the planned use of technology, as well as inclusion of performance events.
- Usefulness for the assessment data is enhanced by expanding reporting capacity as well as stated commitments for professional development in this important area with teachers and principals.
- Indication is provided that standards and achievement level descriptors will be internationally benchmarked; however, there is no indication as to how such an important element might be undertaken or what is meant by this terminology.
- Assessment for accountability purposes is only dealt with superficially in key deliverable #3, but is referred to as a "primary purpose" in (A)(3)(b)(iv). On the other hand, assessment for the purpose of improving student achievement is detailed exceedingly well.
- With 17 states committing to the governing role, SBAC has strong representation from around the country's geographic regions. Strength of this consortium is limited because it is lacking more states with large urban populations, like Michigan, being part of the governing group.
- Identifying a lead state enhances accountability for managing the fund and procurement activities, and Washington's record of 5 years without an audit finding enhances its selection as lead state.
- The Steering Committee is a key component in the project's governance; yet, there is no indication as to what level of office a state's representative should hold. Members of the Steering Committee may be influenced in their perception of comments at the meeting table by the rank of the individual making the observations. The submission refers to this project as "radically reshaping the education systems" which suggests that the representatives on this committee should be key decision-makers.
- The submission is strengthened by committing to a Policy Advisory Committee with key stakeholders already alerted and on standby.
- There is clear delineation of responsibilities and entitlements evident in the structure of the consortium.
- Since achieving consensus on decisions is the goal, use of simple majority vote is a questionable alternative when strong endorsement and commitment from states is so desirable.
- Producing a consolidated reporting system is a major element of an enhanced assessment program and is identified as such in the submission. The submission is vague about this aspect and, therefore, weakened by not providing some examples of what the consolidated reporting system might contain.

- State MOU's accurately reflect rights and responsibilities in the submission, and they are proactive in their requirement to identify barriers as well as action plans for overcoming barriers.
- The submission allows states to exit but the process is somewhat confusing. States may "request" to exit "without cause" which the steering committee will "act upon" within a week. Clarity is lacking in this description as to which body has the final decision.
- The indication is that the Executive Committee will involve co-chairs but there is no indication as to why this structure is deemed necessary.
- The submission commits states to expectations and responsibilities but is silent on consequences for accountability, where there is lack of follow-through once the grant funding has been entrusted.
- The accountability aspect requires tracking student achievement by using student and staff identifiers. The proposal is weakened by omitting reference to how this accountability need will be addressed.
- This section is rated at the upper end of medium.
- Recommendation: If not committed to another consortium already, including at least one of New York, California, Florida or Texas as a governing state would further enhance the consortium's credibility, because they have large, urban populations, and/or persuading some of the larger urban states listed as member states in this consortium to become part of the governing group -- e.g. Ohio, Pennsylvania.
- Recommendation: That consideration is given toward identifying a higher standard of agreement than 50% + one.
- Recommendation: That steering committee members for each state be identified as the leading educational officer.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	4

(A)(2) Reviewer Comments:

- Credibility and support for assessment featuring increased accountability is enhanced by proposed teacher engagement in both the development and scoring of assessments.
- The potential for improved student academic outcomes will be increased because of the ongoing, annual cycle of formative and summative assessments combined with the use of multiple types of measures including performance measures.
- Committing to developing a multi-media communication plan featuring score reports will increase stakeholder understanding and support for the assessment program, because the various groups can see what benefit will be accrued to their respective needs.
- The teacher-moderated scoring concept will enhance the potential for fairness to students while, also, providing ongoing professional development for participating teachers.
- The submission raises the specter of incentivizing right behaviors which, then, suggests there are potentially negative elements from the use of incentives, but it, then, is silent on how the current culture incentivizes ineffective and inefficient behaviors. For example, the current system has pushed students through without requisite skills, knowledge and attributes (i.e. social promotion), and there have not been sufficient consequences to put pressure on the system to do better.
- SBAC's significant strength is its dedication to using technology for test delivery in a broader range of outcomes; increasing access to students requiring accommodations; and provision of useful and timely reporting of assessment results.

-- The Theory of Action places a great deal of emphasis on supporting activities (e.g. professional development, formative assessments) which are necessary for moving into the pressure end of the support-pressure continuum. On the one hand, support for professional development is frequently supplied by funding participation in and release time for in-service; however, merely providing support is sometimes insufficient to motivate people to participate in professional development, and some pressure by holding them accountable for results may also be necessary. The submission would be enhanced by a more balanced focus on these two significant aspects.

-- The Theory of Action is very strong on the educator-friendly elements but its lack of focus on the accountability of the service provider merits a low on the high rating. In other words, the grant emphasizes improvement through formative assessment, but also expects teachers and principals to be accountable for student achievement in a summative fashion. This latter aspect receives relatively minimum reference.

-- Recommendation: Moderation is an essential activity whenever humans mark student work, and it is also has significant value in teacher professional development. Implementing a policy that ensures all teachers are cycled through moderation over a specified period of time will enhance each teacher's assessment capacity.

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	35

(A)(3) Reviewer Comments:

-- SBAC emphasizes the concept of fairness yet reduces its commitment to this principle by implementing a schedule with a 12 week assessment window at the conclusion of the school year. Fair testing practices, as well as the submission's words in (A)(3)(b)(iii) concerning "appropriate student access to the grade level content", require assessments being based on what students have been taught, and this elongated testing schedule reduces the amount of time students will have had for learning the grade level outcomes being assessed.

-- The types of assessments projected are consistent with the Theory of Action, and include a plan to measure standards that have been traditionally difficult to assess. This type of assessment will provide a more accurate measure of student achievement, and improve the validity of the assessment program. Certainly face validity will be enhanced throughout the education community.

-- The usefulness of summative and formative test results is being enhanced by the submission's commitment to build capacity in teachers and administrators in their understanding of information and how it can lead to more informed decision-making during the instruction process.

-- The notion of investigating distributed summative assessment as an option is interesting; however, the submission would be strengthened by indicating how this concept may actually work toward the ultimate goal of college and career readiness. There is a need to frame this investigation with a philosophical perspective on the merit of measuring enduring learning so that fairness to students is not compromised. Without considering the complexities of the relationship between short and long term learning, the assessment may drift from being truly summative, and may actually conflict with SBAC's balance continuum (figure A3-1) at the "standardization" end.

-- Reliability of the summative assessment component is greatly enhanced by the computer adaptive feature which intends to sample learning content above and below grade level for each student. Therefore, the capacity to measure high and low achieving students within a grade as well as the requirement to measure growth in achievement is significantly enhanced. In essence, computer adaptive testing is not typical standardized testing because it responds to the individual's answers. The submission is silent on this and could have emphasized this unique aspect to a greater extent.

- Basic requirements set by the department are being exceeded by making adaptive summative assessments available in grades 9 and 10. Not only does this additional component distribute teacher accountability across 3 grades for teacher evaluation, but it also enhances the reporting process for both students and parents.
- The suggested 4 point scale (i.e. below basic, basic, on track to college and career ready, and advanced) is confusing because the term "basic" implies that the level of learning is adequate. This conflicts with the stated vision of having all students being college and career ready.
- In (A)(3)(b)(iv), SBAC indicates a primary purpose of the assessment as being to provide timely feedback to students, teachers and principals and, by use of parentheses, that this purpose is "along with accountability". The notion of accountability is commendable and its presence is duly noted, but is not explained in the consortium's vision, goal and role. This is a significant understatement because explanation of how accountability relates to assessment is as fundamental as how assessment relates to improved student learning.
- Automated scoring for immediate results and a commitment to score performance events using AI technology where possible are significant strengths of the submission, because of the potential for quick turnaround of assessment results.
- Committing to developing and utilizing a vertical scale in both the summative and formative assessments is a significantly positive factor for generating utility of the assessment reports by educators during their decision-making on individual students.
- A commitment to provide test results in several languages is a significant positive feature in achieving parental interest and support, as well as understanding their child's achievement.
- In this section of the submission, SBAC references many complex concepts introduced earlier in the submission, which would have benefited from incorporating some examples to assist knowing what might be contemplated.
- The submission has a strong basis for support because of its capacity in delivering formative and summative assessments. Issues identified above keep the submission in the middle range but the design proposed is sufficiently strong to be at the high end of the middle range.
- Recommendation: That the distributed assessment option be based on a philosophical perspective that is founded on enduring learning so that fairness is not compromised. Without consideration of this perspective, the assessment may not be truly summative, and may actually conflict with SBAC's balance continuum (figure A3-1) at the "standardization" end.
- Recommendation: That a communication plan be developed regarding the ongoing use in SBAC of selected response items demonstrating how questions, when properly constructed, can be written to assess higher order thinking.
- Recommendation: That a communication plan be developed regarding the intended use and reliability of AI scoring to mitigate distrust in its use with high stakes assessments.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	28
(A)(4) Reviewer Comments:		
-- SBAC's submission benefits from alignment with an organization having experience in assessment, including the use of adaptive testing which is the foundation of its assessment strategy. That being acknowledged, the submission lacks clear identification of the types of skills that personnel involved are bringing to the quality control aspect.		

- The submission presents a clear understanding of how to construct a purposeful set of formative assessments or interim benchmarks (I/B), so that there is a large set of test items to measure small, incremental differences in student achievement. Without this approach, the I/B would be of limited value especially when measuring student achievement at the high or low end of the graded achievement scale.
- A substantial efficiency benefit of having so many states involved in the consortium is evident in their plan to collect existing test items and performance events for building both the summative and formative test banks.
- The plan to utilize regional item development meetings may not be efficient use of financial resources but will be an effective strategy for cross-pollinating support and enthusiasm in achieving critical mass across the consortium.
- Having a bank of quality test items is essential to any assessment program, and the check and balance approach of having at least two states review submitted items as well as sending 25% of items for independent review acknowledges the significance of this bank.
- The Theory of Action emphasizes the significance of involving teachers in all aspects of the assessment process as well as integrating technology to the greatest extent possible. Section (A)(4)(c) on scoring provides the relationship of these two elements as well as how checks and balances will be incorporated in both elements, thereby increasing the reliability of the scoring process. This is an extremely important process with suitable plans in place to achieve the objective.
- Test results must be accompanied with a sophisticated reporting system including drill down capacity as well as training in their interpretation, and the consortium commits to providing both.
- The submission has considered and planned all of the necessary processes for developing an effective assessment system but has neglected to adequately present the qualifications that personnel will bring to their tasks. It appears that the previous experience of the partnership involved is to be accepted as evidence that qualified personnel will be involved.
- Commitment to Universal Design is articulated and the appendix demonstrates guiding principles for ensuring this is addressed.
- There are many concepts referenced in the section, and examples of issues needing to be addressed within these concepts are demonstrated in the appendix.
- The submission is evaluated at the low-level of the high range.
- Recommendation: That the consortium include in its policy development a perspective on whether a summative assessment should award partial marks for multi-step responses. While this is a must for formative assessments, a decision to incorporate partial marks in a summative assessment should be based on a philosophical debate that is separate from the debate on formative assessments.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	28
(A)(5) Reviewer Comments:		
(A)(5) Research and Evaluation		
-- The submission addresses all elements of the goal by identifying processes to ensure validity, reliability and fairness in the assessments so that they are used for their intended purposes and accessible for all		

student subgroups. Processes are also identified to ensure that assessments are being implemented as designed.

