



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b)(6)

(A)(1) Consortium Governance

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

(A)(1) Reviewer Comments:

PARCC has laid out a comprehensive assessment reform agenda that covers the components required by this selection criteria. The application narrative and additional information included in the Appendices provide detailed responses to subsections i through v under Structure and Operations. The Consortium is comprised of 26 member states--11 of which are Governing and 15 Participating members. The role that any particular state plays is determined by the level of its formal commitment to the partnership and carries with it distinct rights and responsibilities. Governance will be carried out through a Board comprised of 15 chief state school officers or their designees from Governing States. A Generally Assembly, as a mechanism to promote sustained engagement, will meet 2 times per year bringing together all PARCC members along with 5 person leadership teams from each of the member states. A concern is that these organizational arrangements are not likely to produce participation that is representative of diverse stakeholder groups across the 26 states beyond the role Chief State School Officers. The application does not clearly distinguish between the kinds of decisions that will be distributed across the different governing and advisory entities. Additionally, given the significance of this undertaking, the proposed governing structures do not adequately address needed oversight at the operational level where a number of critical decisions will need to be made outside of the governing board quarterly meeting timeframe. The agreed upon super-majority voting process will create enhanced opportunities for a variety of voices to be heard in making important decisions. The Consortium is commended for its plan to work collaboratively with the Smarter Balanced Consortium to develop common level performance descriptors.

Florida will assume fiduciary responsibility in accordance with (A)(1)(b-vi) and for ensuring that the Consortium carries out the project in accordance with federal requirements including related administrative functions, subject to the direction and approval of the Governing Board regarding the expenditure and disbursement of all funds. In fulfilling the role of Fiscal agent, FL will have no greater decision making authority regarding the expenditure on disbursement of grant funds beyond that of a single Governing State. The Consortium's expectations for managing funds received under this grant category and issuing RFPs are described in each MOU under Section VII (Section A.2.i-viii/Membership Types and Responsibilities/Fiscal Agent).

MOUs submitted by each of the PARCC states adhere to a standard format along with addenda. Except for the allowable variance in membership roles (i. e. Lead Procurement, Governing and Participating) all state MOUs conform to the same terms and conditions. By signing off on MOUs, all 26 states have agreed to the Binding Commitments and Assurances outlined in X A-6 including periodic reviews of its State laws, regulations and policies to identify any barriers to implementing the proposed system and address any barriers prior to full implementation of the summative assessment. Barriers presented in the addenda provided by states vary in level of specificity, focus and inclusion of timelines. Some states have listed only barriers while others have addressed implementation of the full system (e. g. adoption of standards, engagement of stakeholder groups, field testing, removal of legislative and fiscal barriers). The application language asking states to address this issue is confusing. MOU section X A-6 speaks to barriers and the steps to be taken to overcome them. The specific title on the Addendum 2 page is ASSURANCE REGARDING PROCESS AND PLANS FOR IMPLEMENTING PROPOSED ASSESSMENT SYSTEM.

The range of barriers described in MOU/Addenda A include: available fiscal resources, translations in multiple languages, promulgating changes to existing codes and regulations, availability of district level technology infrastructure, and parental access to constructed response items.

Information submitted by the applicant satisfactorily addresses procurement (Addendum 3 to each state MOU) as required by this criterion. Assurance regarding participation in the PARCC procurement process has been validated by signatures on these addenda from chief procurement officials in each state. Consortium members' commitment to the described procurement process and details pertaining thereto are included in Appendices (A)(1)-A-ii through A-xxvii. The Lead Procurement State is Florida.

The application requests a total of \$2.7 million to support the governance structures, activities and travel for meetings across the 4 years of the grant period. The application also requests \$2 million for FL to assume the responsibilities of fiscal agent and \$7.3 million to provide each of the states with funds for in-state support (approximately \$117,000 per state during Year 1). The beginning budget calculations are based on the current 15 Governing States. PARCC anticipates an additional 10 Participating States will transition to Governing status by the time of assessment implementation. If such a change were to occur the current overlapping states would be required to withdraw membership participation from other consortia. There are several concerns regarding the resource allocations for Governance which are noted in the budget section of this review.

Recommendations

PARCC should:

- Provide guidelines to ensure balance and diverse stakeholder representation in the Governance Working Group/Committee formations. 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 governance structures (i.e. a membership slots on the Design, Advisory and Working Groups).
- Prepare a detailed Project Management Plan which delineates more specific tasks under each of the major plan components along with associated potential risks.
- Clarify confirm and synthesize potential barriers in each state and processes/timelines to address them.
- Explore more technology oriented meeting and communication mediums to reduce governance related costs

(A)(2) Theory of Action

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

(A)(2) Reviewer Comments:

PARCC has set forth a bold goal: building collective capacity to dramatically increase the rates at which students graduate from high school prepared for access to college and the workplace. The application provides descriptions and rationale which address criteria a-d for this selection area. The application focuses primarily on the summative assessment components of the system. The summative component will consist of a year end assessment in each subject area along with assessments distributed throughout the year—4 components in English Language Arts and 3 in mathematics. "Ready to Use" formative tools will also be developed under this grant by PARCC, as well as, within states.

Four essential features make up the proposed assessment system: Use college and career readiness as an anchor, measure rigorous content and students' ability to apply that content, measure learning throughout

the school year and leverage technology for innovation, cost efficiency and speed. PARCC states expect that the implementation and use of the proposed assessment system will produce 4 key impacts:

- Reporting results based on clear knowledge and skill definitions will provide students information throughout their education experiences so that deficiencies can be addressed early and promote greater alliance between the pre-collegiate and higher education systems
- Incorporating results into accountability policies and educator evaluations will drive teachers administrators, schools and students toward the goal of readiness
- Providing teachers with timely actionable assessment data will allow much greater instructional focus and adjustments where needed along with individualized interventions
- Making data from the assessment system accessible to parents, policy and the public will allow for cross state and international comparisons and hopefully make the case for reforms needed in other supporting policy areas

The PARCC assessment system is viewed as key lever that will transform today’s education system. The applicant acknowledges that an improved educational system requires additional support efforts and strategies including strong professional development, high quality classroom instruction, and the involvement of parents. The Theory of Action (TOA) is conceptually drafted based on important assumptions about teaching and learning but the follow-on components fail to include significant strategies specifically directed toward improved learning. The proposal for a new assessment system as articulated by the PARCC TOA assumes positive educational effects based on the high standards embodied in the Common Core Standards. The application outlines many potential benefits. The TOA, however, does not signal potential pockets of resistance that will need to be addressed or substantial challenges/obstacles that are likely to be encountered in critical implementation activities, e. g. reaching consensus on achievement standards, unanticipated consequences, availability of adequate funding for sustainability, creating opportunity to learn resources/standards, implications of college placement decision using the new assessments and technicalities associated with scoring and reporting.

(A)(3) Assessment System Design

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

(A)(3) Reviewer Comments:

The PARCC summative assessment system will consist of four components distributed throughout the year in ELA/literacy, three components distributed throughout the year in mathematics, and one end-of-year component in each subject area. Overall, the assessment system will include a mix of constructed-response items, performance tasks, and computer-enhanced and computer-scored items. This technology based "distributed" strategy for timing each assessment component will offer substantial advantages over traditional summative assessment systems. The Partnership will also initiate activities that will support teacher’s formative assessment needs both under the direct auspices of the partnership, e.g. development of model curriculum frameworks and exemplar lesson plans, and also leveraging tool development within states. This aspect of the overall assessment design is not addressed in detail and lacks operational details.

In its application PARCC provides a detailed description of the overall approach and specific steps that will be taken in each content area to ensure the full range and depth of standards are assessed. ELA/Literacy and Math will be administered in an on line computer based model in grades 6-11. For the earlier grades prompts will be delivered online but the student responses will be paper recorded. The ELA/literacy focus will include an extended research/writing assessment, a speaking and listening assessment and an end of year assessment. In the area of mathematics students will be presented with contextual problems and will be expected to apply mathematical reasoning skills. Other components will include assessing standards within each of mathematical domains and constructing modeling tasks—connecting classroom mathematics to

everyday life situations. The assessment content will be driven by the priorities set forth in the Common Core Standards and focus on measuring a few essential topics in depth. The final approach to high school assessment will require partnering with higher education. Due to the current disparities in public education, it is to be expected that in the beginning large numbers of students taking the tests will score at the lower end. Useful information about what all students know and can do is needed. PARCC therefore, will need to ensure for both subjects that in the assessment design students at the low end of the achievement scale are appropriately targeted and assessed with tests that produce accurate measures of performance. Also see A 4 discussion of the need to include high and low end students in the pilot and field tests.

The assessment system will produce student performance data in the form of achievement scores, information on college and career readiness, student growth measures and item analysis of released items. Three essential steps will be taken to ensure that the assessment system produces the required performance data to signal that students are college and career ready or on track to be college and career ready. PARCC will establish a set of performance level descriptors (PLDs) that clarify target performances for each subject area and at each grade level in collaboration with Smarter Balanced Consortium, identify a CCR performance target as an anchor, and articulate the achievement standards across all assessments in the system to establish meaningful on-track benchmarks for each grade level. The descriptions of performance and test items for reading should reflect the widely accepted interactive model of reading--higher-order and lower-order processes operating together. Higher--level processes (e.g. critical evaluation, making inferences) should be present at all levels from basic to advanced levels of achievement.

PARCC plans to go beyond the traditional standards setting methods. The Partnership will build on deep research-based knowledge and existing practices and at the same time generate input from the international and high education communities to create an innovative standard setting process. As noted in the application, the final decision regarding cut scores will be a "hefty one--filled with political and educational consequences." It will be important that both the Standard Setting Committee and the External Review Group involve a broad group of stakeholders that are inclusive and representative of the diverse student populations in the member states. In preparing for their decision, the Governing Board should develop a pro-active plan that responds to the analysis of student and institutional consequences with particular emphasis on certain populations or geographic regions where there is the most need for improvement.

Ensuring accessibility of the assessment system to all students is a foundational principle strongly embedded in the Partnership's TOA. Each component of the system will be designed in a manner that allows students with a wide range of cognitive abilities, content knowledge and English proficiency, to demonstrate what they know and can do. PARCC intends to address this challenge with a proactive, front-loaded plan. The Partnership's approach will be guided by the best thinking among experts and cutting-edge research. PARCC will also work with member states to ensure that IEP teams in local schools receive information about the common assessment components and training to apply common accommodations guidelines. Finally, the committee will work with the Research Strategy Group to establish protocols to determine whether the guidelines are being followed during field tests and full operational implementation.

The first three through-course components in ELA/literacy and mathematics will be administered after roughly 25 percent, 50 percent and 75 percent of instruction. The fourth through-course component in ELA/literacy, a speaking and listening assessment, will be administered after students complete the third component (described in more detail below). The end-of-year components in each subject area after roughly 90 percent of instruction.

As part of the professional development plan described in Section A 6, teachers will have timely and accessible data to differentiate and target instruction and fine tune their instructional plans throughout the year. Teachers will use the *Periodic Feedback Report* and the *Interactive Data Tool* to understand how their class and individual students are performing.

Summary Tables for A 3: English Language Arts and Mathematics include comprehensive descriptions by grade levels for the relevant components requested including:

- types of data produced/how used

- frequency and timing of administration
- number and types of items and distribution of item types
- administration mode
- scoring method and estimated turnaround time
- reports that will be produced and intended use

The application does not raise any issues about the potential misuse of assessment data, e.g. as a single measure for high stakes decisions, retention or college level course placement.

PARCC has presented a strong conceptual plan addressing major design features of the new assessment system but with causes for concern noted above.

Recommendations:

- A solid management plan will be needed to ensure the work is properly sequenced, properly resourced and stays on track.
- The PARCC Governing Board should develop explicit guidelines and practices for use of data from the new assessment system and, in particular, relative to high-stakes decisions about teachers, classrooms, and schools.
- A more detailed plan for the formative assessment component should be provided
- Test design specifications should ensure low-end of the achievement scale items and tasks are included
- Membership on the Standard Setting Committee and the External Review Committee should be broadened to include local level representatives
- The PARCC Governing Board should develop a pro-active plan created simultaneously with the standard setting process and that responds to the research analysis of student and institutional consequences. A special advisory structure should be formed to work with the Board in formulating a set of recommendations.