- The listing of all topics for research and evaluation is extensive and emphasizes the full range of issues evident in the assessment field.
- All requirements for the grant have been addressed and this aspect of the submission is evaluated in the high range.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	11

(A)(6) Reviewer Comments:

- Capacity building is effectively addressed by recognizing that the communications plan must be tailored to the needs of different groups rather than a one-size fits all approach. That being stated, the communications could be strengthened by building in a feedback loop following the communication so that it becomes a two-way process.
- Efficiency needs are addressed effectively by tapping into established school system networks providing training to teachers and principals both in the short and long term.
- Communication technology such as webinars and videoconferencing will be utilized to minimize travel costs, which could be a significant budget expense especially since the effort is to incorporate regional meetings rather than state-only meetings.
- Outreach is a significant element for success in expanding the assessment aspect, and stakeholders have considerable influence with the public on how this project should be perceived. Therefore, how perceptions are formed will be important to success, and the proposal is weakened by an insufficient emphasis on keeping those in the political arena adequately informed beyond traditional written communications.
- All of the grant requirements are addressed in the submission; however, additional detail by using examples would add clarity to intended actions. Therefore, the submission is evaluated in the high range and at the low level.
- Recommendation: That the communications plan emphasize how the project will focus on two significant aspects: the quality of learning experiences being provided to students, and the improved level of valid and reliable assessments of student achievement now available.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	10

(A)(7) Reviewer Comments:

- SBAC is intending to incorporate all elements of technology that are currently available, and recognizes that some aspects are currently in their infancy and experiencing ongoing improvement (e.g. AI scoring). The technological hallmarks of the proposal were identified in the Theory of Action and will provide efficiency in test administration through CAT, as well as in information recovery and distribution through the system portal.
- Testing documents will undergo peer review and the process will be managed efficiently by using electronic means.
- In essence, technology is effectively and dynamically incorporated into every aspect of the submission including scoring student's written responses: albeit, scoring will be handled by both human and technology means.
- Software developed in the project will be available to non-participants.
- The credibility of this submission's efforts in maximizing technology is enhanced by a partnership with an organization having extensive experience in this critical area. The submission is awarded all points available because of the extensive use of technology being planned for all aspects of the project.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	25

(A)(8) Reviewer Comments:

- Basic requirements of (A)(8)(2) have all been addressed and a chart in the appendix provides a detailed overview of WestEd's project management activities.
- In the event that this submission is approved for funding, the preliminary project management company has excellent credentials including a long-time track record in related areas. Should a different organization then be selected through the competitive process, it can only result in a strong management component becoming stronger.
- Frequent and transparent progress reporting is integral to effective project management. The plan commits to a weekly process and is strengthened by the use of front-loading on work plans and, further, is supported by daily 'stop light' reports. Such a process instills a form of healthy comparison between what the project management team members are accomplishing and the time-line identified in the project plan.
- CAT is a significant strength in this submission because of the sophisticated technology platform being used. Failure of this system provides a significant risk and the budget includes a paper-form contingency in the event of system failure.
- Project management is enhanced by the plan to hire an external third party to audit activities.
- The project work plan and accompanying timelines for major milestones are sufficiently detailed and commensurate with grant requirements.
- Efforts to increase efficiency and control costs are evident in the plan to use existing technology through the Michigan item authoring and banking component.
- Currently it is likely that expenditures on professional development are occurring within each state and the submission is silent on leveraging these resources to the project.
- The Theory of Action commits to developing the professional capacity of teachers and administrators as well as communicating to separate groups. The budget set aside to accomplish these activities is of a sufficient amount to make a significant difference.

- The cost of assessments will be a significant issue in sustainability of the project for the states involved. Projected assessment costs of \$19.81 for the summative and \$7.50 for the formative (per pupil) is a cost savings for the majority of states which should translate to sustainability requirements set by the department.
- Project management requirements are at the mid-range of the high level.

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	20

Competitive Reviewer Comments:

- In their signed letters of support, IHEs have provided strong commitment to participate with the consortium in the design of the assessments and ensure they measure college-readiness.
- IHEs are committed to implement policies exempting students from remedial programs upon successfully meeting the consortium-adopted achievement standards for each assessment.
- The grant requirement of 30% of direct matriculating students in public IHEs in the consortium's member states is significantly exceeded to 74%.
- Commendably, the initiative has already been taken to orientation meetings with IHEs, and the consortium has committed to including one IHE representative on the Executive Committee. Further, two IHE working groups will enhance communication and facilitate a smoother transition for students at a significant point in their education careers.

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes

Absolute Reviewer Comments:

- The consortium is committed to all requirements in the actual assessment program and, because of the use of CAT, has capacity to readily and more accurately measure the achievement of both high and low performing students.
- The intent to implement a vertical scale facilitates growth measurement.
- There is a commitment with strategies to measure all standards, including those which have been difficult to measure in the past.
- Accommodation issues are addressed to measure achievement of all sub-groups.
- A grant requirement to include at least one high school grade is exceeded because states can opt into optional assessments being provided for grades 9 and 10.
- The submission commits to provide data that meets requirements of accountability, evaluation of staff, as well as identifying areas where program and personnel improvement are required.

-- Assessments provide for both summative and formative data; however, the summative component requires a questionable length of time in the testing window.

Grand Total	220	173
-------------	-----	-----

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

-- Level 1 budget is within the total dollar amount available from the grant but does not allocate any funds from state internal budgets, nor does it have any commitments from additional external sources.

-- Budget expenditures are being minimized by extensive use of on-line meetings and utilizing in-kind services from within state personnel where ever possible. However, it is unclear in the submission as to whether teachers from the participating states will be paid for their contributions in the critical task of item development. If this task is, indeed, perceived as a cost-saving measure, it becomes a risk item for the budget.

-- Accuracy of the submission's budget is made more reliable by the many references to rates on labor-related functions published by the Department of Labor.

Level 2 Budgets

Name: Assessment Design - Translations

-- This submission relates to accommodations and could fit into the level 1 submission if the additional funds are not available.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b)
(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	18

(A)(1) Reviewer Comments:

The Smarter Balanced Assessment Consortium's proposed governance structure includes detailed criteria, processes and relationships among the member states as to how it intends to successfully design, develop and implement the proposed common assessment system. There are 31 states in the Consortium, 17 of which are governing states and 14 are advisory states. The state of Washington acts as the Lead States.

Although the organizational structure of the Consortium is complex and involves many different groups, differentiated roles and responsibilities, operation protocols, and other key elements are clearly spelled out and delineated to ensure that the Consortium operates in a consistent and coherent manner. A Steering Committee, for example, composed of one representative from each governing state, has a comprehensive role overseeing the overall direction of the Consortium, while the Executive Committee, a subset of the Steering Committee, has responsibility for the development of the assessment system, major purposes and uses. Committee and working groups provide support based on skills and expertise. The Technical Advisory Committee, in particular, plays a crucial role providing input and feedback to the Consortium in the areas of curriculum/instruction, assessment design, and technology. A group of nationally-recognized organizations, such as the American Association of School Administrators (AASA), the American Federation of Teachers (AFT) and many others, have committed to serve on the Policy Advisory Committee. The state of Washington will manage funds and perform procurement on behalf of the Consortium as the Lead Procurement State. Consensus is the goal for all Consortium decisions.

Guiding the work of these groups is the Consortium's compelling vision for a valid, reliable and fair common instructional and assessment system that can fully measure the knowledge and skills represented in the Common Core State Standards as a means to prepare all students for post-secondary success in college or career. In pursuit of this vision, member states have committed to support the work of the Consortium as specified in their Memorandum of Understanding (MOU), and have agreed to terms and conditions for participation in or exit from the Consortium.

There are however areas of concern that can impact the Consortium's ability to evolve as the organization grows and changes over time. Examination of each member's MOU shows that the level of commitment to adopting and implementing the assessment system may be dependent on continuous funding and/or legislative approval to address conflicts between local policy and the implementation of the assessment system. Among the Governing states, for example, three have yet to adopt the Common Core State Standards (ID, OR, NM), six identified uncertainty about the level of state contributions or risk of funding shortfall as barriers (UT, CT, ID, NC, OR, MI), and three had issues regarding the alignment of testing components with the proposed assessment system (WA, CT, NC).

(A)(2) Theory of Action

	Available	Score

(A)(2) Theory of Action	5	4
<p>(A)(2) Reviewer Comments:</p> <p>The Consortium's proposed theory of action, that <i>all students leave high school prepared for post-secondary success in college or a career through increased student learning and improved teaching</i>, appears to be well grounded. The Consortium makes a good case that it will accomplish this goal through an integrated system of standards, assessments, curriculum, instruction, and professional development that should be in place by 2014-2015.</p> <p>Consistent with the Theory of Action, Consortium states have committed to adopting the Common Core State Standards to guide college, career and grade-level expectations for students across the range of performance as well as the knowledge and skills required at each grade level to meaningfully articulate progress towards these expectations by 2011. Additionally, member states have committed to a common measurement for student performance, including performance level descriptors and achievement standards, as well as common assessment administration procedures, common definition of English language learner, student participation in testing, and accommodations for English language learners and students with disabilities. Finally, use of innovative technology approaches serve as the common denominator to effective communication to districts and schools, engagement and training of teachers, and the delivery, scoring and reporting of assessment results.</p> <p>While extensive use of computer-based technology is expected to provide timely, useful and effective assessment information, there is an area that could prove problematic to the Theory of Action: how realistic is it to expect all Consortium states to be at a stage in their technology infrastructure that can conducive to a seamless implementation of the assessment system <u>online</u>? At least one Governing state (NC) has voiced its concern in its MOU about having the necessary technology infrastructure to deliver the assessments online.</p>		

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	48
<p>(A)(3) Reviewer Comments:</p> <p>The Consortium does a good job in describing the technical specifications that characterize a well grounded assessment system. As designed, the system will provide evidence of student performance on complex tasks as well as on specific concepts, facts and skills through integrated summative, interim and formative components featuring a range of items designed to measure higher-order thinking skills that are aligned to a common set of state standards. By sampling skills critical toward attaining benchmark goals several times per year, the proposed system design could provide useful information to identify students at risk who may need additional/different instruction. Furthermore, the proposed assessment system design includes explicit procedures to establish the sine qua non elements to testing: validity, reliability and fairness. The proposed system also takes advantage of technology advances in online adaptive testing to provide accurate measurement of individual student achievement and growth, as well as application of data mining and other technological tools to increase teachers' assessment literacy, and maximize distribution and meaningful use and interpretation of assessment data to inform instruction, guide curriculum and professional development decisions, and further parental involvement.</p> <p>Although the Consortium does include provisions in their Level 1 Budget module related to Technology architecture, it would be nice if the proposal provided more detailed information as to how the Consortium intends to address some issues associated with the computer delivery of the assessments, reporting of student data, central archiving of records, and professional training that could potentially make or break the system. Considering that the system will accommodate the needs of a student and teacher population in up to 31 states, the concern here is how will the system specifically be designed to scale and handle load?</p>		

What redundancy will there be to deal with network outages, and to handle peaks in client traffic? What will be the technical support processes schools and districts will receive, such as incident management, problem management and escalation processes?

Another area that could benefit from more detailed information is how the system will keep track of and provide access to all current and historical student assessment data, whether teachers will be able to customize their use of the system, including ability to score tests based on their specifications, ability for any report to export data in user friendly manner to various formats (Excel, PDF, XML, CVS, etc.), freedom to use reporting tools of district's choice to write reports directly against the system's database, ability to provide a history of cumulative testing by student, and ability to incorporate in the system their own specific decision-making rules as to when they should intervene instructionally.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	30

(A)(4) Reviewer Comments:

The Consortium's approach to the development of the proposed assessment system is consistent with industry standards, both in scope and sequence. Specific attention is given to the initial task of identifying the learning outcomes to be expected from classroom instruction with input from key stakeholders and content specialists as a first step in the development process. The activities it envisions for item and test development, scoring, equating, field testing, assessment accommodations, standard setting, collaboration with external sources of expertise and quality control measures are appropriate for the constructs to be assessed. Of interest is the formulation of a model (responsible flexibility) to support balance, comparability and flexibility and the proposed use of within-year learning progressions as a rationale to ensure that the various components will fit together. Finally, all of these activities are predicated on the inclusion in testing of all types of students.

One area that will require special attention once the system is in implementation is to validate the claim that current technology for building authentic problem-solving tasks—in the form of technology-enhanced adaptive items and performance events—makes possible the use of this type of items in ways that are appropriate for large-scale assessment. Because this type of leading-edge assessment tasks often involves advanced computer-based simulations, some states may not have the technological capacity for implementation.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	30

(A)(5) Reviewer Comments:

The Consortium has taken several steps to ensure that the assessment system has integrity and is characterized by high technical quality and psychometric soundness. As summarized on Table A5-1, the Consortium's specific short-, medium- and long-term steps related to collection of evidence of validity and fairness are fairly exhaustive. The proposed work includes IHEs and representatives from the workplace in defining college- and career-readiness, developing benchmarks in grades 3 to 8 in pathways to college- and career-readiness, and alignment activities. A similar picture emerges from the proposed steps to establish measurement reliability and precision. The Research and Evaluation Group, for example, will conduct an evaluation plan based on a reasoning-from-evidence research model to collect and evaluate data to support evidence of different types of validity (including content, construct, predictive, and

consequence-related validity as well as measurement reliability and precision) on an ongoing basis during the all phases of design, development, and implementation. The research plan includes close collaboration with the Consortium's Technical Advisory Committee as well as IHEs, workplace representatives, and content experts in English and mathematics who have an understanding of the Common Core State Standards and associated college and career outcomes.

Of particular importance are psychometric research and evaluation activities that will be conducted for the summative assessments (achievement and growth measures) and the optional interim/benchmark assessment in order to help validate the use of these data for teacher and principal effectiveness. Equally important is the expected research to address the computer adaptive platform as an effective tool for high-stakes purposes, including independent alignment studies of items in the item pools at each grade, item-level analyses to examine the validity of claims of strong item-to-content and item-to-cognitive complexity match, and the extent to which items and performance events address the full range of the Common Core State Standards. Finally, within-year and across year studies of comparability across diverse states in terms of students achieving the end-of-year performance standard at each grade level and across key groups in each state is commendable.

The Consortium's decision to allow States to select a model to measure students' changes and growth in performance over time that is best suited to their particular needs acknowledges the difficult task involved in selecting an appropriate growth model, especially in light of the controversy surrounding the validity of using vertical scales for measuring student growth. It also reflects the need to ascertain what type of growth is best measured with computer adaptive testing strategies. This flexibility, however, does not preclude Consortium-wide use of growth data for student and teacher evaluations or other desirable purpose. The Consortium will still conduct research on growth modeling (in conjunction with IHEs) related to learning progressions, how much growth should be expected at each grade level, tying growth measures to formative tools to target professional development, growth as measure of gains in acquisition or retention, growth measures of higher-order thinking skills, and feasibility and reliability of advanced automated scoring.

By way of recommendation, evaluating the predictive validity of performance standards in grades 3 to 8 and high school (on track to college- and career-readiness) should be considered as an immediate research step (i.e., performed during the assessment design and development stages) instead of a long-term step. This would ensure that the information the proposed assessment can provide to IHEs on college readiness has been verified and validated by the time IHEs begin admitting students using data from the assessment system.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	12

(A)(6) Reviewer Comments:

The Consortium's professional capacity and outreach plan, while ambitious, recognizes limitations in capacity and resources. Its focus on providing access to authentic learning experiences, fostering interdependence and opportunity for building assessment literacy is wide-ranging and consistent with the theory of action. The plan includes a Professional Capacity and Outreach Working Group collaborating with existing professional development structures in each member state and with existing regional and federal networks, such as the Comprehensive Centers and Regional Educational Laboratories. It also envisions the use of various technologies (a portal, webcasts) to maximize dissemination of professional development information across all levels within and across members states.

While it is clear that the Consortium expects member states to carry the burden of the plan, it offers few details as to how the plan will actually work at the local level. It assumes that local professional development organizations will have the resources and evaluation experience to successfully build capacity

along the stated standards, policies and practices. Yet, considering the breadth and depth of the proposed initiative, which include computer adaptive assessments that make use of technology-enhanced item types and teacher-developed and scored performance events that are likely to be unfamiliar to many teachers at the local level, initial cycles of professional development and review will almost certainly require sustained expert guidance and support beyond the provision of training modules. It will also require meaningful feedback in order to ensure consistency and fidelity of training across the Consortium. Little information is available regarding who or what will coordinate efforts of this type at the Consortium level until member states have reached a sustained level of trained professionals that can effectively take over the task of building capacity.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	7

(A)(7) Reviewer Comments:

The Consortium's approach to using technology to support and connect all aspects of the proposed system is highly innovative. Computer adaptive technology, for example, is at the heart of assessment development, delivery, and scoring. Technology is also the engine supporting analysis of data as well as sharing and disseminating information for professional development and public support purposes. Much of the proposed technology is available; some, such as web sites, dashboards, data warehousing, and electronic clearing houses and portals are widely used.

While the creation of a portal to provide access to a centralized data repository, reporting, and professional development delivery system is an efficient means to distribute information, the challenges of deploying an electronic means of connecting school districts to such resources are almost always exacerbated at the district level due to inequalities in infrastructure, resources and technical expertise. The time and effort required to get staff working together to integrate existing technologies with the portal technology should not be underestimated. Getting all the parts to work together is usually more challenging than anticipated and requires increased IT resources. In this regard, the Consortium offers little insight into how it intends to build capacity among its members and partners and how knowledge will be transferred from the contractor to the local staff. The Consortium's plan to identify stakeholder needs and the goals for a single point of distribution for information and services for each of the identified stakeholder groups is a step in the right direction.

In addition to the above technology-related questions, there are also questions about how this information will aid teachers in determining when instructional modifications/enhancements may be necessary. What will be the method that the portal will provide to teachers to help them monitor student progress (e.g., comparing a trend line against a student's goal line?) and what will be the rules to decide if there's a need to alter instruction or raise the goal? What kinds of recommendations or analysis of skills will be provided? What will be the critical elements that the portal will provide to teachers to help them with decisions to alter/enhance student performance (instructional strategy? size of instructional group? time allocated for instruction? materials used? reinforcement?) Will the portal allow evaluation at multiple levels of comparison groups? Will it document the effects of interventions? For students with an IEP, will it translate their IEP goals into expected rates of progress automatically? Will teachers be able to chart expected rates of progress and quickly compare to actual rates of progress?

There's also the question of the use of artificial intelligence algorithms to score items and performance vents on the summative assessments. This is still a technology that is maturing. Close attention should be paid therefore to the results of the evaluation of the reliability of measurement of this technology.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	26
<p>(A)(8) Reviewer Comments:</p> <p>The Consortium is in the process of procuring a Project Management Partner. In the meantime, it has contracted the services of WestEd to serve as interim Project Management Partner. This is a company that possesses experience and expertise suitable to the task. WestEd, for example, has worked directly with multiple SEAs on large-scale assessment issues, including providing support to states as they conduct crosswalks between the CCSS and applicable state standards or that are transitioning to instruction based on the CCSS. The company also has experience with state consortia and is the lead advisor to the Bill and Melinda Gates Foundation College Readiness Assessment Pilot.</p> <p>WestEd's proposed work plan includes the appropriate milestones and timelines to achieve the project's deliverables. The timeline is very tight, although consistent with the amount of work to complete and the complexities involved in bringing together the technology, assessments and other key elements into a logical, integrated and balanced assessment system. A key feature of project management is the inclusion of a "Plan B," or a paper-form contingency plan in the event of adaptive computer system failure. Another is the use of an external third party to conduct quality control of project plans, applications, and documentation.</p> <p>The budgeted amount requested (\$150 million) will probably suffice to cover the development costs, but absent from the estimated budget is the amount required for maintenance, licenses, and support that will be needed following the development stage of the project. Neither is any figure included to address any upgrades in equipment and/or infrastructure that may be required at the local level in order to fully take advantage of the proposed system. Appendix A8-7, Survey of Operational Costs for SBAC States, lists the estimated annual contracted expenditures for mathematics and ELA Assessments, but no information is provided for equipment replacement/upgrades.</p>		

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	20
<p>Competitive Reviewer Comments:</p> <p>In their letters of intent, participating IHEs formally commit themselves to the Consortium's goal of creating final high school summative assessment in English and mathematics to ensure that the assessment measure college readiness. Furthermore, they all concur with the ultimate goal of implementation of policies to exempt students from remedial courses and placement into credit bearing college courses based on jointly agreed upon achievement standards.</p> <p>The total direct matriculation students in participating IHEs as percent of state total is 74 percent.</p>		

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score

Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes
<p>Absolute Reviewer Comments:</p> <p>The SMARTER Balanced proposed assessment system not only meets the requirements of the Absolute Priority, but also exemplifies the type of collective state actions that can provide students with the foundation to live and compete in today's global economy. Beyond the significance of the Consortium's collaborative efforts to implement a consistent approach to rigorous standards that ensure students' readiness for college and careers, of particular note is the state-of-the-art technology approach that the Consortium is using to accomplish its goal. While there will be bumps along the road to a fully operational, integrated, balanced system to measure student achievement and growth based on common state standards--quite a historical accomplishment in itself, the approach is likely to further elevate the expectations for each and every student in the Consortium states.</p>		

Grand Total	220	195
--------------------	------------	------------

Budgets

Level 1 Budget
<p>Name: Level 1 Budget(s)</p> <p>May be adequate to cover the development costs. Lacks cost estimates for sustainability beyond the initial costs, including licenses, maintenance and support.</p>
Level 2 Budgets
<p>Name: Assessment Design - Translations</p> <p>The system provides for the translation of materials to Spanish and at least two other languages to be determined on the basis of need.</p>



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b)
(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	17

(A)(1) Reviewer Comments:

Consistency of consortium's vision, goals, role, and key deliverables with the theory of action. The consortium's vision, goals, and key deliverables are closely aligned with the theory of action for the assessment system. The vision is to build an assessment system with good validity, reliability, and fairness qualities to assess deep disciplinary knowledge and higher order thinking skills. The goal for the assessment system is to facilitate improvement in teaching and learning in order to better prepare students for post secondary success. Four deliverables are identified as (1) a comprehensive assessment system, (2) an online adaptive test, (3) an effective reporting system, and (4) a system of professional development. Jointly, these vision, goals and deliverables align very well with the nine components of their theory of action which broadly include policies and strategies for improving learning and instruction, assessments and involvement and professional development of teachers.

(Theory of action components are : (1) consortium and state policies; (2) assessment system; (3) communication of policies and standards; (4) curriculum and instructional materials and professional development; (5) use of technology in achieving goals; (6) summative assessment system; (7) interim and formative assessments; (8) involvement of teachers; (9) use of information for improving teaching and learning).

Structure and Operations. The consortium has put together a thoughtful governance structure. The excellent elements of the governance structure include a comprehensive vision of involvement of consortium members, clear roles and responsibilities for the consortium members, identification of state barriers and a plan of action in order for states to fully participate in the consortium. The Governing and Advisory States are clearly defined and identified. The members, roles and functions of each of the two key committees, the Executive and the Steering, are clearly described. The Memorandum of Understanding (MOU) signed by each participating state presents a vision and principles, responsibilities of states and management of consortium funds and organizational structure of the consortium. In particular, detailed documentation of barriers and an action plan for participation in the consortium demonstrates active involvement of states in the planning and development of the consortium.

Same level of clarity and detail is not provided on some other aspects of the consortium. The emphasis in planning and describing this governance structure has been on the involvement of states rather than the overall functioning of the consortium. Very little information is provided about decision making mechanisms across committees, collaboration and interaction among committees and task forces. For example, no details are provided about how the decisions across these committees will be coordinated and finalized, how some of the ongoing decision making for the assessment development will be approved by the Steering Committee or the state membership in general.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	4
(A)(2) Reviewer Comments:		
<p>Nine components of Theory of Action are identified: (1) consortium and state policies; (2) assessment system; (3) communication of policies and standards; (4) curriculum and instructional materials and professional development; (5) use of technology in achieving goals; (6) summative assessment system; (7) interim and formative assessments; (8) involvement of teachers; (9) use of information for improving teaching and learning. The proposed assessment system is intended to be more than a good measurement of student knowledge and competencies. Inclusion of all of these components demonstrates the comprehensiveness of the proposed assessment system and potential for being effective in making a real difference in education.</p> <p>Description of 8 of these 9 components display the thoughtfulness that went into the strategies and steps identified for each component, however, the first component lacks information about how it will be achieved. For example, the last component about the use of information by teachers, students and administrators specifically identifies three strategies for achieving the goals of this component (involving teachers in score report development, interactive resources for teachers for understanding student performance information and how students will be engaged). On the other hand, little information is given about the first component on how creating supportive policies will be accomplished. It is stated that "SBAC has committed to creating a policy environment that can support the innovative systems described in the design section of this proposal." This is indeed essential for the success of the consortium. However, no strategies or steps that will be taken to achieve this goal are presented.</p>		

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	40
(A)(3) Reviewer Comments:		
<p>The proposed assessment system is innovative and is based on best research knowledge about the role assessment can play in improving learning. The innovative elements of the assessment system include the CAT summative assessment, inclusion of performance events (computerized assessments that may take 2 classes) as part of the summative assessment, interim benchmark (I/B) assessments that are based on learning progressions, and facilitation of collaboration among teachers.</p> <p>The close link between the Summative and the I/B assessments are expected to enhance the effectiveness of both the summative assessments as source of information to guide learning, as well as for I/B assessment for tracking students' progress over the year. The common scale for and similar items between Summative and I/B assessments will facilitate this strong link between the two assessments.</p> <p>Immediate results and feedback through computerized testing, as well as score reports that will report on learning progression is promising to be useful for formative purposes.</p>		

The inclusion of sample items in the application are useful and demonstrate clearly the range of types of items that will be used.