(A)(4) Assessment System Development

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

(A)(4) Reviewer Comments:

An **evidence-centered design** (ECD) approach will be used to develop the system of summative assessments while **universal design** methods will be applied within each component of the system. In designing the full system, PARCC will draw evidence from several sources and assessment types to more fully understand a student's knowledge and skills. The ECD and universal design approaches are substantiated in research conducted by Messick, Levy, Steinberg, Almond, Thurlow et al. The development processes are closely aligned with these two approaches and follow standard test development procedures--content analysis, development of test blueprints, selection of a test vendor, item development and review, pilot and field testing, and data and alignment reviews. Subject matter experts will work with assessment and technology experts to develop test item templates and models of each item type to be developed. The selected vendor(s) will be responsible for actual item development and tasks based on the established *Blueprint*. Once the items are developed they will undergo intense scrutiny by item review committees and will also be subject to bias and sensitivity protocols. A number of committees, panels, work teams and advisory structures will be established to ensure the quality and fairness of the assessment. Membership will be drawn from a diverse group of stakeholders across PARCC states including assessment and subject matter content experts, higher education representatives, business and community leaders and experts in accessibility, accommodations, cultural differences, and socio-economic issues. The Partnership has assured that care will be taken to ensure that all groups have adequate representation. The application does not address possible conflicts and confusion that could arise from efforts directed at national and local stakeholders involved in multiple assessment consortia.

The application lays out an extensive and detailed process for designing and developing accommodations which is grounded in cutting edge research and based on the deep experience of member states. A special committee on Accessibility and Accommodations will lead and guide efforts in this area. The expectation is that PARCC members will adopt a common set of accommodations and accommodation guidelines for students with disabilities and English language learners. PARCC plans to launch a range of pilot tests and research studies covering both sub-groups to investigate whether the accommodation enables valid results for different populations and implementation challenges that may need to be addressed.

Additional elaboration is needed to indicate the types of research questions that need to be explored. Given the scarcity of research available on accommodations for English language learners, this area of research should receive high priority on the list of the targeted studies. In reviewing the application narrative under this selection criteria and the Research section, it is not clear where expenditures have been included or the level of resource allocation for the accommodations related research studies. The application does not address staff assignment, needed qualifications or contractual arrangements that will lead or facilitate work in this important area. PARCC's proposal also does not speak to language translations of the summative assessment components and yet this appears as an important area for at least for one of the largest Governing States.

PARCC envisions using a combination of computer and distributed human scoring. The end-of-year component in ELA and mathematics, by design, will be 100 percent computer-scored. To ensure accuracy and reliability for the through-course components, PARCC will implement an approach that includes double scoring and read behinds. The approach includes a built-in process for ensuring the continued reliability of both AI and human scoring. The goal is to provide a mechanism for teacher engagement in the scoring process while allowing for state flexibility in the level of teacher involvement. States will have the option to determine whether teachers will be formally involved in the scoring process, the number, involved and whether incentives will be provided for participation. States may choose to use 100 percent vendor scoring while other states may prefer to include greater participation of their own teachers in the scoring process.

As stated in the application, the reporting system will include a vendor data hub, Partnership and state data systems along with Partnership level and state level reporting systems. PARCC will develop or leverage existing platforms for a centralized *Interactive Data Tool* and reporting platform/system that will work in tandem with current state longitudinal data systems. In order to better track students and teachers through the system and allow data to follow individuals across member states, a unique identifier for each student and teacher will likely be required. A set of common identity management and data technology standards will be established for member states. National experts will provide guidance on platform design, the pros

and cons of development versus acquisition, and how to capture and compare data across states. States will provide the designated vendor with student demographic information to create individualized student access and assessment forms. Once a student has taken an assessment component, the data will funnel through the vendor and be made available to stakeholders, as appropriate, through the *Interactive Data Tool*.

The Partnership is committed to ensuring quality in all aspects of the assessment system and the processes leading to its development. To this end, several key strategies and activities will drive the work:

- A Technical Advisory Committee will oversee the work ensuring the highest standards of quality are adhered to in designing, developing and implementing the new assessment system.
- Key advisory committees comprised of technical and psychometric experts along with educators will conduct item, media and passage review to ensure content accuracy, alignment to standards and grade level appropriateness; determine item suitability for use and review materials for bias or sensitivity protocols issues.
- Findings from the committee reviews will be used to inform ongoing item development.
- Recommendations which address 8 areas of test development processes issued by the Council of Chief State School Officers' Large Scale Assessment Group will be implemented as standard procedures.
- A significant research agenda will also support continuous improvement and validity associated with key development constructs

The Partnership's approach to field testing will involve 2 phases. Early access to some assessment items will be made available as early as 2011-12 in order to provide opportunities for practice with the new item types. Formal field testing will begin during school year 2012-13 and continue through 2013-14. The Partnership will use a stratified matrix sampling approach for both the pilot and field testing. The large size and diversity of the student sample (3,500 students in the field test) will enable PARCC to conduct robust studies of various accommodations as there will be sufficiently large samples of special subpopulations of students with disabilities and English learners. The field tests will be structured to ensure that sufficient samples of students with specific disabilities and levels of English proficiency are included. The application does not address the selection criterion with respect to representative sampling of high-low end regular students. The special population category should be expanded to include representative numbers of high and low end students in the pilot/field tests.

Recommendations

PARCC should:

- Elaborate and prioritize specific research questions to be explored in the areas of appropriate accommodations for English language learners
- Clarify adequacy of staff assignment for the research work
- Include high-end and low end students in sample for pilot and field tests

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	20
(A)(5) Reviewer Comments:		

The research studies outlined in the application fully comply with the research classifications established by the NIA under this selection criterion and will address construct, consequential, and predictive validity, external validity, reliability, fairness, precision and comparability. One of the most significant aspects, and likely the most controversial, of the assessment system will be the establishment of achievement levels and cut scores that are internationally benchmarked. PARCC's efforts, therefore, to support the proposed non-traditional standard setting process with research studies to verify the content and consequential validity of the assessment system are deemed to be critical. PARCC has also presented a conscientious plan for evaluating intended impact and effect. While much of this research will need to await the final implementation of the system after the grant period, the plan calls for gathering early evidence on the social, operational, fiscal and technical impacts of designing, developing and implementing the proposed assessment system. Additional research questions proposed include: Is the system improving instruction as intended? What impact is the system having on student achievement in general and on rates of student readiness for college and careers? All of these areas are judged to be important to establishing credibility of the system over the long term.

The plan would benefit from the delineation of clear time frames describing when each of the studies would be initiated along with projected completion. Both the research and project management budgets include expenditures for research. It is not clear from the application and budget what specific research would be undertaken during the grant period versus after 2014-15 nor how funding will be generated to cover the full range of the research topics some of which will require longitudinal data collection. The proposed collaborative research study with other consortia and listed as #17 in the Research and Evaluation budget @\$750,000 needs further explanation and elaboration. It is also not clear that the Project Management plan includes sufficient staff to both support the various advisory groups and coordinate PARCC's substantial research agenda even though much of this work will be assigned to external contractors. According to Summary Table for (A)(8)(a)Key Project Management Personnel, only a 25% FTE from existing ACHIEVE staff has been assigned responsibility for coordination of PARCC research studies.

Recommendations:

PARCC should:

- Prioritize the series of papers, studies and literature reviews presented in its application responding to this selection criteria and a specific action action plan for how this would actually roll out.
- Ensure that all plan components have an evaluation strategy built into their operational plans. Two components stand to profit from more explicit evaluation inclusions: Professional Capacity and Outreach.
- Give high priority to research targeted to ELL students, the articulation of the potential educational effects of the new assessment system and the development of a program of research and evaluation to learn whether and how the system when actually implemented actually contributes to improved educational outcomes.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	8

(A)(6) Reviewer Comments:

The primary focus of PARCC's plan is directed towards helping educators make use of the assessments through released items for instructional purposes and developing educator knowledge gained in assessment

administration and scoring. PARCC's plan for supporting teachers and administrators includes 4 key components:

- Building a leadership cadre of content experts
- Developing on line training tools to help implement the system
- Developing tools to interpret and use assessment results
- Sharing advice to help educator understand and address the curricular and instructional implications of the assessment results

PARCC will develop the specific tools and strategies under each of the components but will not supply the training directly. Instead PARCC will work within member states on a strategy to leverage existing funding streams for training and support once the modules have been developed. PARCC will also build model curriculum frameworks, share exemplar lesson plans and student work and conduct gap analysis between state standards and the Common Core Standards.

Projected expenditures for content experts (\$3.5 million), the professional development coursework (\$2 million), expert standards writers and consultants to help analyze the Common Core Standards (\$1.2 million), model curriculum frameworks (\$500,000) are spread across 2 budget modules—Professional Capacity and Project Management. Budget specifications and the application narrative should align.

This aspect of the application targeted to building professional capacity is the weakest among all the application components. It is substantially underdeveloped conceptually and operationally. While teachers will play a large role in the assessment design and development components, it cannot be assumed that this experience will necessarily be automatically transitioned as the proposal states "to model good instruction and support professional development". PARCC plans to develop a set of "assessment course" products that will be distributed electronically. The key concepts embedded in these courses focus on helping educators understand the new assessments, how to administer the various components and interpret results but fall short in addressing how these assessments will be used to drive improved teaching and learning. Additional tools related to the Common Core Standards and the common assessments will be produced at the individual state level. PARCC indicates in its application that it will help states conduct gap analyses related to the Common Core Standards, create curriculum frameworks and share exemplary lesson plans and student work. The application does not sufficiently describe how what is proposed will be implemented or how key challenges will be addressed, e.g. working with the states to ensure delivery of the training, production of curriculum modules, creation of the online courses, ensuring widespread use of the assessment tools, collaboration with states to address broader state policies, etc. There needs to be stronger justification for work and allocation of RttT Assessment resources proposed in the areas of standards and curriculum. The PARCC effort should be directed at leveraging funding sources and collaborative approaches with other organizations. It is also not clear who among existing or new ACHIEVE personnel will provide direct leadership to this aspect of PARCC's work and the percentage of dedicated staff time. Mechanisms for state guidance, engagement and oversight appear to be missing. For example, specific membership structures are not described that will oversee this work such as a working group or advisory committee. A much more detailed operational plan addressing professional capacity is needed to clarify objectives, timelines and responsible parties.

The application's plans for formative assessment are sketchy. On line assistance and tools will be provided to teachers on re-use of released items and performance tasks (mentioned in Assessment Development section). Aside from this on line training, formative classroom assessment does not appear to be a key feature of the PARCC application for the grades 3-high school component, although the case advanced in budget 2 module (Priority #1) strongly argues for an additional \$10 million to develop K-2 formative assessments in literacy and math. It is not clear the extent to which this area will be addressed as part of the effort to work and fund assessment supports in each of the states.

PARCC has developed a comprehensive Public Outreach and Engagement Plan. The plan is well conceived, for the most part, and designed around a phased and multi-tiered strategy targeted to multiple stakeholders. The purposes of this effort will be to help the various audiences understand the new

assessment system, to address early on any real or perceived misunderstandings and to build broad based support for the Consortium's work. The PARCC communications strategy will be carried out at both the national and state levels. Oversight will be provided by two committee mechanisms: a national Communications Committee comprised of at least one representative for each state in state communications teams. Major communications and outreach activities will include: development and dissemination of tools and strategies to inform the public, targeted coalition building in working with state teams, a college ready outreach for students and their families and a higher education engagement strategy involving college and university leaders.

Given the history of inequality in the nation's public schools, and the fact that substantial efforts will be needed to level the playing field, a major concern is that the plan underestimates the impact and possible consequences of the proposed new assessments on the most vulnerable children. There is much hope that the new assessments will stimulate action in other arenas of needed educational reform, e. g. teacher quality and distribution, resource allocations, opportunity to learn enhancements, etc. The application emphasizes that, while having widely recognized standards and assessments can provide focus for planning and a common language for students, parents, and teachers, what is most important is what parents, students, and educators actually do with that knowledge. Substantial dialogue will be needed, to articulate the range of ways in which the information provided by PARCC might or might not lead to positive changes. Outreach strategies will need to be developed which address the believers and the critics, those who will argue that the new national tests will simply perpetuate the beliefs about failing students and might not leverage any improvement in educational processes.

Recommendations

In the area of Professional Capacity, PARCC should redesign its professional support strategy to focus on how assessments will be used to support improved teaching and learning and develop a detailed operational plan to support this concept, including goals, objectives, target audiences, measures of effectiveness, resource needs and personnel assignments.