The vertical scale across grades will facilitate the comparability of scores across grades, and the measuring of change and growth.

The key weakness of the application is the plan for determining progress towards college/career readiness. The proposed method for determining progress towards college/career readiness is based on standard setting. The standard setting process is intended to identify cut scores for Below Basic, Basic, On track to college- and career-ready and Advanced, for each of the grades 3 through 8, and grade 11. Basing progress to college/career readiness on standard setting *only* is problematic. The planned validation of these cut-scores does not include establishing how students who perform at a certain level on the vertical scale do or are expected to perform at first year university, or on SAT/ACTs, or any other criterion that connects the performance levels to performance at college or careers. Furthermore, the the third level of the proficiency scale, "On track to college- and career-ready" communicates a very different meaning about student competency levels than the other three performance levels. While, *Below Basic*, *Basic*, and *Advanced* communicate status on a learning progression, *On track to college- and career-ready* communicate standing against an external criterion. The inconsistency between the third performance level and the others is problematic for guiding standard setting meaningfully as well as for appropriate interpretation of the performance results by the stakeholders.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	28

(A)(4) Reviewer Comments:

The applicant's plan for developing the proposed assessment system is very well designed for the assessment to be ready for wide-scale administration in a manner that is consistent with the proposed design and incorporates a process for ongoing feedback and improvement. The proposed plan is comprehensive, up to date with most current measurement and assessment knowledge and has many innovative aspects. However, plans for development of formative assessment tools and resources is missing from the current plans. The missing area and all the positive aspects of the proposed plan for assessment development are discussed below.

Most current measurement knowledge and innovative aspects. Overall, the assessment development plan is based on most current measurement knowledge in the field about developing assessments and includes many innovative aspects. Two examples of innovation in the plan are the emphasis on assessing complex learning outcomes through innovative item types (e.g., performance events) and the use of learning progression for development of items as well as for guiding learning and score reporting. The commitment for designing assessments that will enable educators to improve learning is evident in the plans described for aligning assessment tasks to standards, and scoring rubrics to the learning progressions. The plan for aligning assessments with standards involves specifying standards in ways that identify what is intended to be assessed and defining how proficiency in relation to that content is defined. This plan is clearly described with specific examples in content areas.

Involvement of states. Involvement of states in the development of the assessment system is essential in order for assessments to provide meaningful information about learning in a diverse set of states, as well as for proper implementation of the system. In the proposed plan, states are involved in the development of the assessment framework, item development, as well as in the item review process. The responsibility of assessment development is distributed across regions and by experience and expertise of some states in developing certain item types, such as technology-enhanced items. These are evidences of effective

collaborative aspects of the proposed plan that is necessary for the successful implementation of the assessment system.

Roles and responsibilities of key committees. The plan clearly identifies the roles and responsibilities of the key groups: the Assessment Design Working Group, the TAC, and the Research and the Evaluation Working Group. The primary responsibility of the development will be on the Assessment Design Working Group. Ongoing review, monitoring and feedback from the TAC and the Research and Evaluation Working Group are essential aspects of the ongoing feedback and improvement.

Involvement of a diverse group in item development plans. A variety of groups will be included in item development plans. The groups will include classroom teachers, IHEs and workplace representatives, content experts, measurement experts, and specialists in assessing ELs and SWDs. Each of these groups contribute perspectives that are necessary for the development of assessments that are aligned well with classroom contexts and realities, with good measurement properties, provide appropriate measurement for ELs and SWDs, and are designed to measure college- and career-readiness. In addition, a contractor will conduct an independent review of test items.

Missing from the assessment development plan. The assessment development plan does not provide any information about the development of formative assessment tools. These tools are expected to play an important role for teachers and students in improving learning. In order for these tools to be aligned with the same set of content and performance standards as the summative and B/I assessments, their development needs to be part of the assessment development plans for the assessment system.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	25

(A)(5) Reviewer Comments:

The proposed research and evaluation plan targets validity, reliability and fairness issues in the assessment system. The plan is comprehensive and covers all key psychometric issues, takes into account the diversity of states, different perspectives and research on vertical scaling and modeling growth and presents an ambitious research and evaluation plan. This plan is promising to provide information and guidance to improve and revise the assessment system on an ongoing basis.

The proposed plan includes psychometric analyses of the summative and the I/B assessments. Particular attention is paid to psychometric models for measuring growth. The plan identifies different perspectives on growth modeling and proposes to test appropriateness of the optimal growth modeling approach for the assessment system.

The validity and fairness evidence gathering is presented as an ongoing effort, from the development phase to the implementation phase and on an ongoing basis as part of research and evaluation. The plan proposes appropriate methods for gathering construct, consequential, predictive external validity and fairness evidence.

The proposed plan considers many measurement accuracy related issues in the assessment system. These include the efficacy of CAT, the precision and comparability of precision of college- and career-readiness scores across grades and the broad performance ranges and item scoring reliability.

The evaluation plan proposes ongoing data collection for determining whether the assessments are being implemented as designed and the theory of action is being realized, including whether the intended effects on individuals and institutions are being achieved. The plan is divided into three phases: Development Phase, Implementation Phase and Long-term Steps. During the Development Phase, data will be collected regarding assessment development, including cognitive interviews to try out different types with students,

small-scale pilot studies, equating studies and expert validation of new item types. The Implementation Phase focuses on monitoring implementation fidelity, including scoring and reporting functions. The Long Term Steps evaluate the long-term effectiveness of the system.

These are very well thought out plans for ongoing research and evaluation of the proposed assessment system.

Missing from the plan. Two key issues that are expected to have an important impact on validity of interpretation of data from the assessment system are missing. These are (1) an ongoing evaluation and revision of some of the innovative aspects of the assessment system, such as the Artificial Intelligence (AI) scoring of the open-ended questions and (2) an evaluation of the alignment of the assessments with classroom instruction by collecting data on opportunity to learn. The consortium is proposing innovative and efficient technology for administering and scoring assessments. Potential problems associated with these innovations, e.g., the AI scoring not capturing student performance levels accurately, may affect the validity of score interpretations in serious ways that may require the consortium to make changes in plans, e.g., require manual scoring of some items or avoidance of inclusion of certain item types on CAT. Currently, the proposed plan does not describe such contingency plans for revising assessment design, development and the system.

The opportunity to learn information, that is information about how well classroom instruction is aligned with the core standards and the assessment, is critical for making valid interpretations. Such information is essential in order for assessment results to guide plans for improving learning. No plans are included in the proposed plan for gathering opportunity to learn data.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	7

(A)(6) Reviewer Comments:

With regards to professional capacity and outreach, the proposed Theory of Action includes a system of professional development that focuses on assessment literacy for teachers and administrators. The proposed plan for professional capacity development is consistent with the Theory of Action. The plan includes five components: (1) assessment literacy, (2) alignment of curriculum, instruction and standards, (3) formative assessment, (4) constructed-response and event development scoring, (5) ongoing dialogues about improving student learning.

The consortium is cognizant of the large numbers of teachers and administrators included in the consortium and has identified cost effective strategies for distributing capacity building efforts through existing education networks. The proposed plan identifies strategies for building capacity in relation to summative assessments, I/B assessments and formative assessment tools and strategies. The strategy for both summative and I/B assessments focuses on involving teachers in developing items. Involvement of teachers in item development will provide opportunities for teachers to become familiar with the types of items that may be in the assessment system. However, this strategy is not expected to be sufficient to enable teachers to administer assessments, interpret scores and identifying strategies to improve learning. In addition, such opportunities will be limited to a small number of teachers.

As part of Outreach, the consortium plans to develop a communication network to facilitate communication about the Common Core State Standards and the assessment system with the public and key stakeholders in each participating state. This communication is expected to provide key information about many aspects

of the assessment system: its purpose, components, test items, scoring rubrics, etc. In addition, each state in the consortium will develop a communication plan and a set of communication tools targeted to the public and the education community. These plans are sufficient for keeping key stakeholders informed. However, this communication is one-way and does not allow feedback from all stakeholders. In addition, these plans do not include strategies for building support for the assessment, for example, by providing information about the effectiveness of the assessment system on improving learning.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	7
(A)(7) Reviewer Comments:		
<p>The proposed plan uses technology effectively to improve the quality, accessibility, cost-effectiveness, and efficiency of the proposed assessment system. However, since the consortium has not documented technology related barriers in the participating states, the proposed plan may not address some of the challenges that will arise.</p> <p>The consortium is planning to use technology to support all aspects of the assessment system. These include developing, administering, and scoring of the adaptive summative assessments, performance events, and adaptive interim/benchmark assessments; developing and using formative assessment tools, processes, and practices; and accessing the SBAC system portal. A centralized system portal for all these functions will optimize monitoring quality, cost effectiveness and efficiency of different related processes and products. For example, the System Portal will support an online authoring environment that will allow training of teachers online, item submission, feedback, review, and approval in a cost effective and efficient way. The centralized aspect of the portal will facilitate maintaining quality assurance procedures for all steps of the assessment development.</p> <p>Assessment system software will be developed using open-source software and proprietary software. The system will be designed to be browser and operating system independent which will minimize user specific problems and make the system accessible to all schools.</p> <p>The consortium plans to address technology-related implementation or deployment barriers by first identifying risks across states and developing mitigation strategies to address these risks. This presents a comprehensive strategy to identifying all barriers for implementation. Some of these barriers could have been identified by the consortium in the application development phase in collaboration with the participating states. This would have enabled the consortium to be able to propose plans that would take into account at least some of these barriers in the initial planning stage.</p>		

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	18
(A)(8) Reviewer Comments:		
<p><i>Project management partner.</i> The project management partner will be identified through a competitive process by October 1, 2010. The consortium has identified a partner (WestEd) using a similar competitive</p>		

process for developing the application. This partner will also serve as an Interim Project Management Partner. The information given in the application about WestEd and the quality of application put together for SBAC demonstrates that it has relevant experience and capacity for managing large assessment systems. In particular, it has experience in large-scale project management and progress tracking, working with states on standards and assessment issues, State consortia, and current large-scale assessment projects. The consortium's intention to proceed with a competitive process for identifying a long-term project management partner demonstrates their commitment to identifying the best-qualified and economically viable Project Management Partner. Their selection of WestEd as a proposal development partner and the interim partner also demonstrates that they have a good selection process in place and a good option for a Project Management Partner.

Project work plan and Timeline. The proposed work plan presented a detailed timeline for all key processes and deliverables identifying major milestones, associated tasks, start and end dates and the responsible parties. The proposed work plan is very ambitious but doable within the time frame and the consortium presents a carefully thought out plan to meet the goals identified in the Theory of Action. There is one area that requires further thinking and revising for the successful implementation of the plan. The psychometric analyses, which include calibration and vertical scaling of items that are essential aspects of development of the test, are not explicitly identified in the included schedule. The timeline for field testing of items is scheduled to be from 2/1/2013 to 6/1/2013. The field test analysis file completion is given as 7/1/2013. This only allows one month for the calibration and scaling and preparation of the file for the item bank. For a complex assessment system, this length of time is insufficient. Field test item performance is not scheduled until 8/15/2013, therefore, revisions can be made to shift the end date for field test data file completion to a later date without affecting the rest of the schedule.

Sustainability over time. The application presents estimated costs per pupil to be \$19.81 and \$7.50 for summative and interim/benchmark assessments, respectively. These estimates indeed present great savings for most states that have per pupil costs ranging from \$9 to \$116 (with an average of \$31). However it is not clear what these estimates are based on. It is stated that "After the grant is completed, the per-student operational cost of the summative and interim/benchmark assessments are projected to be \$19.81 and \$7.50". These estimates are for the period after the grant is completed. Therefore, these costs are expected to be for maintaining and updating the system. Since no details are provided about how the states will engage in the maintenance and updating process, it is not possible to determine the accuracy of these estimates.

Budget. The consortium presents a detailed budget that is consistent with the Theory of Action and the project plan, timelines and deliverables identified in the application. The budget is presented by different categories (personnel, equipment, etc.) as well as by year for the four years of the project. Level 1 Budget clearly identifies costs by Governance (\$10 million), Assessment Design and Development (\$98 million), System Design (\$428,000), Research and Evaluation (\$5 million), Professional Capacity and Outreach (\$7.5 million), Technology (\$27 million) and Higher Education Engagement (\$1.5 million). The allocation of costs to these components is proportional to the emphasis indicated in the Theory of Action, with the majority of the budget (85%) being allocated to the development of the two major deliverables, summative and interim/benchmark assessments. A close review of estimated salaries, travel, supplies and other budget categories indicate a reasonable estimation.

Level 2 Budget includes costs associated with translation of the mathematics assessments into 4 languages and sign language. The budget is presented with the same level of detail as the Level 1 Budget. However, less detail is provided for the rationale for the Level 2 Budget. It is not clear what will be translated above and beyond the test items. No details are provided about whether the whole assessment system will be translated, and whether the cost includes language comparability research studies. Close to \$10 million is estimated for translations. This cost estimate is disproportionately high given that only \$5 million is estimated for Research and Evaluation and \$7.5 million for Professional Capacity and Outreach.

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	18
Competitive Reviewer Comments:		
<p>The consortium has received Letters of Intent (LOI) from 162 public IHE/IHE systems in the participating 30 states. These LOIs indicate the commitment of IHE/IHE systems (1) to participate in the design and development of the consortium's final high school summative assessments in mathematics and English language arts; and (2) to implement policies that exempt students from remedial courses and place them into credit-bearing college courses. Signatures by the State's higher education executive officer (if the State has one) and the president or head of each participating IHE or IHE system have been obtained. Twenty-seven of the participating states obtained commitments from IHEs that represent large proportions (46% to 100%) of the direct matriculation students in the respective states. The rest of the three states (New Mexico, Vermont, and New Jersey) obtained commitment from IHEs that represent a small proportion of direct matriculation students in these states (2%, 0%, 16% respectively). Most states signed generic LOIs that stated their commitment without any other information about their strength of commitment or enthusiasm for the consortium. A small number of states (e.g., Maine, Wisconsin, Kansas and South Carolina) also provided very enthusiastic LOIs developed and signed by IHE/IHE system representatives in these states.</p> <p>The consortium has received support from IHE/IHE systems across 25 states that represent the majority (more than 50%) of the direct matriculation students in these states, and two states (Washington and Missouri) in the consortium have IHE/IHE systems that represent close to half (47% and 46% respectively) of their direct matriculation students. This is a great level of support for the consortium that was obtained in a short period of time. The potential impact of the consortium on college- and career readiness in over half of the states in the country is tremendous.</p>		

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes
Absolute Reviewer Comments:		
<p>The consortium demonstrated in its application that it will develop and implement an assessment system that measures student knowledge and skills against a common set of college- and career-ready standards in mathematics and English language arts for all students, including English learners and students with disabilities, in grades 3 through 8 and grade 11. These assessments will be administered at least once during the academic year.</p> <p>The consortium has developed processes for alignment of assessments to the full range of standards as part of their assessment design and assessment development.</p> <p>Innovative assessment types, such as performance events and technology-enhanced items are proposed for assessment of deep disciplinary knowledge and higher order thinking skills. The consortium also</p>		

proposes validation studies, such as cognitive labs, that will help determine the cognitive demands of items and whether they are capturing higher order thinking skills.

Computerized adaptive testing (CAT) is proposed for summative assessments as well as for some of the interim benchmark assessments. CATs provide measurement with similar levels of measurement accuracy for all ability and competency levels. Therefore, the proposed assessments are expected to provide accurate measure of the full performance spectrum, including high- and low-achieving students.

The vertical scaling proposed by the consortium will facilitate accurate measurement of growth across grades. A standard setting process will identify cut-scores for each grade for performance levels of students in relation to their progress towards college- and career-readiness. The standard setting process is intended to identify cut scores for Below Basic, Basic, On track to college- and career-ready and Advanced. Based on these performance level scores, student's achievement and growth towards college- and career-readiness can be determined. There are some limitations to determining whether individual students are college- and career-ready or on track to being college- and career-ready using only a standard setting process. In addition to the standard setting, a statistical linking to an external criterion that defines readiness or on-track-to-readiness is needed.

The student achievement and growth data from these assessments can be used to inform school effectiveness decisions, and for individual principal and teacher evaluations. These data should not be the only source of information for such decisions and evaluations.

Both the benchmark and interim assessments are designed to inform teaching and learning. The score reports for these assessments will include students' performance in relation to a learning progression and performance on specific learning outcomes. These types of scores will be informative for guiding teaching and learning for teachers and students.

Grand Total	220	164
-------------	-----	-----

Budgets

<p>Level 1 Budget</p> <p>Name: Level 1 Budget(s)</p> <p>The consortium presents a detailed budget that is consistent with the Theory of Action and the project plan, timelines and deliverables identified in the application. The budget is presented by different categories (personnel, equipment, etc.) as well as by year for the four years of the project. Level 1 Budget clearly identifies costs by Governance (\$10 million), Assessment Design and Development (\$98 million), System Design (\$428,000), Research and Evaluation (\$5 million), Professional Capacity and Outreach (\$7.5 million), Technology (\$27 million) and Higher Education Engagement (\$1.5 million). The allocation of costs to these components is proportional to the emphasis indicated in the Theory Action, with the majority of the budget (85%) being allocated to the development of the two major deliverables, summative and interim/benchmark assessments. A close review of estimated salaries, travel, supplies and other budget categories indicate a reasonable estimation.</p>
--

<p>Level 2 Budgets</p> <p>Name: Assessment Design - Translations</p>
--

Level 2 Budget includes costs associated with translation of the mathematics assessments into 4 languages and sign language. The budget is presented with the same level of detail as the Level 1 Budget. However, less detail is provided for the rationale for the Level 2 Budget. It is not clear what will be translated above and beyond the test items. No details are provided about whether the whole assessment system will be translated, and whether the cost includes language comparability research studies. Close to \$10 million is estimated for translations. This cost estimate is disproportionately high given that only \$5 million is estimated for Research and Evaluation and \$7.5 million for Professional Capacity and Outreach.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b)
(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	15

(A)(1) Reviewer Comments:

The consortium's vision as described in the theory of action is based on compelling principles. The goals seem aligned to the theory of action. The key deliverables include a required state summative assessment and optional interim and formative assessment tools, which seems helpful for providing states with flexibility within the Consortium governance. However, the optional nature of the non-summative two sets of tools seems not completely aligned with the theory of action, which places much emphasis on assessment for learning principles. This is described more fully in the Assessment Design and Development sections but is mentioned here because of how flexibility for Consortium governance interacts with other system objectives.

A distributed leadership model is described for the governing structure and operations. Representation on the various governing structure for Lead, Governing and Advisory states seems clear. Rights and responsibilities for these different roles seem clear. Leaving processes are identified. Statements of commitment to common policies and definitions as required in the request for proposal as well as plans for identifying state existing barriers and establishing procedures of procurement are present.

There is a great deal of strength and flexibility in the governance design for this proposal. However, some concerns that do arise include:

- Joining processes seem to depend on the approval of the Executive Committee, without complete specification of the parameters of approval to be employed in the admit decision.
- The consortium's methods and process for making different types of decisions seem somewhat unclear at times, or at least open to questions of representation. Clear lines of decision-making authority are not always present among the various committees and working groups, short of referring back to the Executive Steering Committee. A particular concern is how the decision-making authority will be divided among the Executive Committee, its larger entity of the Steering Committee, the Working Groups, and the polls of Total State Membership described. It appears that the Executive Committee decides what is referred to the Steering Committee. The Executive Committee has fairly limited representation, with eight proposed members and only four seeming to be *required* to be from different governing states. Assuming a majority rule, this could mean potentially very few of the 31 consortium states controlling decisions. However, the "culture" of the consortium is stated as collaborative, so perhaps the Executive Committee would share more of the decision-making authority.
- Addressing existing state barriers in the particular point of management of student, teacher and principal universal identifying information does not seem to be addressed.
- The timeline to adopt a common set of performance level descriptors and achievement standards seems rather late to best situate the consortium operationally. It may leave insufficient time to evaluate the impacts of the policy decisions and to examine any challenges that may arise.
- Regarding one aspect of roles and deliverables, the relationship of professional development and the role of the consortium seems somewhat inconsistently described in the proposal. In one portion of the text, teacher professional development is described as "outside SBAC's direct scope of work"

but in other portions of the proposal a professional development model based on encouraging assessment literacy is described. This is described more fully in the Professional Capacity section but is mentioned here in the governance section because it includes that all consortium states are promised a deliverable of a "system of professional development."

Recommendation: Have the Executive Committee establish a set of bylaws that describe how decision-making will be delegated, and that makes the decision-making process more clearly operational on an ongoing basis for day-to-day basis.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	4
(A)(2) Reviewer Comments:		
<p>The Theory of Action does communicate a set of logical, coherent and credible principles, that if well and fully implemented seem as if they could be considered from a research-based perspective to perhaps result in improved student outcomes, as called for in (A)(2). These principles include standards-based learning progressions; an articulated set of assessments for summative, interim and formative uses; multiple measures; and an integrated role for teachers. Accommodation policies and a stated commitment to providing schools with sufficient technology infrastructure and approaches to teacher professional development were discussed.</p> <p>Actual components of the proposed system are less clearly stated in the narrative text, but are more adequately described in the Appendix A3-1.</p> <p>One major concern overall with the theory of action is that a large amount of flexibility seems to be allowed for the cross-state adoption of components of the system. Minimalist adoption could seem to undermine the theory of action principles above. Also, it is unclear how "incentivizing the right behaviors" among the consortium states described in the theory of action section either (i) relates to the theory, or (ii) will be achieved through the consortium approaches. Additionally, these incentives are not described.</p> <p>Recommendation: Consider encouraging or requiring states to adopt more than current minimum requirement in this proposal, to better support the theory of action successfully implementing across states. Also, consider how the language about incentivizing behaviors can be more clearly described and tied to the theory of action.</p>		

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	35
(A)(3) Reviewer Comments:		
<p>As mentioned in prior sections, the consortium plan and supporting theory include summative, interim/benchmark, and formative components. In this proposal section, these facets are implemented in a design that shows elements of being innovative and is somewhat feasible and consistent with the theory of action, but substantial issues and concerns do arise.</p> <p>As mentioned previously in this review, all but summative components are optional. For states selecting the minimum compliance, it seems that substantially less information would be available to support student learning and guide decision-making, and thus the proposal is much less innovative under these</p>		

conditions. Additionally, the supporting theory that undergirds this as an innovative assessment system is less consistent with the assessment design if the optional components are not implemented.

Computer-adaptive approaches are being used to help provide a more accurate measure of student achievement, including for high- and low-performing students. Above and below grade measurement ranges will help allow more students to show what they know.

Some of the performance tasks and technology-enhanced tasks seem as if they may help with those standards deemed more difficult in the past to assess. Universal design for learning approaches are not well clarified in the proposal. Artificial intelligence approaches for some types of automated scoring are mentioned but not sufficiently detailed to evaluate plans.