The outreach plan could benefit from the following additional considerations:

- Further elaboration of the coalition building effort directed to generating input into the assessment design and development process and ensuring a two way communication over the 4 year period
- Strategies for engaging K-12 organizational partners (similar to the outreach planned for higher education) such as the Council of Great City Schools, National Association of Black Educators, National Council of La Raza, Education Trust, Association for Curriculum and Development, National Urban Alliance for Effective Education, associations of elementary and secondary principals, etc.
- Stronger assurances of diversity in the outreach plan components, e.g. inclusion of community and faith based organizations, local level constituents, representatives with knowledge and experience in ELL, SWD and with traditionally under-served student populations, materials and tools developed for non-English speakers.
- Connections to the research described in Section A 5 on the implementation and implications of using the results of assessments, e.g. cut scores on achievement standards, promotion policies, projected accountability impacts and links to teacher evaluation.
- Collaboration with other state and federal grant initiatives with special attention to constituents and leaders involved in RtT Intervening in Low Achieving Schools program component and the state led TurnAround initiatives.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	8
(A)(7) Reviewer Comments:		
<p>PARCC has laid out an ambitious plan for technology development and application. The partnership will work with one or more technology partners and member states to create an open source integrated system to host, manage and supply data in a timely manner based on stakeholders' interests and needs. The technology platform will also support assessment administration and collection of student responses. According to PARCC's application "the system will be hosted through a cloud based, distributed architecture to create efficiencies in cost, reliable data retrieval, deep capture of historical tracking, and timely non-disruptive backups." The data will be maintained in a format that can be uploaded into each state's data warehouse system as needed. Required technology at the local level will be an Internet connection and standard Web browser. Initial specifications for an online assessment system are outlined in the application. A Technology Design Committee, created by the PARCC Governing Board will guide the technology work and provide further specifications for the assessment delivery and reporting platform. To support its vision of interoperability, the Partnership will work with member states, USED and national foundations focused on developing an education technology platform.</p> <p>PARCC and its members acknowledge the technology challenges associated with administering all assessments components in all states by computer. The Partnership believes that the advantages in moving in this direction far outweigh the challenges. Specific advantages include quick and cost effective scoring and use of innovative items only available in computer format. The Consortium plans to work with states which in turn will help their local districts transition to computer based testing. The initial list of barriers covers a wide range of topics including disparities in technology infrastructure, security of testing materials, destruction of infrastructure caused by unexpected natural disasters, multiple computer devices, and accommodations for students with disabilities. To assist states and local districts, PARCC will identify sufficiently early online system specifications and minimum school and district technology requirements. States will design locally appropriate gap analyses and based on the results states and schools will develop collaborative activities and funding solutions to address high priority needs. PARCC will also collaborate with existing national organizations who are addressing gap closures in technology infrastructure and resources. Additional strategies will include: a site certification process to ensure that schools are ready for implementation by 2014-15, publication of necessary student, teacher and administrator data required by the assessment system, making available the paper and pencil accommodation form in case of technology destruction, non-technology based communications, and developing other solutions in concert with specific advisory groups. Overall the plan is judged to be of reasonable quality, however, a major concern resides in the uncertainty that current local technology disparities will be adequately addressed and that all students will have capacity (knowledge, skills and familiarity with operating systems) to successfully access the online assessment system. The level of resources matched to the operational plan in the area of technology do not seem aligned.</p>		

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	21
(A)(8) Reviewer Comments:		
<p>PARCC has already conducted a competitive bidding process and chosen ACHIEVE as its project management partner. ACHIEVE's mission is to provide national leadership on academic standards, assessments and accountability to help states prepare all young people for post-secondary education, work</p>		

and citizenship. This mission is closely aligned with the RtT Assessment grant purpose as defined in the NIA and PARCC Theory of Action. The organization has been in existence since 1966, is comprised of 33 staff members and based in Washington, DC. ACHIEVE has helped shape education standards and assessment practices and policies in a large number of states. It also has considerable experience working in higher education. ACHIEVE's track record of success includes leadership in the design and management of several multi-state consortia: The American Diploma Project Network, the ADP Assessment Consortia, the College and Career Policy Institute and the Common Core Standards. Given this history ACHIEVE is well positioned to launch the new assessment consortium. Staff vitae included in Appendix (A)(8)- B describe a highly talented staff with rich experience and strong credentials in standards and assessment. Nine existing staff will be assigned to work on the PARCC initiative ranging in time commitments from 20%--100%. As ACHIEVE president, Michael Cohen will provide overall leadership in support of PARCC at no cost. Only a 20% staff time allocation is dedicated for the functions described as "assessment design and leadership for PARCC management, state policy integration, communications, outreach and sustainability. This minimal level of staff time does not seem to match the level of work outlined in the application. The application requests funding for the employment of 11 new FTEs including a PARCC director. This number includes 4 additional staff positions beyond what is described in the application narrative under Project Management (Senior Project Associate/Assessment, Project Associate/Assessment, Meeting Planner and Project Associate /Project Management). A substantial amount of the budget is directed towards governance and staff expenses at a disproportionate level compared to the the rest of the plan components.

The application does not speak to the process that will take place to recruit and hire the FTE new staff nor are specific qualifications (project management certification, knowledge and experience in ELL or SWD, teaching high end and low end students) or job descriptions included. It is not clear the extent to which ACHIEVE employment policies embrace specific diversity principles so as to ensure that PARCC staff will be representative of the diverse students residing in the 25 consortium member states. While organizational structure is mentioned in the application, an organization chart of personnel and reporting arrangements is not included. Of favorable note is the staff assignment on the personnel list to address integration of the new assessment system into broader state policies. PARCC acknowledges in its TOA the need for accompanying supports aligned to the assessment system along with local effort skill and persistence, thus, its plan to support a broader set of state reform practices will help achieve the intended outcomes.

Summary Table (A)(8)-B: Project Workplan and Timeline describes key project milestones, deadlines and responsible entities as required by this selection criterion. ACHIEVE has presented a preliminary Management Plan with a promise for finer level details once the grant is awarded. The application describes its general approach to risk management as one which will include progress monitoring and making mid-course corrections/adjustments as necessary. In-kind services of Battelle Institute have been secured to provide program management guidance. The application does not address how risks will be explicitly identified nor specific progress/effectiveness monitoring techniques that will be applied. While some aspects are addressed under the Research criterion, the plan could benefit from a more serious consideration of an overall evaluation strategy.

The PARCC plan consists of nine level 1 budget modules (Governance, Fiscal Agent, Support for Governing States, Assessment Design and Development, Research and Evaluation, Professional Capacity/Outreach, Technology, Higher Education Engagement, and Project Management) and two level 2 budget modules to develop formative assessment tools for grades K-2 and technology enhancements. The overall expenditure projection totals \$169,990,272—a price tag that seems more than adequate to fully fund the new assessments along with various governing and research related activities that surround the system. It is to be expected that more precise expenditure details will be available based on actual contractual engagements and a more specific project management plan. There are some resource allocations, however, that surpass what is articulated or justified in the narrative, e.g. costs associated with multiple state representatives involved in governance meetings, discretionary grants to governing states, the rationale for 5 FTE fiscal staff.

Of the total requested \$19.9 Million is for the Budget Module 2 components. The estimated budget for the K-2 assessments is just under \$10 million. The argument for K-2 formative assessment resources would appear to be a logical extension of the 3-8 and high school design continuum. Given the controversies in the early childhood education field, assessment initiatives in this direction should present a more forceful documentation of need based justification and commitments of support from the field of early childhood including educators and parents. The optional feature of the early grade assessments raises a serious concern in terms of equity and local school district resource capacity. The budget rationale is absent any strategy for how these early grade assessments would be paid for once developed. The request for needed technology enhancements is to: maximize the number of students who will be able to take the assessment components at the same time; accelerate the timeline for scoring of through-course tasks; and expanding capabilities of AI software and improving in computer based test delivery. The budget narrative notes that having an additional \$10 million would create both flexibility and redeployment opportunities. There is much to be learned once the development works gets underway and translates to practical applications. Without more specific experience in actually building the basic technology platform, the technology enhancement request is judged to be premature.

The plan projects operational costs for the summative assessment components in the range of \$17-\$18 per student on the low end to a high of \$45-\$50 per student per year. Based on the scoring model presented in the application, the operational costs are expected to be approximately \$700 million overall and \$32.68 per student. According to the application, this figure is judged to be equal or less than the per student costs of member state's current assessment systems, however, there is no state assessment expenditure data to validate this claim. It is not clear whether the projected costs include the additional costs associated with incorporating accommodations into computer-based testing. The states in their MOU Addenda have not uniformly or explicitly addressed a plan for securing the funds necessary to implementation. In fact several states in their Addendum 2 submissions have identified funding as a major barrier to implementing the new assessment. Other states have indicated an expectation for absorption without additional revenues based on comparable costs for the operational PARCC assessments to their current state program. Financing considerations in the application largely focus on assessment design, development, and start up with a transition to state budgets in 2014-15. The application speaks generally to the need for research and development directed to scalability and sustainability but does not deal substantively with long term issues beyond 2014-15. While there is much work to be done to get to that point, it is not until after the grant that perhaps the most significant step in the assessment development process occurs--the setting of achievement cut scores. The criticality and concerns with this process are covered in A 3.

Recommendations

PARCC should:

- Describe staff hiring process and specific staffing qualifications/needs in the areas of project management skills, diversity, research leadership coordination and leadership of the professional capacity component
- Provide organizational chart and personnel reporting arrangements
- Identify risks, elaborate on barriers and develop plans for addressing them
- Create overall evaluation strategy covering all components in the application
- Address long term maintenance of assessment system and sustainability

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:

PARCC states it has received Letters of Intent (LOIs) from 188 public and private institutions and systems of higher education across 26 states. Based on the Summary Table for Competitive Preference there are 182 public IHE/IHE system partners. Appendix IHE –B-I includes letters from 16 private institutions in AL (4), MD(4) and NJ(8). PARCC should confirm the number of higher education partners who have committed to the new assessment system. Several of the IHE/IHE systems that have signed on to PARCC have also pledged involvement to the SBAC group. Neither applications nor the LOIs address this overlap.

For the most part the Letters of Intent follow a standard format. Collaborating IHEs or IHE systems, with the exception of Western Illinois University, have documented in their respective LOIs:

- 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 the assessments measure college readiness
- 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
- 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.

Note: The LOI from Western Illinois University does not comply with required content, signatures, numbers or commitments specified in the NIA.

Strengths of the proposed collaboration between PARCC and higher education include:

- A substantial number of direct matriculation students in public IHEs will be impacted by the new assessment system
- Additional student impact will be achieved as a result of the participation of 16 private IHEs
- A thoughtful strategy for structuring the engaging a broad group of higher education representatives within states
- Dedicated personnel budgeted and assigned to work with the various advisory and working groups comprised of representatives from the higher education community
- Recognition of supportive policy challenges (e.g. bridge courses, curriculum alignment, targeted college readiness supports) that are required to ensure that the new assessment does in fact serve as leverage to improve all student’s readiness for college.

Recommendations

PARCC should:

- Confirm that the definition used in its application is consistent with the NIA definition of “direct matriculation” students and used consistently for all institutions and IHE systems. Fifty of the LOIs submitted did not include the number of direct matriculation students as required by this selection criterion. The application cites three different sources that it drew on to generate the data for this selection criteria: individual systems and institutions in their LOIs, statewide data systems and IPEDS for NJ, NH and ND. The IPEDS data only covers the time period for high school graduation for 12 months as opposed to the 24 months included in the NIA definition.
- Clarify the number discrepancies for selected institutions in the states of RI, ND, IN, AZ, NH, MD, MA between what is reported in LOIs versus the Summary Table

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:

PARCC has created an ambitious proposal which is consistent with the RtT Assessment intent. The application addresses each of the four priorities specified under the Absolute Priority along with the associated 10 sub-criterion. The key features of this application build on the strong track record of ACHIEVE, the selected Project Management organization, and its experience working with multiple states in the areas of standards and assessment reform. PARCC will also make a special effort to collaborate with the Smarter Balanced Consortium. The commitment from participating SEAs involves 26 partner states with 11 serving as Governing members and 15 functioning as Participating members. More than 180 Letters of Intent from colleges and universities serves as strong evidence of support and commitment from the higher education community. The budget presents an overall and project level fiscal plan for allocating RtT Assessment resources. A particularly laudable component of the PARCC plan is its interest in ongoing and follow-up research to substantiate impact and effect consistent with its TOA and beyond the time frame of the first operational implementation of the assessment system. The proposal is viewed as a starting point for embarking on a plan that will undoubtedly require further definition, elaboration, research and critical decision making. While not all components of the Absolute Priority are addressed at the same level of detail, the over all application presents a strong plan for advancing assessment reform in a powerful way.