It is not clear from the information presented exactly how students will be deemed as on-track for career/college readiness. Benchmarking and standard setting are minimally described, and the proposal leaves overall approaches as to be decided during the duration of the project.

The mode of summative assessment and benchmark/interim assessments are well described, but a major concern with articulation between the two is that learning progressions are proposed for the benchmark/interim assessments but not for the summative assessments. This seems to be a substantial and unfortunate flaw in the concept and design of the proposal. Implications for scoring and reporting when learning progressions are not articulated across components of the system seem substantial.

Finally, the formative assessment strategies seem much less well shaped, lack systems for alignment and equating with the other assessment processes, and seem not included in the full milestones chart except for arranging and bidding the rfp. The formative assessment components seem very much a low valued "add-on" to the system, with limited integration, funding and support.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	23

(A)(4) Reviewer Comments:

The Consortium proposal for developing the assessment system has aspects suggesting readiness for wide-scale administration that can be timely, cost-effective and consistent with the proposed design. However, significant concerns exist.

The phases, processes, teams and roles for initial item development of at least the summative and interim/benchmark items seem reasonably well described, and seem to meet many of the traditional concerns in the measurement field. Accommodation approaches are discussed. Field testing scope and turnaround times are likely too constrained to meet the assessment development needs of the proposal.

Insufficient information is specified for specifics on scoring, such as what artificial intelligence algorithms will be employed and how, so scoring processes seem less possible to evaluate. Additionally, very much less is described regarding the formative components, again making these less possible to evaluate.

A major concern with the assessment development process overall is that the measurement concerns do not extend far enough to evaluate the intended outcomes of the task measures. These are to be measures of career and college readiness. Therefore, a substantial stage in the item development and iterative improvement of tasks seems as if it should encompass whether or not the tasks adequately inform on student readiness *when they enter college and careers*. A stage of assessment system development then should field test performance with college freshmen for instance, and comparison of performance scores on the assessment with college first-year success and early entry career success. This should not be relegated to late stage research, as described in this proposal, but should be an integrated and important part of initial and subsequent task development.

Another key concern is that if learning progressions are to be incorporated in this system, they should be incorporated throughout, certainly in the summative components if in the interim/benchmark. Much more needs to be said about how they will be developed, used and reported. Learning progressions, while a powerful conceptual organizer, are challenging to construct and validate, and require the employment of considerable cognitive science research. This is not described adequately in the proposal.

Also, on another topic, it seems there is very limited time between establishing proficiency levels and test impacts, and the full operational use of the system. To the extent that this is considered part of the assessment development process, if these parameters were established sooner in the schedule, it would better support that concerns and problems could be addressed and still meet the operational schedule across states, for a larger portion of the consortium.

The model of teacher professional development combined with assessment proficiency and scoring moderation seems to be a strong point of this proposal. However, note that it is not clear to what extent teachers in each state will be participating, for instance how many of them and how often, with some of these professional development practices. It is also not clear how they will be compensated and supported in this process, since this is left to the states.

The common electronic platform among the states should be helpful for data collection and reporting. It would be necessary for good utility in schools that this data warehouse would adequately encompass other data collection needed to evaluate the same students (such as high school exit requirements) so that it does not become another "add-on" to the many collection and reporting systems that schools and teachers currently have to learn, manage, understand and integrate for profiles and evidence on students and schools.

Report formats described seem helpful to schools and teachers for understanding individual and group performance, at least for the summative evidence. However, it may be hard for teachers to know what to do with some of the information for instructional decision-making, without additional assistance from the Consortium. Especially states that opt for only the minimal required summative component seem as if they will have very limited information to populate useful reports. Also, to be consistent with empowering the stated theory of action, learning progressions as mentioned above should be used across all components, and it should be possible for interim/benchmark assessment information to be fully incorporated in reports. Much more clarity should be brought to how the formative assessment components can better articulate with the system, and how they will be used and encouraged.

Finally, it is not clear what reports will make possible the other evaluative uses of the system, such as of teachers, principals and schools. But likely this will be part of future specifications once future evaluative characteristics for these areas of the educational system are made known to states and schools.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	18

(A)(5) Reviewer Comments:

The plan for identifying and employing psychometric techniques indicates (a) a composite score based on weights and presumably a sampling scheme over the standards, (b) a detailed score profile that describes student progress on a number of dimensions. Insufficient information is provided on item sampling designs, weights, student time on task, overall number of dimensions, and other characteristics to evaluate the likelihood that (a) and (b) above can be achieved through this proposal. Presumably it would be much more possible to achieve especially (b) if the interim/benchmark data is available to schools and included in reports and data display interfaces. However, since this is optional, much more detail would be needed on the specific characteristic above to determine whether for instance the minimal use of the summative component alone would be sufficient to generate this information, in ways that are valid, reliable and fair.

The strategies for measuring growth also are incompletely specified, and seem to be structured as to be determined later. It would seem important to provide more information at this time.

Comparability again is incompletely specified. More information and specifications would be required to be determined at this time in order to evaluate the approaches. Absence of a plan for putting in place unique identifiers in all the states of the Consortium for students, teachers and principals may jeopardize use of the information for both growth assessments and for school and teacher evaluations..

Throughout the research and evaluation plan, there is little discussion of how the information will be used iteratively, or for continuous improvement through the program duration.

Having the CCSS-aligned items in the non-secure benchmark/interim pool possible to be grouped into customized blueprints to meet the different needs of the State and local level seems promising.

Finally, use of learning progressions again should be more thoroughly worked into the thinking on research and development of the system. For instance, learning progressions can be a powerful approach and when fitting well suited to modeling of data and interpretation of results for teachers and schools. However, the 3-PL model specified for calibration in this system means that due to the item discrimination parameter, item characteristics curves may intersect or "cross" over the latent trait ability distribution. This makes mapping or modeling of learning progressions problematic. Rather than indiscriminate use of 3-PL, a range of more nuanced and innovative modeling approaches have not been considered in this proposals. For instance, models can be combined to capture needs of both scalable dimensions and more flexibility in discrimination parameters. One example of this is modeling the dimensions with 1-PL (Rasch) models, for a subset of items well-fitting to this model, to empirically establish, adjust and validate the learning progressions, then using these items anchored and subsequently including more items equated and calibrated with more parameterized models if (i) the items show the need for more parameterized models (2-PL or 3-PL) in order to offer reasonable fit characteristics, and (ii) the items appear necessary to include in order to encompass the totality of the construct. This example is not a recommendation but shows that many more measurement models and blended approaches are available for measurement today, and might be considered for best practices in a 21st century assessment system.

Recommendation:

SB's principles encourage flexibility for states. In parts of the proposal, this seems to create problems because insufficient requirement across states is specified to ensure adequacy in the system to meet the vision and goals. In this section, SB proposes one approach to achieve flexibility that does still specify sufficient portions of the system. This involves identifying an overall standards map and then building tools that support state's use in customizing the framework to some of their goals in certain areas. Perhaps this is a model that could be extended across more of their system in some way. It would take some thought but the goal would be to allow some state flexibility but not leave the approaches unspecified and vaguely defined, as has been noted in some other comments in this review. Flexibility as an intention is fine, but it needs to be shown how this can be made operational and still coherent for a complex common assessment system such as this, as specified in the proposal guidelines. Perhaps providing "road maps" or guides to the universe of acceptable approaches, from which states can then select some options would be possible in more parts of this system. This would be preferable to leaving the approaches nearly unspecified, such as for growth models and comparability.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	6

(A)(6) Reviewer Comments:

The plan for supporting teachers and administrators in implementing the system seems to rely primarily on capacity building within existing networks. While this is a reasonable overall vision, there do not seem to be

any requirements of any participating states to use or build the networks as described in the proposal, so this seems to bring into question the degree to which the implementation is feasible, and whether it will result in the intended outcomes for students as specified in the call for proposals.

As part of the capacity building, states will be encouraged to employ teacher participation as described for the summative components. However, state inclusion of teacher professional development and use of the other components of the system are entirely optional. Common professional development requirements and goals for monitoring across states are not present, to ensure the necessary capacity building takes place. The formative assessment component seems especially overlooked in this section, if it is intended to offer teachers evidence for instructional decision-making.

The clearinghouse, dynamic reports, and educators dashboard seem useful, especially if they are integrated into the other state-based educational data systems that may need to remain in place or be integrated into the data collection process.

The communication tools as described in Outreach and Communication Plan seem to be primarily focused on community, public and policy-making stakeholders. Perhaps the communications tools and plans for student, parent, teacher, school and administrative stakeholders are assumed to be encompassed in the professional capacity building, but there is lack of detail about their communication opportunities and tools. Language translations and outreach to diverse audiences of stakeholders on the importance, consequences and utility of the new assessment programs do not seem to be present.

Recommendation: Building support for the assessments with teachers, families and communities is important. The Consortium needs to establish the importance of the assessments with the stakeholders, and show how they will improve learning outcomes and opportunities for students.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	7
<p>(A)(7) Reviewer Comments:</p> <p>Proposed technology uses enter effectively into the assessment design, development, and delivery for this Consortium. However, data warehousing and other back-end data tools and systems are not well described or accounted for in milestones and budgets. These include administrative systems, and all the data warehousing needed for scoring and reporting for this Consortium.</p> <p>Some of the technologies to be used are extant and have been effectively deployed in consortium member states to date. Other technologies may need to be developed or adapted for uses of the consortium. Once developed or adapted, if successfully used in numerous of the consortium states, it seems reasonable that they could be re-used by other states and governmental agencies as well. Some are specified as proprietary, though, so perhaps may be owned by vendors and therefore may not be available for free use-transfer.</p> <p>Inter-operable standards are to be used by the consortium, and this will be helpful.</p> <p>In regard to issues of access, the consortium has stated that having the necessary infrastructure in place in schools is a high priority. How this priority will be fulfilled and sustained, however, is not clearly described at least in this section of the proposal. It should be noted that much of this technology infrastructure in schools such as hardware is likely to be reuse, drawn into assessment from other current school uses of technology. While this may improve assessments, it may have unintended dampening effects on other important teaching and learning uses of technology in the consortium state schools, unless funding is identified. The new demands will also impact technology support personnel in schools, and additional plans for funding technology support personnel do not seem to be present.</p>		

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	20
<p>(A)(8) Reviewer Comments:</p> <p>The project management partner for this proposal appears to be specified for the short-term, with a process of selection in place for the longer range. The specified project management partner does seem to have an appropriate mission, date of founding, and size for the project. Key personnel have engaged in substantial assessment related work, although not of this scale in the past. Percentages of time for some key personnel early in the project seem low given the scope and magnitude of the Consortium's goals.</p> <p>The project workplan and timeline represent some of the tradeoffs as described previously in other sections. Specific concerns include:</p> <ul style="list-style-type: none"> • No milestones are apparent and specifically identified for framing and refining learning progressions, yet this is an important part of the theory of action. • The extended time of about one year between the evaluation of Consortium submitted items/gap analysis and the distributed item writing stage delays other key stages. • This long period for evaluation of extant resources pushes back the timing of item writing, and pilot and field testing. • One result is the very late dates for establishing the adoption of achievement standards and evaluation of impacts, as well as for implementing the systems in schools and establishing the capacity for use. 		

- Specifications for milestones on adoptions of growth model approaches, comparability approaches and other characteristics of how the system will operate, which were left open in the proposal description pending further planning, are not included as milestones here.

Regarding funding plans, a large percentage supports the assessment development process. Better examination of current assessment assets held by this large group of consortium states would seem to pose more opportunities for reuse and revision that could substantially lower these costs, making more funding available for other key areas described below.

A relatively small percentage of the funding is going to Professional Capacity and Outreach. Yet there are no requirements that states supplement this or re-engineer practices to directly address the need for teacher and administrator professional development for the new systems. This suggests that identifying, managing and mitigating risks associated with entities responsible for use of the system may be substantially under-resourced.

Though the goal of the assessments is career and college readiness, only a small percentage of the funding is directed to engagement with higher education or career venues. While the higher education aspect is a competitive preference priority, it seems as if goals of accessing actual college and career readiness in comparison to the consortium's measurement results should also be part of the core priorities for the assessment development process. Initial and ongoing validity studies for this aspect of the assessment system performance seem essential, but are not included in the milestones. Plans specify little in the assessment development process to ensure that the system is indeed directly achieving the goals. Reaching targets of theoretical standards describing career and college readiness does not mean these outcomes have actually been successfully achieved, especially when the standards themselves are new and relatively untested.

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	11

Competitive Reviewer Comments:

Most states in the consortium do seem to represent a large percentage of matriculating students within the IHE agreements, although a few states have limited or undeclared student inclusion. The policy statements do seem to agree with the understanding that students who meet the consortium-adopted achievement standard for each assessment and any other placement requirement established by the IHE will be exempted from remedial courses. A commitment to this portion of the language of Competitive Preference Priority (b) and (c) alone does not constitute evidence of a strong commitment from the IHE's, because first (a) must be established, which seems an essential condition to the meaningfulness and success of a commitment to (b) and (c). Support of (a) in terms of the IHEs needing to "ensure that the assessments measure college readiness" is quite weak in IHE commitment to alignment, validity and use of the system to support what could ultimately be considered empirically supported outcomes of college readiness. Following this line of reasoning, therefore a major concern here with the establishment of the relationship with IHEs is that while there is language in the consortium proposal that IHE's will participate to ensure the assessments measure college readiness, there is very limited planning and agreement for the empirical work, validity studies and evaluations that would be necessary to establish this on a sound evidentiary basis. This significantly weakens the commitment of the IHE's to the critical need of credibility of the intended outcomes and vision of this proposal.

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes
<p>Absolute Reviewer Comments:</p> <p>Absolute priority seems met. The plan for how assessment growth will be shown in high school is not entirely clear, due to the flexibility in the system that seems to allow only a single assessment to be administered in high school. Growth may then need to be measured from earlier time points, and universal student identifiers do not seem to be in place in all states.</p> <p>Recommendation: Require universal identifying data be collected for students, teachers and principals in all states, and consider requiring at least two assessment across all high school assessment options.</p>		

Grand Total	220	139
-------------	-----	-----

Budgets

<p>Level 1 Budget</p> <p>Name: Level 1 Budget(s)</p> <p>Did not see where the list of expenses associated with the development of the Formative Assessment components was included in the budget spreadsheets.</p> <p>The rates of human scoring on the constructed response items did not seem to include a full rationale and justification. This budget figure will be low if the AI scoring model does not prove sufficient. However, without more complete explanations of the AI technology to be used and its validation in research studies, it is not possible to determine to what extent the scoring procedures used here might be considered fair, valid and reliable, and whether the read-behind rates and thus funding for them are appropriate.</p> <p>As mentioned previously, the percentage of funding for assessment development might be reduced by additional investigations of current state assets. Also travel costs seem high, and perhaps could be reduced through more virtual interactions. As noted previously, the funding toward professional development capacity seems insufficient, without an additional agreement or requirement of states on how additional or outside monies will be used toward this objective.</p> <p>Also as mentioned previously, it could be helpful if some of the assessment development expenses were moved up earlier in the spending cycle, to allow more time for processes in and subsequent to the field trials. These include setting of standards and evaluation of policy impacts. For instance, it seems that collection of current state assets and gap analysis especially could proceed more quickly, and in tandem with specifications for item writing. This could allow the item writing funding to move up to a year earlier in the cycle.</p>

<p>Level 2 Budgets</p> <p>Name: Assessment Design - Translations</p> <p>The additional budget request for translations seems high, especially as compared to expenses in some other areas, such as research/evaluation and professional capacity development. Costs for the translations do not seem well detailed. Also, given the difficulty of scaling computer-adaptive testing banks across languages, requiring the rescaling and field testing of translated items, it would be important to employ other approaches</p>

as Universal Design for Learning to lower the accessibility barrier in these mathematics items wherever possible without requiring translation. Note that the main proposal as a whole does not include much discussion of Universal Design practices, which can also help lower the accessibility barrier for English Language Learning students and others with special needs. Such practices should be employed even in the presence of translations. Also note that many more languages will be represented in school populations than will be available through the translations. Finally, it may be important to consider translations at least for instructions and directive prompts in the ELA tasks, rather than reserving these exclusively to mathematics, as seems to be suggested in the Level 2 Budget Module.

Recommendation: Reduce the Translation budget to a research & development pilot project, and fund it as a study in just one language, for instance Spanish. Include public dissemination of the results so that the work can be adopted and scaled more widely if it proves promising. Encourage UDL rather than translations wherever possible, and require item banks to be field tested and rescaled if translated.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



SMARTER Balanced Application #SB (b)
(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	12

(A)(1) Reviewer Comments:

(a)

The vision adequately describes key deliverables, including: a comprehensively designed assessment system, an online adaptive test administration, a consolidated reporting system, and a system of professional development. It leaves out the accountability deliverables.

The proposal refers to "internationally benchmarked" standards but does not specify which ones or indicate how the benchmarking would be done.

The proposal suggests that professional development focused on test literacy is feasible via teacher involvement in item development and scoring. While this is a worthy goal, and it will provide benefits, it would involve only a small minority of teachers and neglects fundamental topics, e.g., test interpretation.

(b)

The proposal does not address the need for executive/steering committee by-laws, (quorum rules, attendance requirements, etc).

The executive and steering committees provide adequate representation for participating states. However, the roles of the executive (overseeing development) and steering (overseeing the "broad picture" and "supporting" the executive) committees are not clear. Criteria for executive committee membership are not described.

The executive committee is to have only one institution of higher education (IHE) representative. This means that the interests of both community/junior colleges and 4 years IHEs could not be adequately represented.

The executive committee is to have two co-chairs, self-nominated, and selected by a vote of the steering committee. One co-chair would initially serve a three-year term and otherwise terms would be two years. A structure with two co-chairs, sets up a situation where co-chairs could disagree, possibly compromising the integrity of the project. Multi-year terms do not allow for timely change in leadership, as may be needed.

Although the proposal provides adequate discussion of Policy Advisory Committee membership, it says little regarding the Technical Advisory Committee.

The start dates for the consortium's policy and definition time line are all identical (10/1/2010). Some activities cannot begin until others have ended. Realistic start dates are needed.

The proposal does not provide increased staffing for the lead procurement state that is needed to handle a significant increase in workload.

(c)

Examination of the memorandums of understanding (MOUs) reveals that some states did not provide a plan for identifying and removing existing state barriers.

(d)

The procurement process is adequately documented in the MOUs.

Recommendations

The interests and activities of community/junior colleges are different from those of four year IHEs, and are more focused on career/technical education. There should be at least two IHE representatives: one from an organization representing four year IHEs, and one representing community/junior colleges.

A more coherent structure for the steering committee would have one chair and one vice-chair both serving one-year terms. This would reduce the possible of division between the co-chairs and allow for change in leadership as desired.

It would be useful if some technical advisory committee (TAC) members were drawn from interested organizations of national scope (e.g., CCSSO, NCME, ACT, College Board) and not just IHE faculty.

The plans for identifying and removing barriers should be obtained from all involved states.

The budget should provide for additional dedicated staffing for the lead procurement state to handle increased workload.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	1

(A)(2) Reviewer Comments:

The theory of action reads like a summary of the proposal, not like an underlying theory. The concept of instructionally focused testing as a basis for improvement has some merit. However, absent a strong accountability component, the rationale for improved student achievement and college and career readiness (instructionally focused testing and reporting) ultimately is not persuasive.

The pictorial schematic (Appendix A2-1) is unhelpful. There are no indications of inputs, context, processes, or outcomes. Interpretations of the arrows and boxes is not provided. To state that the theory of action is more recursive and multidimensional than what is graphically depicted begs the question.

The proposal does not make clear how "Teachers, students, and administrators use information from instructionally useful assessments to improve teaching and learning." Lacking is a description of how an accountability system for schools, teachers and principals would operate in this new test environment to produce the desired changes. The NIA states that the assessment system must inform evaluation of schools, teachers, and principals.

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	30

(A)(3) Reviewer Comments:

Because the theory of action is not well specified it is difficult to evaluate the consistency of the design with the theory.

(a)

The proposal suggests administration of the summative test during the last twelve weeks of instruction. A testing window this wide may allow states more flexibility, but poses increased risks to comparability and security.

(b)

The proposal refers to "full" system field-testing in spring 2014, but does not say how the sample will be obtained. If voluntary, it will be hard to obtain a good enough sample for the necessary item calibration, scaling, standard setting, and subgroup analyses.

The proposal mentions the use of alignment studies to ensure coverage of the full range of standards. While these methods are well understood in the context of paper/pencil testing, it is not clear how they apply to computer adaptive testing, where each student presumably sees fewer items that are selected by the computer using an algorithm based on prior responses.

The proposed use of the Access by Design Model reflects a good approach to including ELLs and SWDs because it will build accessibility into the system throughout the development process, beginning with blueprints and continuing through to item and test construction.

The proposed formative tools, administered and scored throughout the school year, should provide timely and useful information for teachers.

(c)

SBAC's proposal is unclear on the use of data to inform determinations of individual principal and teacher effectiveness. For example, the wording of the MOUs states only that scores can be used to "better understand the effectiveness and professional development needs of teachers and principals." According to the NIA: "Systems grants must produce data (including student achievement data and student growth data) that can be used to inform (a) determinations of school effectiveness; (b) determinations of individual principal and teacher effectiveness for purposes of evaluation."

The proposal refers the reader to section (A)(5) for a description of methods to develop a vertical scale. However, (A)(5) does not describe a specific method.

Summary tables for (A)(3) (pp. 59 – 71) indicate unrealistically small numbers of items for the test components, which are inconsistent with the budget narrative (Appendix A8-6).

The proposal specifies a paper/pencil version for states/districts without the technology needed for a computer adaptive administration. While this is a laudable goal, the effort to develop, administer, score and report a paper/pencil test of acceptable quality is likely greater than SBAC expects.

The proposed development of data mining tools (similar to what NAEP now provides) should be useful to teachers, administrators, and researchers. This kind of tool provides intuitive pick-lists for variables and types of analyses, combined with real-time reporting of results. Users need not purchase software programs, e.g. SPSS, or learn how to use them.

Recommendations

Adopt a testing window shorter than 12 weeks. Ideally, a testing window should open after a set number of days of instruction and should close before states or schools can game the system by waiting as long as possible before testing. A wide window helps the test contractor if administrations are spread out, but schools/districts tend to procrastinate until the last possible moment.

SBAC should explain how alignment for computer adaptive tests be assured during item development, population of the item bank, and actual test administration.

SBAC should commit to and describe a method for vertical scaling and measuring student growth. The capacity to measure growth must be designed into the system from the beginning and cannot be added in

an ad hoc fashion at a later date. Early design decisions will determine the capacity to do the required vertical scaling one way or preclude the ability to do it in other ways.

SBAC should commit to developing an assessment system that supports the evaluation of teachers and principals.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	15

(A)(4) Reviewer Comments:

(a)

The item development plan assumes that some selected response (SR) and constructed response (CR) items will be supplied from existing State item banks and contributed by States. This assumes that such items are owned by States (not by vendors) and can be legally donated to the consortium. It also assumes that any such available items will be technically adequate, will suitably align with the Common Core State Standards (CCSS), and can be delivered in a usable format (not in a proprietary item bank). There is a substantial risk of insufficient items to support the assessment system.

The proposal indicates that responsibility for developing, reviewing, and editing items will be distributed to States, with individual States deciding how and when to conduct the work. There is a substantial risk of inconsistency in standards for item development and review.

The proposal indicates that a sample (25%) of items be submitted to a contractor to review quality and alignment prior to pilot- and field-testing. Quality control (Q/C) checking of items should be ongoing and 100% of items should be flawless for field-testing and final statistical review. Discovering flawed items later in the development process wastes resources and damages credibility of the program with pilot- and field-test participants. A sampling approach to item quality is not acceptable.