Grand Total	220	165
--------------------	------------	------------

Budgets

Level 1 Budget
<p>Name: Level 1 Budget(s)</p> <p>PARCC has requested an overall total of \$169,990,272 to implement its proposed assessment system. The budget is presented in summary format and also breakouts by specific budget categories such as personnel, fringe benefits, travel, etc. The application consists of 9 separate Level 1 budget modules for a total of \$149,994,386. Level 1 modules conform with the format outlined in the NIA and each budget module includes a narrative along with a list of activities/deliverables and associated costs. The current expenditure proposal does not exceed the prescribed RtT \$150 million cap. The applicant gives some attention (Technology and Research) 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 overall expenditure projection seems adequate for the core system along with various governing and research related activities that surround the system. More precise expenditure details will be available based on actual contractual engagements and a more specific project management plan. There are several concerns raised by the application and budget modules that exceed what is presented in the application narrative or require further justification and elaboration:</p> <ul style="list-style-type: none"> • Costs associated with multiple state representatives involved in governance meetings (See Project Management Section)

- Support for 5 full time employees assigned to the area of fiscal responsibility (1 FTE at ACHIEVE and 4 FTEs at FL Department of Education @ \$70,000 each over 4 years)
- Support for higher education faculty involvement. The application, does not sufficiently describe what will be implemented through use of 4 higher education experts (Two for years 1-2 and 4 for Years 3-4@ 10 days per month) along with site visits to states. Expenditures are included in this budget module a \$2 million.
- Need for break out of expected meeting attendees versus expected travelers for governance and general assembly sessions, as well as the rationale for budget assumptions based on: a) 30 member states participating in General Assembly meetings during years 1-2 when there are only 25 current PARCC members and b) Costs and rationale for additional personnel to attend Governing Board and General Assembly meetings for Chief School Officers. The application narrative describes the Governing Committee makeup to consist of one representative per state.
- Distribution procedures for discretionary sub-grants to states and the proposed level of expenditure. The application indicates in the budget module that 15-20 sub-grants (\$60,000 - \$80,000 per state) will be made to Governing States. There are only 15 Governing States in the consortium which is comprised 26 members overall.
- Disproportionate amount (over resourcing) of the budget assigned to the areas of personnel, administration and governance (See Project Management Section)
- Underestimation of costs to implement quality activities in the areas of AI scoring and professional capacity
- Research project #17 (A set of Collaborative studies with other consortia to compare the results of student performance) budgeted @ \$750,000 (See Research Section)

Level 2 Budgets

Name: Formative Assessment Tools for K-2

Of the total requested \$19.9 Million is for the Budget Module 2 components. The estimated budget for the K-2 assessments is just under \$10 million. The argument for K-2 formative assessment resources would appear to be a logical extension of the 3-8 and high school design continuum. Given the controversies in the early childhood education field, assessment initiatives in this direction need more evidence based justification and support from the field of early childhood educators and parents. Formal indications of support from member states advancing this assessment component were not included in the application. While as part of its rationale, the application references information published by the National Association for the Education of Young Children (NAEYC), letters advancing support for this assessment component from the major organizations active in the field of early childhood education (e.g. Black Child Development Institute, NAEYC, etc.) were not included. The optional feature of the early grade assessments raises a deep concern in terms of equity and local resource capacity. The budget rationale is absent any strategy for how these early grade assessments would be paid for once developed. The application also does not address a strategy to help teachers use the results of the proposed K-2 assessments in ways that would substantially augment learning toward the goal of having all students prepared for 3 grade.

Name: Technology Enhancements

An additional Budget 2 Module requests \$10 million to create flexibility and redeployment opportunities in the area of technology. Without more specific information and experience in actually building the basic technology platform, the technology enhancement request is judged to be premature.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b)
(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	9
<p>(A)(1) Reviewer Comments:</p> <p>(a) The consortium's vision...</p> <p>The proposal suggests wrongly, however, that a scoring platform and moderation system will "ensure" valid and reliable assessments. Valid and reliable interpretations of test scores require more than a scoring platform and moderation system. Serious investigations and research projects lead to reliable and valid interpretations of the test scores. A validity argument is based on collecting evidence about the claims made for the scores. Those claims are justified through research findings. The language in this section should be aligned with that of the Research and Evaluation section where some of the issues of valid and reliable interpretations are investigated.</p> <p>(b)(c)(d) Structure and operations...</p> <p>The proposal gives states various roles and provided flexibility for changing those roles. It is responsive to the various criteria for such things as differentiated roles, ways to make decisions, etc.</p> <p>Governance, etc.</p> <p>PARCC places major responsibilities on an Executive Committee composed only of Chief State School officers. It allows, however, an Assessment Design committee, and other not defined committees, to make decisions apart from the Executive Committee. The proposal uses test specifications, for example, as an area where decisions can be made about the assessment without resorting to the Executive Committee. There is no clear delineation of what is or is not an issue that design committees rather than the Executive Committee can make. It is unclear how the various responsibilities will be delegated.</p> <p>Recommendation: The National Assessment Governing Board's approach to overseeing an assessment is a better model. Participants on the board include not only those who might have primarily interest in policy issues but also those with technical expertise. In addition there are other constituencies represented as voting members of the board. This broad representation is a much better way to insure responsible decisions.</p> <p>Recommendation: This assessment is highly technical and relies in some areas on assumptions about what will develop in the future. There needs to be someone who is sensitive to these technical issues at the table when big decisions are being made. As presently construed it is difficult to know where in the structure these decisions will be made, when, and by whom.</p>		

(A)(2) Theory of Action

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

(A)(2) Reviewer Comments:

The description of the components of the plan is clear. How those components will be implemented is less clear. Using college and career readiness as an anchor may not be such an easy thing to do. Readiness is a fuzzy construct. However defined it may or may not be the same at every grade level. For example, how to operationalize readiness in the early grades is particularly problematic since students develop at different rates and have a longer time "to change." It is a bit like taking grades in third grade and using them to predict performance in the first year of college. Perhaps they are related; perhaps not. But, under the best of conditions one would hope that each student would thrive and in so doing defy a prediction.

The notion of college and career readiness is referred to throughout but is never fully defined and explicated. How such ideas are to be implemented depends first of a good definition of what it is.

The through-course assessments make strong assumptions about commonality of experiences for students. This may be a bigger issue in mathematics than in English language. In any case, they assume that each state, district, school has a matching curriculum in ways that used to be called the scope and sequence of the curriculum. It is assumed that topics are introduced at the same time and in the same sequence so the assessment results can have a common meaning. Without a common curriculum, then, interpretations of the results are problematic. That is, they are invalid for students who have not been exposed to the material being assessed.

(A)(3) Assessment System Design

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

(A)(3) Reviewer Comments:

The PARCC proposal contains interesting ideas and emphases. It aligns itself closely with the new standards. And it is responsive to the criteria.

(b)(i) Even though this section is very long, the presentation lacks important detail. For example in discussing its "distributed" strategy it says "Tasks and rubrics will be designed in a way to capture and support the full range of student skills..." It would be desirable to have an example or examples in the proposal of what those tasks are and what that assessment looks like. Previous experiences with performance assessments, for instance, have not had the property of one performance opportunity producing results across the full achievement scale. It is doubtful that several of them can get accurate measurement across the full range of proficiencies...

The proposal contains detailed information about literacy and mathematics assessments. The variety of the assessment types is well-documented. An emphasis on releasing items seems good, too. Yet, often the presentation is not sufficiently in depth to allow the reader to know exactly what will be implemented and how. And there appears that it will take a substantial amount of time (the amount is not specified in the proposal) to assess all that is to be assessed.

The proposal suggests releasing items as one way to inform and guide instruction. It does not provide a model for releasing the items.

(b)(iii) The proposal's section on accommodations (and as they are put in place as detailed in the next section of the proposal) is exemplary. The major issues are defined, one by one, and it is specific about how to address those issues. Other parts of the proposal lack such specificity.

(c) (ii) (iii) Individual items play a big role in the proposal and especially in the through-course assessments. Several times it is said that three items will be used for each of two topics to make inferences about whether or not a student is "on track." The proposal states that those measurements are going to be after having completed about 25, 50 and 75% of a course. There are huge assumptions being made about how the assessments can be constructed and the utility of the results. The major assumption

is that students will have been exposed to the same material or experiences reflected in the standards in each classroom in each state. The mathematics standards, for instance contain five strands. In order for the assessment results to be interpreted in defensible ways, it is assumed that the strands will be introduced at the same time in the same sequence for the same amount of time.

In addition, the through-course assessment would contain in theory three items on each of two topics. But three items are unlikely to produce results that can be reliably used to determine whether or not a student is "on track." The through-course assessments assume common content and common sequencing but only have common standards to work with. For the through-course assessments to have a chance of being effective there should be a semblance of a common curriculum.

(c)(iv) (v) There is a mix of item types and assertions that measurements will require complex responses. The exemplar items, however, do not match that description. They tend to be poorly constructed as will be discussed later.

The items types appear to be different depending on when they are assessed. Performance events are used for through-course assessments but not for the end of year assessments. If performance assessments are in place because they provide opportunities to display different qualities of response, there is a question of whether the end of year assessments will be congruent with what occurred during the year. There should be consistency.

Four achievement levels will be set in the early grades. The methods for doing this and the argument that there will be consistency across grade levels is not made in the proposal. Rather the emphasis is on the high school standards.

(c)(vi) An emphasis on AI scoring provides a way to turn around the results quickly. Heavy reliance on the technology assumes continued progress in the development of the technology. The proposal fails to include a thorough description of what technology will be used and what improvements are expected to be made over time in order to insure precise scoring.

(c)(vii)The emphasis on Periodic Feedback Reports is desirable as is the inclusion of parents as recipients. No examples are given. Reporting is a crucial piece of an assessment design and examples or templates for the various reports would make the proposal stronger.

The section on reporting contains statements about how results of the assessments could be used for accountability purposes. It talks about "extraneous factors that threaten causal attribution" implying that the major barriers are merely technical. The proposal neglects to say that teachers have duties to perform that are not necessarily related to achievement outcomes. For instance they are expected to interact positively with parents, school board members, other teachers, etc. In addition for elementary school teachers there are other subject areas for which they are responsible and those areas may not be covered by a state assessment.

Recommendation: In the section on release of items I would recommend using the NAEP Question Tool both as it is and a guide to how items might be released to teachers and students. It is a marvelous application for understanding item results and making responses usable. As an added bonus, when used correctly the Question Tool gives valuable normative information for states and shows results by a number of background and demographic characteristics.

Recommendation: It would be nice to have examples to judge how clear the proposed reports are, especially those that go to parents.

(A)(4) Assessment System Development

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

(A)(4) Reviewer Comments:

(a) The proposal endorses evidence-centered design and universal design as approaches to the development of the assessment. It has a defensible interpretation of evidence-centered design but lacks specificity when discussing other aspects of the assessment. It would have been desirable to be more specific about each aspect of the design. For example when it discusses item types, it throws out lots of labels without adequate definitions. Congruences across all aspects of the design is what evidence-centered design attempts to do.

When discussing the procedures to be used the proposal often does not tie them to the design as developed earlier. For example, there is no direct mention of how the procedures for developing the end of year tests might differ from those of the through-course assessments. The through-course assessments are short. So it is unlikely that they will be able "to appropriately distribute points and items or tasks across the full breadth of the standards for a particular grade level or subject area." There is no mention of the possibility of tailoring the development of the components of the assessment to the different purposes.

The proposal mentions innovative item types in several places. It does not specify what they are and how they will be treated differently, if at all, in the pilot and field testing phase of development. If those item types are to be treated differently, then the proposal should say what characterizes the differences and what evidence will be used to determine their usefulness.

In various sections of the proposal it is stated that there will be accurate measurement across the full spectrum of achievement. Yet it is never clear how this might be done. There is no mention of tailored testing or testlet approaches for instance. The weight of the assertion apparently is tied to the properties of individual items. The question is whether items are that strong. There should be in the proposal an argument about their strength and the strength of the general approach; i.e., evidence-centered design.

(b) The section on accommodations is very thorough and extremely well-done. It is a model for the kind of specificity that makes it easy to understand what is intended to be accomplished by the proposal.

(c) This section is disappointing. It is not tied tightly enough to the preceding sections. It makes broad, general claims about what will be done without giving sufficient specification. For example, IRT models will likely be used for scaling and scoring but there is no mention of which form of IRT or what work will be done to determine the appropriateness of their use. The through-course pieces of the assessment could be sufficiently varied to warrant multivariate IRT models.

(d) The Interactive Data Tool sounds promising. One would like to know more about it and how dependent it is on existing data bases within the states. (States are variously capable of building and managing large data sets.) According to the proposal the tool will be mainly used to disseminate information. Although the reports will be accessible to persons from a variety of positions, their proposal does not say which stakeholders will have access to which reports. For example, if there is an individual student report, for obvious reasons it is important to limit who has access to it.

The nature of the reports should be more detailed. It would be desirable to have an example of a report for a particular piece of the assessment for a particular audience. The proposal says various audiences will be notified about the available reports. How and when each will be notified is pertinent. There are a number of assumptions about access to technology built into the notion that most, if not all, parents, for example, will have access to reports.

(e) The quality control response lists many tasks without comment implying a kind of equal importance and difficulty in accomplishing them. Yet, "review items' media and passages for content accuracy..." is a continuing, time-consuming, demanding process while "score conversions are accurate..." is straightforward and not at all demanding.

What follows is a long but important demonstration of how the proposal fails to link all important aspects of the assessment in a way reflecting an understanding of evidence-centered design. Evidence-centered design in a nutshell says the interpretations you make are based on the logic and structure you use to construct and refine the assessment. That means there must be congruence between the standards, content to be sampled, and test items. The items are the building blocks upon which all interpretations and

uses are based on. Therefore, one must be able to argue from the response demands of the item to the standard being measured.

Page 709 gives an item that asks a student to make a decision about data (which person should be chosen to long jump) using apparently only the mean of a set of scores. (It is not clear whether a student is expected to do more analyses of the data or what aspects of the data she is expected to use to answer the question. An obvious design flaw!). It should be pointed out, however, that the standards expect students to be conversant with centers, shapes and spreads, not just centers.

If one wanted an item that asked students to be able to reason about data, additional displays and calculations are desirable. For example, to answer which student should be chosen to long jump, the following numbers and displays are relevant:

Variable	Count	Mean	Minimum	Q1	Median	Q3	Maximum
Elsa	6	3.6500	2.9500	3.1750	3.7350	4.0330	4.2800
Miki	6	3.7250	3.5500	3.5800	3.6800	3.9025	3.9700
Aisha	6	3.7617	3.6200	3.6575	3.7550	3.8675	3.9200

Note that there is only .04 meters (about 1.5 inches) between the two longest jumpers on the average. But the third jumper has a longer jump (4.28) than either of the other two.

A boxplot that could not be pasted into this response shows how varied the scores are. It shows that Elsa scores are much more variable than the other two. What is it about Elsa the student should ask? Good reasons for choosing a particular person depend on variation in the data as well as the centers of the data. Never a center without a spread!

An item with these data and the boxplots would begin to make the evidentiary argument.

(A)(5) Research and Evaluation

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

(A)(5) Reviewer Comments:

(a)The proposal contains a long list of studies that will be conducted. They cover the issues in the criteria and extend them in some cases. Sometimes the charge is overwhelming: "...TAC's key tasks will be to recommend ... a common assessment system that can provide reliable results to inform valid instructional and accountability decisions." That's everything! What exactly will be done, when, how, and by whom is missing from the proposal. Specificity is needed.

One extremely important omission in the list of studies is that of classification accuracy. At each grade level and for each through-course assessment standards are to be set and determinations made of whether or not a student is "on track." How good the decisions are depends on the properties of the errors of the assessment. The more reliable the scores, the more accurate the decision. More accurate decision, in general, come from longer tests. These tests are very short and likely to produce scores with huge errors associated with them.

RFI-8: Vertical scaling is described as a challenging endeavor. Yet implied in the decisions about "on track" or not is a vertical scale or something that makes "on track" mean the same thing across grade levels and types of assessments.

There appears to be inconsistency between a desire to use the results for judgments about teacher effectiveness and growth models. The proposal suggests that teachers will be evaluated on effectiveness

but leaves upon the question of which model. Yet the section of teacher effectiveness contains language of "true gains." True gains is just one model for growth. Gains adjusted for background characteristics is another. There is an inconsistency in the proposal when in one section it is said that there is no recommended growth model but in another section mentions using true gains.

(b)The proposal suggests evaluating not only the intended effects of the assessment but to also look for unintended ones. Those unintended ones can be either positive or negative and should not be ignored.

Under the title of Social Impact the proposal responds to the "a lot more testing" issues. The commitments there are worthy ones.

A recommendation would be to look carefully at accuracy consistency and its consequences both positive and negative. An inaccurate classification could become a political football. Students are said to be "on track" this year but not next but again on the third year. This could easily happen if the scores are not precise enough to make good decisions. This is particularly acute of the through-course decision since there will be multiple decisions based on the results of short assessments.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	10
(A)(6) Reviewer Comments:		
<p>(a)Again the proposal gives a long list to meet these criteria. One strength of the list is that it refers back to previous sections of the proposal and shows how the capacity building, for example, is tied to the activities already proposed. This, of course, means that what was proposed earlier will be completed on time and done well. With projects as large as these time lines often need to be changed.</p> <p>Particularly important would be work to show that classification decisions are accurate so parents, and other audiences, get a consistent message about their child's progress. The proposal does not address classification accuracy although it is relevant for both the through-course assessments and the final one.</p> <p>The sequence of "courses," Assessment 101, etc., could be quite powerful. It would be nice to know more about the content and the scope of the "courses." Also the notion of gathering feedback during the piloting is a good one.</p> <p>(b)The proposal covers an array of stakeholders and how they will be informed.</p> <p>A recommendation: If the Partnership Communications Committee does not come up with it, there should be a systematic attempt to inform media representatives in ways in addition to press conferences and press releases.</p>		

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	7
(A)(7) Reviewer Comments:		
<p>(a)(i) There is a long list of things that will be done. Often it is assumed those things are done easily. In technology what can go wrong does go wrong. For example, producing an assessment system that must "seamlessly integrate with states' data management systems" may be a daunting task.</p> <p>The proposal contains a number of ways in which technology will enhance item development including both how various item types might be implemented and how item development committees might be</p>		

formed. In both case, the initiatives may be more difficult to implement than the proposal suggests. Creating templates for items that demand complex responses, for example, are not easy to do. And they may be even more difficult to utilize in a Web conference because often more can be done when persons are face to face and can converse freely about what they are doing.

(a)(iii) Past history suggests that states are differentially capable of producing and using large data bases. It is not clear from the proposal whether there will be extensive training where it is needed.

(b)The proposal contains a number of barriers that should be addressed. It would have been desirable for the proposal to contain some sense of how much it will cost to reduce them. For example, the proposal mentions the issue of accommodations and costs but does not give details about either.

A recommendation would be to have the developer have backup plans for issues that develop when technology fails. Note that technology is an hierarchical thing so the failure or barriers can be at any of many levels. A glitch in scoring in one thing; a districts computing needs quite another.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	13

(A)(8) Reviewer Comments:

(a) Achieve, through the American Diploma project, has had extensive experience managing a project that includes state cooperation. Although they have participated in projects that include assessments, they do not have a substantial amount of experience in designing an assessment system. So to the extent that managing a system implies a good deal of knowledge about what can and should be done and the knowledge to choose well additional personnel and contractors, Achieve appears to be a good candidate for project management.

The director of PARCC, the associate director, the director of assessment and director of post-secondary engagement are all to be hired, so there is nothing known about crucial persons of the management team.

(b) The timelines seem adequate. Under Responsible Entity the phrase "design committees recommend, governing board decides, Achieve facilitates" is not as informative as it could be. If the design committee recommends but the board decides against, there is no fall back position.

(c)(i) The proposal meets the specification in terms of clearly defining the costs per module. The allocations are in line although governance seems a tad high and technology a bit low.

(c)(ii)(iii) The budget covers all important aspects of an effective assessment system and reasonably allocates resources to produce the system. These amounts seem justifiable.

Recommendation: Since a substantial number of persons involved with the project are not identified, it would be desirable to learn more about what characteristics and experiences Achieve is looking for in each of the appointments.

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:

(a)-(c) States may differ in the amount of cooperation between the State Department of Education and an Higher Education authority. Often this is because of turf issues but also it can be a matter of sufficient resources. It is difficult to discern whether there is this type of cooperation in the participating states.

The level of commitment is difficult to judge as well.

The proposal suggests that 90% of the matriculates are accounted for.

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 applicant meets the absolute priority.

It could be clearer on the following issues:

(a)(ii) The proposal asserts that it meets this criterion but does not provide concrete evidence of what it will do. So it talks about varied item types, complex response demands, and demonstrations of knowledge and skills but it does not give good examples. In fact, some of the exemplar items are seriously flawed.

(a)(iii) The proposal says it will provide accurate measurement across the range of proficiencies. It uses performance assessments written to elicit a wide range of responses as one way to accomplish it. The through-course assessments will have six items and the proposal does not specify how long the end of year assessment will be. In order to assess across the broad range of performance it is necessary either to have some form of tailored testing or to have a very long test. Whether the combination of through-course prompts and end of course assessments will span the proficiency range is an open empirical question. Past experiences suggests this will be difficult to do as envisaged by the proposal.

(a)(iv) The proposal contains a request for a contractor to produce a growth measure and an argument to support using it. The data produced by the assessment is, of course, the foundation for growth. The proposal is not clear about whether a previous year's end of course assessment may be used as a baseline for a growth measure.

(b)(i) The proposal meets the requirements for annual assessments. But it is unclear if a state is allowed to opt out of the additional secondary school assessments.

(b)(ii) The proposal says it can produce growth data to be used to determine whether or not students are college and career ready. Often it is not clear about how it will be done. A specific technical issue is how well the through-course assessments assign students to categories; i.e., the classification accuracy.

(c) The section on accommodations is the strongest part of the proposal.

(d) (i-ii) The foundation for each of these accountability settings is a strong assessment. According to the AERA/APA/NCME test standards one score, however, should not be used as the only basis for an important educational decision. If each of the components of the assessment were as powerful as suggested by the proposal, the scores would be one important piece of an accountability system.

(d)(iii) There are a number of ways cited in the proposal to gather information about what school personnel may need to be better assessors and better users of assessment results. The presentation would have been more powerful if one concrete example had been discussed thoroughly.

(d)(iv) The proposal designs a system that will be used both for accountability and instructional purposes. It does not discuss how those two purposes may not be completely compatible. Past experience with state assessments that attempted to both improve instruction and provide accountability have experienced an excessive focus on accountability and less emphasis on improving instruction. There are those who say that teaching to the test, one unintended consequence of accountability pressures, makes instruction worse, not better.

Grand Total	220	133
-------------	-----	-----

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

The budget seems reasonable. Perhaps one could argue that smaller amounts could be spent on governance and larger amounts on assessment design (the heart of the project) and technology.

Level 2 Budgets

Name: Formative Assessment Tools for K-2

The budget appears defensible.

Name: Technology Enhancements

The budget seems reasonable.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b) (6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	16

(A)(1) Reviewer Comments:

The applicant, PARCC, proposes a governance structure that appears conducive to the successful design, development, and implementation of the proposed assessment system.

PARCC's vision is rooted in the recognition that at present, high school graduation is not an effective indicator of readiness for college or a career—what they call a "readiness crisis." They propose that collaboration between K-12 and higher education across their partnership will drive educational reform and improve the likelihood that students will make progress—getting on and staying on track to not only graduate high school, but to do so sufficiently prepared to be successful in subsequent endeavors in college or the world of work.

This section of the proposal clearly lays out in brief the major components of the PARCC system, each cross-referenced to the section(s) in which more detail is set forth. These key components are all addressed and elaborated upon in the Partnership's Theory of Action.

PARCC's proposed governance structure appears suitably designed to support the design, development, and implementation of a common assessment system.

The intent of the Partnership is that every member state will participate in the design and development of the new assessment. The proposed governance plan is consistent with PARCC's goal to create "a cutting-edge assessment system that states can fully implement statewide no later than 2014-2015," although the ultimate success in reaching that goal obviously depends on far more than thoughtfully established roles and responsibilities.

The choice of roles—Governing State or Participating State—requires different levels of investment of staff time. Participating States are "invited—and encouraged" but not required, as are Governing States, to commit staff to committees and working groups, but all must participate in piloting and field testing of assessment components. Although, according to the proposal, Governing States will have the most decision-making authority, it is unclear whether official status alone will drive decision-making processes/protocols if degree of involvement is intended to be a factor. This ambiguity is based on the consortium's expressed intent to leave the most critical decisions to states making "the greatest commitment to help create and ultimately implement the proposed assessment system." This is of some concern, particularly given the moderate number of Governing States at the outset (10 states plus the District of Columbia), since without the "carrot" of greater authority, Participating States not yet ready to switch their status may be unwilling to contribute the substantial number of staff (and amount of staff time) likely to be necessary in order for the various committees, working groups, and task forces to function effectively. If in fact Governing State status is required in order to wield greater decision-making power, it can only be hoped that Participating States will either switch status or be generous with staff, nonetheless, in order to achieve and maintain the necessary capacity to reach the Partnership's goals.

PARCC's procedures and policies for joining the Partnership are reasonable—essentially to make all the same commitments as existing members. The proposal makes clear that states joining after the application for RTTT Comprehensive Assessment Systems grant funds may not reopen settled decisions or engage in

other retroactive measures. In addition to outlining procedures and policies for any state wishing to join the Partnership, change status, or opt out, PARCC explicitly identifies a recommitment requirement which reveals a realistic understanding of the possible impact of changes in political leadership during the tenure of the grant.

The proposal identifies governance and decision-making protocols, as well as plans for adopting key policies and definitions, that appear feasible, in general. Given that the plan is for the Governing Board to make major policy and operational decisions on behalf of the Partnership, it will be critical that they have sufficient and timely input from the various design, advisory, an working committees. The proposed alternative, when committees cannot reach consensus on major proposals--presenting multiple options supported by descriptions of the pros and cons of each--is appropriate, but decision-making may be hampered by holding Governing Board meetings only quarterly. One sign of forward thinking was the Partnership's selection and identification of the first Governing Board Chair. Another was the indication of the intent to have charters drawn up for committees and working groups. Some details, such as the goal of "maximizing involvement [in committees] across the Partnership while keeping groups to the optimal size for completing their changes" showed realistic thinking clearly borne of experience; however, other details (e.g., treating test specifications and a calculator-use policy as calling for the same level of decision-making) did not.