(b)

SBAC's commitment to universal design principles is praiseworthy, and the proposal notes that the necessary work to develop test accommodations will be done one way or another. However, the proposal refers to a different grant proposal to implement an accommodation study. The proposal may not be awarded and/or the work performed under the grant may not be useful. (See discussion of the budget narrative below regarding the allocation of \$1.5 million for an accommodation study, potentially funded by a different grant.)

(c)

The proposal describes the use of state of the art procedures for scoring. However, regional scoring centers are to be established to train teachers, followed by the development of a web-based portal to support distributed training, calibration, and scoring. The initial use of regional scoring centers could pose problems of logistics and consistency across centers.

The section on monitoring refers only to teacher scoring of performance items. The proposal does not describe processes needed to check for irregularities in the administration of SR and CR items by teachers.

(d)

Procedures for reporting laudably refer to innovative technology, which will be needed for administration and reporting. However, it does not commit to a specific schedule for timely reporting of results.

The proposal indicates that score reports are to be provided in only three languages. The limit of three languages neglects the needs of a large number of students and parents who speak other languages.

(Note that this is distinct from the Level 2 Budget proposal for translated versions of the tests, which may not be as important.)

(e)

The plans for Q/C and field-testing appear to be adequate, except with regard to monitoring the participation of students with disabilities (SWDs) and English Language Learners (ELLs) and the provision of appropriate accommodations for these students. The proposal suggests that over-sampling these populations will be sufficient. However, it is likely that more aggressive measures will be necessary. Schools and districts are often reluctant to test these students, either because their scores tend to be low or because extra work is required to provide and document appropriate accommodations.

Recommendations

The Smarter Balanced Assessment Consortium (SBAC) should centralize item development and review processes and should decrease its dependence on states for contributed items.

SBAC should conduct 100% independent quality/control testing on test items.

Instead of regional scoring centers, there should be a single scoring center for the summative test.

The proposal should describe processes used to check for irregularities in test administration.

SBAC should commit to a specific schedule for timely reporting results.

Test results, test administration directions, and interpretive materials should be provided in more than three languages for students and parents who are not fluent in English. These translations are relatively easy to do (as distinct from translating the tests - a much more difficult task) and should be provided for language groups of students who exceed a given size threshold, e.g., 1,000 or more.

Field-testing should include procedures for monitoring administration to SWDs and ELLs to assure adequate participation and provision of accommodations as intended.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	18

(A)(5) Reviewer Comments:

(a)

The proposal lists a variety of desirable steps/studies designed to verify validity and fairness during design and development (beginning), transition (implementation) and long-term.

The proposal suggests that States serve as laboratories to examine the strengths/weaknesses of various growth models. This means that the selection of a growth model could lie far off in the future, possibly subject to political considerations irrelevant to good measurement. The successful implementation of any growth model likely requires early design decisions that in turn will allow necessary kinds of item development, calibration, scaling. Postponing the selection of a growth model risks having a suboptimal test design needed to support the actual measurement of growth.

Allowing States more flexibility in implementation, scoring, reporting, etc. will create problems in standardization and comparability. The CCSS and the use of a single assessment are intended to reduce the current levels of confusion and non-comparability across states.

(b)

Given the vagueness of the theory of action (see above), it is difficult to determine whether the "theory" is being realized. That said, the proposal does list a number of evaluation steps/studies that will be useful in determining whether the intended effects on individuals and schools are being achieved.

Recommendations

SBAC should select and implement one growth model based on the best currently available research. Rather than procrastinate on the selection of a growth model, a model should be adopted, the assessment designed accordingly and growth scores reported with professionally acceptable caveats.

SBAC should standardize administration, scoring, and reporting across states.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	7

(A)(6) Reviewer Comments:

(a)

The proposal laudably indicates that SBAC's Professional Capacity and Outreach Working Group will collaborate with existing professional development networks in each state. However, the networks will not fully cooperate without direction from state and local leadership.

While the involvement of teachers in development and scoring can be a good thing, these activities reach a small part of the population of teachers and do little to support assessment literacy.

(b)

SBAC's communications plan appears to focus on providing audiences with information about the assessment system, its components, reports, etc., but appears to lack a focus on the benefits of a Comprehensive Assessment System: how it will improve teaching and learning, improve schools, etc. The plan does not sufficiently address building support for the system.

Recommendations

SBAC should request SEA and LEA leadership to support and encourage professional development networks to support the CCSS and the assessment system.

SBAC should develop a training plan with instructional modules that the local professional development networks can implement as needed. Specialized plans and support will be needed for teachers of SWDs and ELLs. These plans should be consistent with the test administration and score interpretation manuals provided by the assessment.

SBAC should include in its communications plan a priority on informing the public and parents about the benefits of the assessment system.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	4

(A)(7) Reviewer Comments:

(a)

The proposal notes that assessment system software will be developed using a combination of existing newly developed open-source and proprietary software. New software development tends to be open-ended in terms of cost and schedule. The proposal does not describe how SBAC intends to develop new software on-time and within budget, e.g., by thorough up-front definition of requirements and avoidance of scope-creep.

The description of the technology approach does not include a discussion of student identifiers – crucial to the calculation of growth scores – needed to automate registration, administration, scoring and reporting. The same applies to identification of teachers and principals, necessary for evaluating staff. Identifiers are essential for tracking and data correction.

The technology approach suggests that system software is to be distributed onto multiple state platforms, which would probably be unworkable. The proposal states that the system "will require minimal local IT staff involvement to install, operate, or update any software applications." Each state has a unique software, hardware, and system administration environment. A distributed approach would require a large system requirements study, which in turn would likely complicate and delay system development.

(b)

The proposal states that a Technology Implementation Task Force will utilize the collective wisdom of the Consortium and member states to address deployment issues. Few school districts possess the kind of experience needed to address these issues. Most school districts contract out work of this scope and complexity. Deployment across multiple states constitutes a quantum leap in complexity. A relatively centralized and standardized approach implemented by an experienced IT contractor seems more likely to succeed.

Recommendations

To the extent possible SBAC should rely more on existing technology and less on new software development.

The assessment system should provide a uniform and centralized system for identification and tracking of students, teachers, and principals.

The administration, scoring, and reporting of the assessment should be centralized onto one platform, not distributed across states.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	20

(A)(8) Reviewer Comments:

(a)

The ultimate project management partner is to be selected by an RFP (Appendix A8-2). The requirements and scope of work are appropriate. The interim project management partner (PMP) now under contract is fully qualified to perform necessary tasks until the PMP is selected. The amount of time that key interim PMP personnel are dedicated to the project is adequate. There is potential for selection of a new PMP. If this happens, the proposal states that the handoff to a new PMP will be "seamless." However, the proposal does not describe what precautions have been taken to assure a smooth transition (e.g., required timely delivery of documents, required staff cooperation, continuation of services during the transition period).

(b)

Consistent with the narrative in (A)(4)(a), the proposal indicates that some item writing will be distributed to the governing states. This approach risks not having enough usable items that meet technical and quality specifications to populate the tests.

The start dates that appear in the summary table for the project work plan and time line in many cases are unrealistic and do not take into account end-dates of prerequisite activities.

The summary table does not allow enough time for the development of RFPs and assumes that they can be quickly written and released. RFPs require detailed content, legal, policy, and fiscal review and revisions by the lead procurement state and by consortium governance before they are final.

The milestones/tasks for the test application do not appear to include necessary administrative tasks needed for accurate reporting, e.g. systems for identifying and tracking examinees and correcting demographic information.

The milestones/tasks allow less than a month for assessment administration training for the field test. Unless the field test is very small and very simple, more time will be needed.

The milestones/tasks allow one day for adoption of achievement standards. Preparation for standard setting takes months and it is unlikely that the actual standard setting can reasonably be conducted in one day. More time is needed.

The proposed common definitions and accommodation practices for SWDs and ELLs are commendable. States now have different criteria for defining "SWD" and "ELL" and have different accommodation practices. The variation across states in definitions and practices tends to reduce the comparability of assessment results. The proposal does not provide enough specificity or assurance for states' adoption of common definitions and practices.

(c)

The budget appears adequate to support the proposed work. The budget for management and governance is 6.7% of the requested total, which is not unreasonable.

A level 2 budget module is proposed in order to translate the mathematics assessment into 5 languages, including sign language. This is a worthy activity, and the funds appear to be adequate for translation. However, each translated test will require additional logistics for administration, and additional research will be needed to validate them. The proposal does not appear to recognize that significant additional work is needed after translation.

(d)

The proposal indicates that the ongoing testing with the new system will be competitive with the amount that states are already spending on their current programs. It is reasonably expected that states will allocate their current funding to the new system at the appropriate time.

Recommendations

SBAC should reconsider asking states to donate old assessment items. Uniform, consistent, centralized development of new items is preferred.

Regarding the plan to translate the tests into 5 languages, it would be preferable to begin with only Spanish and later expand to less frequently used languages. Spanish is the most common language used by ELLs. Starting with one language would reduce costs. The English language test is by itself an ambitious undertaking. The experience acquired by translating and administering a Spanish language version of the test could profitably be applied later to other translations. (Note that this is distinct from SBAC's plan to provide score reports in multiple languages, which is highly desirable, and should be expanded to include test administration directions and test interpretation materials.)

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	18
Competitive Reviewer Comments: (a) SBAC obtained appropriate letters of intent from 162 public IHE/IHE systems from 30 states in the consortium. The total number of direct matriculation students in the Consortium's participating IHE's represents 74 percent of the total of all such such students across all participating states. Note, however, that the commitments given by the executive leadership of the IHEs will be shaped by the faculty senates that deliberate on matters of admission and curriculum. (b) The letters are appropriately worded. (c) The letters were signed as required.		

Absolute Priority – Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.

	Available	Score
Absolute Priority - Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards.		Yes
Absolute Reviewer Comments: The lack of clarity regarding the capacity for evaluating teachers and principals is troubling, and the work required to build that capacity can be substantial, particularly if it must be developed later in the assessment development process. Moreover, SBAC's proposal is reticent on the topic of accountability. However, SBAC does state that it is merely waiting for specific requirements to emerge from the reauthorization of ESEA. SBAC indicates that it will respond as needed once the requirements are described in the law. While it would be more efficient to be proactive, it is likely that SBAC will be able to implement accountability as required when the time comes. Recommendations SBAC needs to clarify that the data produced by the assessment system can be used to evaluate teachers and principals (item (d)(ii)). The proposal does not make clear that teacher and principal identifiers can be collected and linked to student assessment results. A system for managing identifiers is needed to support tracking and evaluation of teachers and principals. The proposal does not provide a specific description of the vertical scale and growth model needed to measure student growth (item (a)(iv)), also needed to support evaluation of teachers and principals. Ideally, the system should be designed from the beginning to support specific methods of scaling and measurement of growth. Postponing these decisions could result in a mismatch between what the system provides and what is needed to support the growth model eventually selected.		

Grand Total	220	125
-------------	-----	-----

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

The budget detail under governance describes the time that Washington staff would dedicate to the project. A fiscal agent project director would dedicate 15% of time and a fiscal agent project manager would dedicate 50% of time. It may be unrealistic for the manager to be dedicated less than 100%. Consider, for example, that the federally funded state NAEP contacts are funded 100%. By comparison, this project will be more demanding than state NAEP.

The budget detail under assessment design provides \$1.5 million for an accommodation study that "will be procured through Washington's process if the associated Enhanced Assessment Grant is not awarded." If the EAG grant is awarded the budget should be reduced by \$1.5 million.

The budget detail under assessment development describes the costs of item development. The budget indicates that 25% of SR items and 15% of CR items will be supplied from existing State item banks at no cost. This may not be realistic.

Level 2 Budgets

Name: Assessment Design - Translations

A level 2 budget module is proposed in order to translate the mathematics assessment into 5 languages, including sign language. This is a worthy activity, and the funds appear to be adequate for translation.

However, it would be preferable to begin with only Spanish and later expand to less frequently used languages. Spanish is the most common language used by ELLs. Starting with one language would reduce costs. Validity studies would be needed for each translation. The English language test is by itself an ambitious undertaking. The experience acquired by translating and administering a Spanish language version of the test could profitably be applied later to other translations.