The PARCC proposal included all necessary tables and supporting documents for this section. It was commendable for the timeline to include plans for interaction between the two consortia submitting proposals (although no acknowledgement was made that this presumes awards to both consortia); references to "the RFP development process" and "the RFP" were non-specific and thus not helpful.

MOUs for all member states were included in the proposal. Included plans demonstrate that many member states are already in the process of engaging in necessary preliminary steps (e.g., adopting the Common Core State Standards [CCSS], conducting gap/match analysis between CCSS and state standards, etc., taking steps to revise legislation as necessary). Few states identified any barriers to participating in the proposed assessment (with only two mentioning the potential barrier that might occur due to insufficient funds, one implying that insufficient technology resource readiness might be an impediment (this, a Governing State), and two states identifying both funding and technology as potential barriers).

One potential barrier the overcoming of which may be somewhat more of a challenge than acknowledged or anticipated is changing item and scoring rubric release policies of some states (e.g., Indiana). This is one area in which clear and timely information to parents and other members of the public at large will be critical, so that no one is led to feel that "the wool is being pulled over their heads."

Based on the overall soundness of this response and limited number of issues/concerns it evoked, the response was scored in the low end of the "high" range.

(A)(2) Theory of Action

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

(A)(2) Reviewer Comments:

Overall, PARCC's Theory of Action is coherent and credible; one of its strength is that this narrative provides context both for the particular goals of the assessment system and evidence of understanding the limitations of even the most innovative and effective means of evaluating student performance in order to improve student achievement.

The components of the system all have the capacity to result in improved student academic outcomes. Particularly noteworthy is PARCC's plan to include distributed "through-course" assessments both because of the increased accuracy of measures of learning in closer proximity to instruction and at a time that can

result in more immediate (and therefore, more likely effective) action. Again, the proposal includes some detail to provide context (e.g., current use of interim or benchmark assessments and their weaknesses).

Based on detail in the Theory of Action, technology will play an important role but is not the linchpin upon which the impact of the assessments rests. It is realistic to anticipate that based on current advances in technology and testing, the Partnership will be able to incorporate items/tasks that can be both delivered and scored by computer. At the same time, at least some of the through course assessments (which will contribute to annual combined scores) will rely on methods of delivery and scoring that reflect and respect the standards being assessed (e.g., aspects of speaking and listening).

The dual themes of identification and remediation support the idea that this assessment is intended to function as more than a tool for accountability (although that is one of four key impacts identified).

The description of the intended assessment system (not only in this section but in the Executive Summary) as one that classroom teachers will find “provides as much for them as it asks from them” is powerful and appealing; however, while the narrative provides some general examples of what it will provide, it is curiously silent about what will be asked of teachers. This raises a “red flag” because although teacher involvement is likely to be both necessary and extremely valuable, to be fully credible any proposal must instill confidence in the capacity of all contributors (teachers, instructional leaders, SEA personnel, others) to do all they are being asked and are expected to do. In addition, in and of themselves, neither the availability of “clear targets” nor “access to actionable assessment data” will necessarily be of utility to teachers; some elaboration upon how utility will be ensured would have been helpful.

The mechanisms described for achieving the intended outcomes are logical and realistic; of some concern, however, is the tendency in the proposal to claim what will happen as a result, rather than more realistically describing what the assessment creates the potential to happen. The desired end-results depend on far more—and a high quality assessment system is necessary but not sufficient, something that the Partnership itself acknowledges elsewhere in the Theory of Action (noting that the proposed assessments are “only one lever for improvement in the broader educational system”).

The Theory of Action contains many engaging ideas about the nature and uses of data derived from this assessment system. One that stood out was the use of results from the through-course assessments to inform mid-year professional development (a much-needed change from the “flavor-of-the-month,” test-prep or text-book orientation focus of far too many schools and school systems today).

PARCC’s Theory of Action ends with an acknowledgement of the challenges posed by implementing a performance-based assessment system that is “scalable and sustainable.” This recognition of what has worked well in the past and merits serving as a building block is more compelling than any blanket condemnation as inadequate of existing assessments or claims to create something brand new. The partnership states bring valuable experience in designing and implementing performance assessment—doing so in the service of readiness for college or careers, and in a financially sustainable way, is a commendable goal.

As a result of all of these observations, the response received a score in the high “medium” range.

Suggestions to the DOE:

In spite of clear efforts to address career readiness as well as college readiness, the proposal appears to attend more to post-secondary education than the workplace. It would be useful to hear more about career readiness (ways in which it may be unique and the many areas of overlap with college readiness).

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	41

(A)(3) Reviewer Comments:

In this portion of the proposal narrative, PARCC reiterates the four features of what they believe will be a sophisticated summative assessment as well as the components of the assessment system and timeframe for their administration. They also reiterate the key strengths of their "distributed" strategy; each of these has considerable merit, particularly the resulting actionable data and source of models of student work—a resource that has been shown quite effective (if used meaningfully) to make targets clear and illustrate what "good enough" and "really good" look like. Decisions about how to best combine results from the through-course and end-of-year components will be driven by research to be conducted during piloting and field testing, and the focus of annual reports will be on whether students are on track for or have demonstrated career and college readiness. The overview is followed with more specific detail that generally (although not consistently) expands upon, and provides a rationale for, particular design features.

It is commendable that the Partnership aims to increase transparency through the public release of ELA prompts. They have offered only general ideas regarding issues and possible solutions, and it would have been helpful to have seen both more fully fleshed out, perhaps by offering some precedents—which do exist—for public release of prompt banks of operational assessments. Clearly outlining the implications for the timing of test development, piloting and field testing, leading to conducting equating studies, would have strengthened this particular notion and shown that while innovative, a public prompt bank is indeed feasible.

Initial confusion was created by the statement that all of the through-course components will be administered in an online, computer-based mode, but cleared up quickly by the description of the speaking and listening component as a "live" classroom performance task. PARCC's thinking about grade appropriateness of typed versus paper-and-pencil responses parallels is cautious, but defensibly so, as it is in line with the National Assessment of Educational Progress (NAEP) which has set 2019 as the target date by which students at the elementary level (Grade 4 in that case) will shift from written to key-entered responses on that writing assessment.

For each of the ELA components, the proposal sets forth a rationale that ties in closely to PARCC's Theory of Action and establishes the significance of what will be measured (and how it will be measured) to college and career readiness. The uses of technology anticipated in the delivery of these items is realistic based on current capabilities and should only become richer and more sophisticated as time goes on. For the end-of-year ELA component, the uses of technology are also realistic as described, assuming that item types involve answer keys, defined menus of answer cues (key content and linguistic elements), or other easily programmed scoring schema such as marked visual field or manipulation of online graphic or textual components.

Within these pages of the proposal, there is some minor "clang" that occurs due to different ways of talking about the CCSS. Differences, for example, between describing prompts in one section (ELA-1 and ELA-2. Writing from Sources) as being modeled "on the CCSS" and in another (ELA-4. End-of-Year Literacy Assessment) as being modeled "as closely as possible on the CCSS" may reflect a mere inconsistency in language or instead, a real (in which case problematic) distinction.

Standing out as different in various ways is the proposed ELA-5. Speaking and Listening Assessment. Without any explanation, it is possible to come up with various plausible reasons why, according to the proposal, the plan is for this component not to contribute to the combined summative score in ELA/literacy. However, as it stands, without explanation, this detail is troublesome. As virtually everyone involved in one way or another with assessment knows, what counts is what gets addressed.

PARCC intends for the assessment system to measure the full range of the CCSS, including those less often measured or difficult to measure. Since that historic difficulty is mentioned, it behooves the applicant to address (at least in brief) the source of those difficulties and to explain ways that the new assessment system will succeed where others have failed. Basic requirements for minimum number of measures/score points for scaling purposes, coverage not just at the standard level but across objectives, and variety of stimuli for example create challenges that are not really addressed in this proposal. Only a few of these are the following:

- The very richness and authenticity of some of the assessment task types proposed also lead to many scoring issues. For instance, comparison of a specific work to a much larger body (the “American Literary tradition” is the example cited in the proposal) would seem to require not only human raters but those who have considerable specialized content knowledge.
- Use by students of multiple sources of information requires scorer access to those sources.
- The opportunity to measure multiple skills and understandings within and across standards is not the same as an occasion that explicitly cues for demonstration of those skills and understandings. Scoring strategies can be—and too often are—devised that erode validity.
- Experience (captured in research) has demonstrated the dangers of confounding evidence of different skills and understandings when incorporating items/tasks that are intended to provide multiple measures. Care will need to be demonstrated that rubrics/scoring criteria can accurately and efficiently discriminate between/among these.

More evidence would have been useful to demonstrate that the assessment system will truly measure the full range of the CCSS, since the implication of various statements in the proposal runs counter to the claim that PARCCs assessment will do so. For example, the description of the writing assessment components suggests that they are always text-based, while the CCSS includes narrative writing (accounts of real and imaginary events). Nevertheless, the Partnership’s careful thinking about this topic is illustrated in various ways, for example their remarks regarding the challenges posed by genre effect and their plans for addressing those challenges.

The overall plan for this system—which would seem to assign more complex and labor-intensive tasks (from the perspective of administration and scoring) to the through-course components and more easily assessed standards to the end-of-year component—makes good sense.

The Partnership’s thinking about mode of administration of mathematics components is realistic and appropriate, taking into consideration familiarity with, and developmental appropriateness of, technology that may support delivery. Their design supports the curricular structure underlying the CCSS and reflects sound thinking about how learning occurs. The use of technology for the end of year mathematics assessment is both meaningful and realistic. It is worth noting that the “next generation selected response items that signal not only whether students provided a correct answer but also help analyze why some students might have provided an incorrect answer” need not be limited to the mathematics assessment; similar items have been developed and used for reading as well.

The proposal includes a detailed description of plans for formative assessment tools and resources guided by experience and research. Among these, two innovative supplementary resources PARCC plans for are a text complexity diagnostic tool and developmentally appropriate resources for assessing K-2 literacy and mathematics knowledge and skills. Examples of resources developed by Partnership States that they are willing to share gave some “meat” to PARCC’s plans for formative assessment.

PARCC’s scoring plans are overly general, and this seems a central weakness of this portion of the proposal. This is perhaps understandable in that much detail will only evolve as item and task types are fleshed out; however, evidence of capacity would have been enhanced by some evidence of understanding the issues and implications of different scoring methodologies under consideration. In particular, without more information about plans for Artificial Intelligence (AI) scoring for ELA, there is no evidence that the Partnership has addressed the complex preparations for programming for AI scoring of many types of extended constructed responses (ECRs) and particularly for scoring the products of the sorts of multi-day extended performance tasks proposed by PARCC for grades 6-11. There is nothing in the proposal to explain why PARCC anticipates that a greater percent of the mathematics assessment will be scored by human raters (50% AI-scored, 50% human-scored) than the ELA assessment (75% AI-scored, 25% human-scored). It is in this regard that the Partnership’s otherwise feasible plan for integrating technology may be questioned.

Extended constructed responses have the capacity to yield multiple measures, yet there is nothing in this proposal to address ways of approaching scoring of the limited number of ECR items in each through-course assessment to capitalize on the opportunity to obtain data on a wide array of different skills and

knowledge. Some standards lend themselves to AI scoring, while others, less clearly so (and the technological capability for scoring some standards is far more advanced than for others). Detail in the Summary Tables about the percent of AI scoring and human scoring does not establish whether these describe the pool of responses or use of different methodologies to obtain different measures from the same responses. This is a critical distinction.

The note that Partnership states may not wish to use teachers as scorers because one use of results will be to inform evaluation of teachers and instructional leaders showed insight and anticipated what will very likely be an issue in the future. Even if student work is randomly assigned to raters, some teachers may be unduly influenced by knowledge that their scores may affect the standing of another teacher somewhere; therefore, assigning scoring to a contractor's pool of trained and qualified raters may be, as PARCC suggests, a reasonable alternative.

The Partnership documents relevant experience with setting standards and shows deep understanding of these processes. The account of kinds of information to be produced is detailed and thoughtful. Without any sort of rationale for which data is intended for each of various stakeholders (Tables 2 and 3), this information was difficult to evaluate. The Partnership has given considerable thought to the uses of available data.

Plans for increasing accessibility for all students are thoughtful, feasible, and clearly informed by research and current notions of "best practices." Anticipated uses of technology towards this end seem realistic.

The Summary Tables for (A)(3) for both ELA and Mathematics provide detail to support the narrative. Some of that detail works well as "stand-alone" information, but some—such as the number of items likely in the end-of-year assessments—could benefit from some explanation, without which those numbers are questionable (e.g., leaving readers to wonder on what basis they will ensure full and adequate coverage of all the standards, taken in conjunction with data from the through-course assessments).

One final note: Given the pervasive sense that when speaking about college and career readiness, the former is often given more weight than the latter, it was good to see throughout the narrative various explicit references not only to knowledge and skills fundamental to the success of students but also to employees, and to the expectation not only of college faculty but also employers.

Weighing concerns about scoring (including but not limited to the proposed uses of technology for scoring) and the omission of detail to support some key assertions or observations against the many strengths of this response, it was assigned a score in the low "high" range.

Suggestions to DOE:

In every assessment, one key challenge is not allowing faulty interpretations of what's important and what's not to spread based on the impressions of students, teachers, and others. Careful consideration needs to be given, therefore, to how the public is informed about "what counts" (that is, what measures will contribute to different reports and data analyses).

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	24

(A)(4) Reviewer Comments:

PARCC begins this section of their proposal with an explanation of their theoretical approach to assessment development; this approach supports accessibility of the assessment from the outset, and allows for a reasonable variety of item presentation formats and response modes--both conventional and electronic. Within this section, the Partnership specifies the use of task design templates to allow for timely and cost-effective development of items that "validly measure the same construct in multiple ways." This is an approach that has been found to work well in other contexts (although success will always depend on

the quality of work that goes into template design). The partnership plans to initiate development work immediately upon award of a grant, however, which puts some pressure on them to develop and test templates within a very narrow timeframe.

The Assessment Design Committee proposed consists of a suitable array of participants; there is nothing to indicate whether the teachers from each member state selected are expected to have at least some experience creating assessment items (with or without the use of technology as a delivery platform and or as a means of capturing and/or scoring resultant responses). This may be critical given the timeframe proposed for assessment development. Furthermore, virtually all phases of the assessment development process would be strengthened immeasurably by setting requirements for prior experience (whether with classroom/formative or large-scale/summative assessment). The same can be said of teacher involvement anticipated in other groups (e.g., test item review committees).

Some activities described in the development process suggest "putting the cart before the horse." For example, it may very well be premature for panels to recommend the score point range for each innovative item and performance task type until and unless the templates for those items have been developed and tested. To do otherwise may lead (as it has in other assessments) to arbitrary score scales. As a general rule, if item development is less constrained, score ranges are going to be more varied if they fit—as they should—the particular item for which they'll be used. If consistency among score scales is desired (e.g., all brief constructed response items (BCRs) using a 0-2 scale), then item development will necessarily be more constrained.

In the section headed "Test Blueprint Development" it was good to see the Partnership demonstrate thinking about not only fullness of coverage of the standards but balance/focus among those standards. While not explicitly identified as a requirement in the NIA, this is a critical element of assessment design and development that sadly, is often overlooked, or insufficiently attended to during the development process even if intended at the outset.

The Partnership's plan to procure vendors for key components of the PARCC assessment system seems realistic and appropriate, so long as stakeholders are involved at all stages and processes are sufficiently transparent (which may "bump up" against some vendors' customarily proprietary practices). One area in which more initial vendor involvement may be advisable is in the development of prototype items involving technology, since there is at present fairly sparse state experience in that type of innovative item development.

The Partnership plans for enough items to be developed or obtained to build eight parallel test forms at each grade level; the narrative does not make clear whether this means eight forms of each of the components envisioned for use over the course of a year in ELA and mathematics, nor does it include an explanation or rationale for that specific number of forms (something that it would have been helpful to know rather than being forced to infer).

PARCC's planned use of cognitive labs to test new items and tasks (particularly those that are more innovative) is commendable; the Partnership might consider applying a similar approach to learn more about any approaches to hand-scoring that are new (e.g., application of multiple rubrics/evaluative criteria to assign scores, and/or the scoring of responses delivered online that require following students' "footsteps"). The Partnership recognizes that adequate representative sampling of special subpopulations of students with disabilities and English learners, while always important, is especially so given the need to evaluate innovative item types delivered by computer and appropriate testing accommodations. It makes good sense to initiate pilot testing as early as PARCC proposes to do, but this will require a development timeline that may not be realistic. Their consideration of two options to field test computer-based items shows careful thought, and there is certainly wisdom to consulting with the Research Strategy Group (RSG) and Technical Advisory Committee (TAC) before making a decision about which route to take.

Review of field test results should not only focus on quantitative data but on qualitative data sources—specifically the close analysis of unexpected/aberrant responses (that is, ones that do not clearly fit along the continuum of expected products/performances). This will require careful programming for "kicking-out" from AI scoring responses that are unusual, and identification of the need for fine-tuning or even major overhaul of rubrics/evaluative criteria used by live raters. Examination of item statistics is certainly

necessary but by no means sufficient to prepare for an assessment system that must yield valid measures of what students know and are able to do.

PARCC outlines clear and comprehensive steps to design and develop accommodations for special student populations, and presents suitable criteria for determining comparability of scores with and without accommodations. Research is planned in order to validate new accommodations in a timely manner.

As suggested by comments on (A)(3), scoring is one aspect of the assessment that may require more attention than anticipated, even with as detailed an approach to scoring as is proposed by the Partnership. The proposed scoring plan seems to underestimate the amount of prescoring required by human raters (ideally "expert" raters) for AI "training" purposes. The process of resolving scores that "diverge" lacks clarity since divergence is not defined as adjacent disagreement (differing by one point) or discrepant agreement (differing by 2 or more points).

While the Partnership describes and provides some rationale for the source of ratings in both field test and operational administrations of the through-course components, there is minimal information about plans for scoring training of human raters, monitoring of those raters (although there is at least a reference to inserting prescored validity responses back into the pool of student responses to be scored), recalibration procedures and procedures for rescoring in a timely fashion those responses scored by less-than-adequate raters), and procedures for checking year-to-year consistency. Specifically, attention needs to be paid to the selection of responses used to check year-to-year consistency among raters. It may be problematic to fold back in previously scored responses if the score was assigned by AI alone or by one rater (who may or may not have assigned the correct score). Far preferable would be the insertion of responses for which a "true score" (score agreed upon by multiple expert readers) has been established. AI scoring also needs to include procedures for flagging and rerouting to one or more human raters unusual, "low confidence" responses.

It is curious that variations in the quality of handwriting is given as a reason for a higher rate of read-behinds at grades 3-5 (footnote to Table 5), since research conducted by various states and testing companies considering the switch from handwritten to word-processed responses for writing assessment purposes has demonstrated that there are in fact more issues associated with human scoring of typed responses (e.g., underrating sufficiently developed responses that in typed form appear quite brief)—issues that can lead to discrepant agreement among raters.

Again, it was good to see that the Partnership recognizes that there may be issues associated with using teachers as scorers (even in a distributed system) because of the potential use of the data as part of teacher and instructional leader evaluation. However, PARCC does not provide any detail to explain real or perceived benefits of teacher engagement in the scoring process.

Use of vendor services related to scoring is a realistic approach, particularly in terms of conducting training; however, it is questionable whether the development of training tools and scoring protocols should be left to a scoring contractor (as implied in one part of the narrative but assigned to others in the section on score quality assurance) rather than being addressed during item and task development. At a minimum, there should be some opportunity for collaboration. Since items will be recycled on subsequent test forms and/or released as part of a formative assessment toolbox, there needs to be a clear plan for validating or fine-tuning the annotations and other information used to guide scoring during and after each scoring enterprise. It is likely that a detailed system of communication and collaboration between educators and experts from Partner states and the scoring training contractor's senior staff will need to be devised.

Given the significant role that technology is anticipated to play in the PARCC assessment system, a source of concern is the limited reference to quality control measures in this regard; the proposal does note plans to develop processes to ensure that computer scoring of innovative items is accurate, but fails to discuss quality control procedures for AI scoring of writing, for example, or procedures for ensuring that whenever technology informs the design and delivery of items it does not precipitate unexpected/unwanted consequences.

In many respects, the Partnership's assessment development plan is quite sound, but various issues related to the use of technology have the potential to delay full administration of operational summative

assessments based on A-3 by 2014-2015 or require a shift to less dependence on technology (at least initially) as a "Plan B." For this reason, and based on some concerns expressed in this review, the response was scored in the high "medium" range

Suggestions to DOE:

One avenue that PARCC fails to address for building a sufficiently large and varied item bank that meets all their criteria is to examine pre-existing item pools from state and vendor sources that could be meaningfully converted to technology-assisted/enhanced items. This might require that some sort of orientation/training protocol be offered by the vendor(s) dealing with item development on characteristics and requirements of computer-enhanced items, with review of sample items currently in use or still in the "pipeline."

PARCC's proposal includes the engagement of teachers in the scoring process through distributed scoring, but is amenable to vendors providing trained raters to conduct scoring. Particularly since the notion of teacher scoring as a valuable form of professional development continues to be expressed (and may even be gaining in currency), the Department may wish to have PARCC explore the issues and implications of teacher scoring further, to aid states in making a decision about who ought to serve as trained raters. The literature on scoring includes not only research that promotes scoring as professional development but research that identifies important caveats. These sources should be consulted and used to inform any decisions in this regard.

Although not called for in the NIA, I would strongly recommend that any cooperative agreement call for the establishment of an outside team of assessment specialists and others to address issues and concerns related to quality assurance. A useful model might be the NAEP Quality Assurance Technical Panel led by HumRRO. With the likelihood of involvement by multiple vendors, each of whom may regard certain resources as proprietary, it is critical that all involved be assured of everyone "playing together nicely." Furthermore, such a team could evaluate objectively whether inappropriate shortcuts were being taken (whether in test design, development, scoring, and/or data analysis) that might put the quality of resultant data at risk.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	24

(A)(5) Reviewer Comments:

This portion of PARCC's proposal is subdivided into two topics: the Partnership's research advisory structure and the Partnership's research agenda.

In the first section, several advisory groups are identified, their membership and role(s) briefly outlined, and frequency of meetings given. This latter information is of limited use, however, without some explanation or rationale for the initial and reduced frequency of meetings of all three groups over the years of the grant. In the case of Issue-focused task forces, it is unclear why meetings in all cases will be reduced to once annually in the last three years of the project when in fact some issues may not be defined or become relevant until partway through the years covered by the award.

There is merit to each of the three options the Partnership is considering for carrying out the research and evaluation program. It might have been useful for the proposal to have included at least a brief description of pros and cons of each option and/or what plans—if any—exist for determining which route to take.

The Partnership's plan to issue RFIs on critical topics (e.g., assessment design/development, strategies to maximize validity, reliability, etc.) immediately upon notice of award is a realistic and necessary step to expedite work on assessment design and prepare to issue one or more test development RFPs. Their acknowledgement of the need for some of the RFIs to lead to research studies based on the adequacy of information obtained shows careful forethought.

The descriptions of RFI-1, RFI-2, RFI-3 and RFI-6 reveal that the Partnership has not underestimated the likelihood of issues and implications related to technology, and wishes to "get out in front" of these. The Partnership raises some good questions about developing technology to meet the needs of the proposed assessment system; what is lacking here or elsewhere is any evidence of thinking about alternatives in situations where technological solutions (existing or able to be developed within the necessary timeframe) wind up being deemed inadequate or inappropriate.

PARCC's research agenda is consistent with, and supports their theory of action. The preliminary program of study—the phase encompassing research and design— addresses key components and requirements of the assessment system and is also intended to inform standard setting. The proposed study related to AI scoring addresses some necessary questions; however, it is likely that as the Partnership explores this aspect of the proposed assessment system further (e.g, as an outgrowth of RFI-3), many more questions will arise that will need to be addressed through additional research, including the potential, inadvertent impacts of AI scoring on item and task design, or the capacity of AI scoring to get at more complex, traditionally hard to measure skills and understandings.

In conjunction with proposed study #8 (efficacy of younger students taking the various components online) it would be advisable to explore the implications of variations in technology tools customarily used by students in Grades 3-5 nationwide. The very mention of "keyboard and mouse," for example, implies that students are used to the same means of navigating the computer screen, while clearly some use a touchpad or other device. More is likely going to need to be known about students' familiarity with, and preference for using keyboard-based commands and/or cursor-selected icons on a drop-down menu, for example. As research on the use of technology in other assessments has made clear, it is critical that measures of the intended construct not be confounded by the degree and range of familiarity with technology resources. These additional dimensions to the study clearly fit under the NIA criteria.

Appendices for this section were detailed, clear, and extremely informative. They provide evidence of the Partnership's careful and extended thinking about their research agenda.

The second phase of the research agenda has as its focus the social, operational, fiscal and technical impacts of the design, development, and administration of the proposed assessment system. The areas of research were described in only general terms, creating curiosity (but not concern) about what sorts of monitoring approaches are anticipated. The Partnership unequivocally expresses commitment to pursuing refinements to the design and development strategy whenever evidence for each of the areas identified raises concern. This leads right into their expressing a commitment to pursue studies to evaluate the implementation and impact of the assessment system and inform decisions about mid-course adjustments needed to address PARCC's theory of action.

The Partnership rounds out the description of their research agenda with brief accounts of various consequential validity studies they intend to conduct. They demonstrate a realistic view of their capacity to obtain the data necessary to validate that the assessment system is meeting its ultimate goal—to prepare all students to succeed in college and 21st century careers. PARCC makes clear the need for the U.S. Department of Education to manage aspects of some studies, but expresses willingness to work with the DOE to provide the data needed to do so.

On the basis of all these observations, the response was scored in the "high" range.

Suggestions to DOE:

Far too often, technical advisory committees convened to support assessment systems are heavily weighted with psychometricians but light on other types of technical expertise, most particularly in the areas of assessment design (including scoring methodology) and implementation of scoring. Particularly given the plans for innovative item/task types and new and only now emerging uses of technology in these regards, it would be not merely useful but very likely essential to ensure that a TAC includes one or more members with expertise in these areas.

PARCC's description of anticipated predictive validity studies contained the first and only reference to the military as a career choice; it seems worthwhile to infuse in materials for stakeholders references not only

to "college and career readiness," but to higher order skills that are important to, and valued by the US military.

(A)(6) Professional Capacity and Outreach

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

(A)(6) Reviewer Comments:

The Partnership's professional capacity-building plan focuses on increasing educators' understanding of the system, their participation in key roles/activities, and their meaningful use of results to inform and improve instruction. Four components are outlined which are intended to address this focus. The Partnership makes clear, up-front, that they see their role as developers of the tools and strategies needed, but they will not be directly involved in training/professional development. Their expectation is that existing funds will be redeployed to provide teacher training.

Although involvement in such activities as "unpacking" the CCSS, participating in alignment activities, and developing scoring rubrics can certainly enhance educators' familiarity with the CCSS, it is far too much of a leap to assume, as PARCC evidently does, that involvement of educators from Partnership states in the design and development process will result in expertise related to designing and implementing instructional practices that help students develop the skills and knowledge articulated in the CCSS. Targeted support and professional development is very likely to be needed in order for teacher-participants to take away meaningful and useful skills and understandings that they can apply in instructional contexts.

This section of the proposal contains descriptions of a sequence of training programs intended to build the level of expertise required of educators in different roles. The intent is for these programs to be administered online, supported by resources available on the Partnership's Web site. The programs will run from a course on the assessment system "basics" (what, how, why, what then) to the essentials of test administration (including related uses of technology) and a more targeted training for those responsible for administration. Field testing and piloting are planned in a timely manner to inform subsequent training and ensure that teachers and administrators are prepared for the first operational administration of the assessment.

Besides training tools to prepare school system personnel to implement the assessment, the Partnership plans to develop and provide training tools for interpreting and using results. Under this heading, the proposal includes the development of training tools to enable them to score student responses, and it is in this regard that various concerns arise. According to the proposal, "upon successful completion of this training, they will become certified scorers." This statement implies that scoring training is a one-time event, when virtually every judgment-based assessment scoring enterprise (whether formative or summative, high stakes for individual test-takers or not) requires not only retraining prior to each event, but ongoing recalibration throughout the duration of scoring. The statement thus demonstrates an oversimplified view of the rigors of scoring. In addition, the unofficial scoring to build capacity may not deliver the promised benefits; experience from various state and district-wide assessments employing teachers as scorers points to many opportunities for oversimplification and/or misinterpretation, which only becomes more pronounced when information is passed from those teachers serving as official scorers to their colleagues who have not had that first-hand experience.

The Partnership proposes to offer additional tools and assistance to member states to help them understand and act on curricular and instructional implications of the assessments (and the CCSS with which they will be aligned)—something that is of merit and demonstrates sensitivity to state and local spheres of responsibility.

It is curious, and perhaps of more than a little concern, given the intention to deliver these various training protocols online, that no mention is given of professional development to ensure adequate capacity to deal with technology as it is infused throughout the proposed assessment system. Recent studies of

instructional technology literacy among teachers and administrators and up-to-date reports from various sources about access to technology in schools throughout the U.S., all make clear that what has been described as a "digital divide" still exists in American schools—not only in terms of student access and knowledge, but that of teachers and administrators as well. Training will only be as effective, ultimately, as educators' facility with technology permits.

The Partnership recognizes the need for a timely and comprehensive public outreach and stakeholder engagement plan. PARCC gives evidence of the intent to be proactive, particularly by identifying and addressing likely misunderstandings among parents, educators, policymakers, and the public at large. Their examples of both anticipated misconceptions and system selling points are right on target.

The communications strategy set forth by PARCC balances what they describe as "broad based momentum" state-level efforts that are specifically tailored to in-state audiences. They plan to time outreach based on the various phases of the project in a manner that suggests that communication will be proactive rather than reactive. While use of electronic media is laudable, real life conditions and differing degrees of access among various stakeholders and the public suggest that multiple modes of disseminating information are well advised.

Outreach is also intended to be timed to build support from the earliest stages in the assessment design and development process, and examples of "critical junctures" are included in the narrative. Considerable attention is given to targeted outreach to parents which will emphasize college and career readiness, a strategy that seems promising. Nevertheless, given the likely perception by teachers, parents, and others that this assessment system will consume more instructional time than ever, it will be critical that PARCC not underestimate the sources and scope of resistance to the proposed assessment system and that they be proactive in this regard as well.

Acknowledging the critical role that higher education plays in the meaningful use of assessment results, PARCC sets forth a four-year engagement strategy for those stakeholders. This strategy includes soliciting regular feedback from members of higher education associations and involving them to ensure alignment with the work of the Partnership. The opportunity to serve on working groups is also anticipated to ensure that the assessments are used as intended, based on PARCC's theory of action. The Partnership anticipates that participants in working groups will in turn engage their colleagues in activities and discussions that will lead to institution-level engagement—a worthy goal, if perhaps just a bit too optimistic.

PARCC's plans are generally feasible, clearly consistent with their theory of action, and can be implemented in a cost-effective manner; therefore the response was scored in the low end of the "high" range.

(A)(7) Technology Approach

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

(A)(7) Reviewer Comments:

The Partnership very accurately describes their technology approach as ambitious—and indeed it is. More uncertainty surrounds whether their approach is feasible, in the allotted time, and what alternatives have been considered in the event that one or more elements of their technology plan cannot come to fruition as scheduled.

PARCC acknowledges up front (and must be given credit for doing so) that their goal—a 100% computer-based administration for most grades in 2014-15 will require many states to "accelerate efforts to expand technology infrastructure in their schools." This appraisal mirrors the most current figures on technology resources in U.S. schools. How this acceleration will (or should) occur is not addressed in the proposal, however.

PARCC here, and elsewhere, in their proposal links this not just to later cost-savings but to the ability to not limit the types of items that can be included in the computer-based assessments, their proposal never addresses whether it is actually advisable for all products and performances to be created using technology. Until such time as the review of the CCSS confirms that all standards can be accurately and fully addressed through online assessment, the full technology target may not be appropriate.

PARCC seems to be quite confident that market forces alone will ensure that sufficient innovation and improvement in existing technologies occur to use technology in the many ways envisioned in this proposal. Perhaps this is feasible in the long-term; whether or not it is in the window of time established under the terms of the grant is another matter.

Some proposed uses of technology are quite reasonable, based on the current state of affairs—among them, development of “next generation” data management and reporting systems using an inter-operable platform approach. Here, the Partnership’s support for modular and open-source solutions makes sense. In the areas of item/task development tracking and banking, as well as in technology in support of test security, the Partnership’s goals appear realistic.

While it is commendable that the Partnership plans on closely monitoring feasibility related to technology-based features of its proposed system during the design and development phases of work, doing so provides no assurance that high-quality, efficient, and/or cost-effective solutions will be developed (especially by 2014-2015). Identifying solutions already developed and those that need to be developed “from scratch” does not ensure the system can or will be brought to scale as intended.

PARCC’s list of general requirements for the delivery of an online assessment system by its technology development partner(s) is a good one, and demonstrates the preliminary thinking that has taken place.

In regard to the technology tools the Partnership intends to use:

- As far as item development, technology can support the creation and scoring of many innovative item types; it remains to be seen if these alone “measure the full depth and breadth of the CCSS,” however, the delivery and response modes selected must reflect the whole range of ways students use technology (e.g., and not advantage those who use a mouse over those who use a touchpad, as just one example).
- Technology will clearly facilitate project management and the conduct of work by many committees.
- Item management at all phases (development, review, piloting, etc.) for both secure and public release items can certainly be facilitated by technology; no attention has been given in the Partnership’s proposal, however, to the significant need for human management of item development and tracking systems by content, assessment, and technology experts.
- Test administration via technology is less problematic, but attention will need to be given to the efficacy and impact of tutorial sessions, since students’ at-home and schoolhouse use of technology may be extensive but sufficiently different than uses required by the assessment that results are compromised. The Partnership presented some great specific ideas (e.g., for an online dynamic calculator and tech-based accommodations) related to test administration.
- Scoring is one of the most problematic uses of technology proposed by PARCC. SR and BCR items lend themselves to scoring via computer; AI scoring of ECRs and performance tasks are another matter altogether. Simply issuing an RFI and pushing the field to develop a better system than currently available is not likely to advance this technology sufficiently to support PARCC’s assessment design. However, technology does support hand-scoring by trained readers in terms of providing an efficient system for distributed delivery of all components—new responses as well as those embedded for training and performance-monitoring purposes.
- Data management and reporting (along with data availability) are clearly areas that can be improved through judicious use of technology.
- Professional development can be facilitated through the development and provision of online training modules, recognizing that the still-existing “digital divide” described elsewhere (and documented by reputable organizations and agencies) may mean that not all teachers and instructional leaders may have access to PD resources. In the section of the response that attends to ways various implementation barriers will be addressed, PARCC identifies various outreach and funding

opportunities that may alleviate some of the gaps and inconsistencies in schools' technology resources.

On the basis of these observations and concerns, the response received a score in the high "medium" range.

(A)(8) Project Management

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

(A)(8) Reviewer Comments:

Central to the Partnership's management plan is their selection of Achieve as the project management partner for the consortium. While a relatively young organization (founded in 1996), Achieve has rapidly built a solid reputation (referenced in their application) as an educational reform organization engaged in important work in the area of standards, assessment, and accountability. Achieve has worked by way of Alignment Institutes with many of the Partner states, as well as others, to revise and refine their standards, curriculum and assessment instruments. Their commitment to college and career readiness is demonstrated through many projects and partnerships in which they are, or have been, involved (e.g., the American Diploma Project), and they have a proven track record of managing and working with multi-state consortia-based projects. Of most relevance, perhaps, to the issue of qualifications as a project management partner is Achieve's partnership with the National Governors Association and CCSSO to develop K-12 standards in English and Mathematics that are rigorous, research-based, and support college and career readiness. The detailed account in Appendix (A)(8) of Achieve's qualifications serves as assurance that they are up to the challenges of serving as PARCC's project management partner.

The curricula vitae of key personnel demonstrate their substantial direct experience in relevant projects. The lists of selected publications generally demonstrate their contribution to ongoing "conversations" about school restructuring, educational reform, and college and career readiness. What is far less evident is the extent to which they have designed and participated in related research studies (and lists of publications reveal relatively few peer-reviewed, academic publications); this does not argue against their management capability but does have bearing on which of the options for carrying out the research and evaluation program (see [A][5]), specifically, whether or not to "enlist the project management partner to implement some subset of the studies." It should be noted that the proposed percentage of time (20%) dedicated to the project by two key staff members—Matt Gandal and Laura Slover—appears insufficient given the scope of their duties as well as the many ways that, based on their experience, they can support the work of the project.

It was good to see that Achieve is realistic about the need to expand capacity to be successful in an endeavor of this magnitude. Their proposed blend of existing and new staff into integrated project teams (supported by expert consultants as well) should ensure delivery of expected services. However, the proposal left unclear what will be expected of the many team members still to be hired; more detail as to their anticipated roles and criteria for their selection would have added strength to this proposal.

Achieve's plan to convene both live and remote meetings and conferences is both efficient and cost-effective. Plans for reporting are sound but seem incomplete—there being no mention of documentation and dissemination of virtual and live meeting notes, an important resource.

It is in the area of risk management that the proposal is rather general. Aside from the intent to monitor progress and make "mid-course corrections and adjustments as necessary," there is little besides mention of general categories in which adjustments might be needed (e.g., assessment design, development process) to demonstrate that there has been sufficient advance thinking about not only what significant aspects of the project are most prone to risk but of possible directions that mid-course corrections might take. It is unclear from the narrative what role Battelle will play, and missing from the proposal is any sort of documentation supporting use of Battelle's services (e.g., qualifications, key personnel, etc.).

The Project Timeline makes clear the intent of the Partnership to begin work immediately upon award of the grant. Some key observations about the timeline are:

- A rather narrow window is indicated for both developing the test blueprint and establishing common policies and procedures.
- Initial item development will be underway at the same time that the technology platform is being developed. Assuming this involves item delivery and computerized scoring of some items, it is not clear that issues and opportunities that are identified during development of the technology platform can adequately inform test development.
- If the intent is for established technology capability to inform continuing item development, then it's unclear that sufficient innovative items/tasks and scoring strategies required by them will be fleshed out in time for field testing according to this timeline.

The Partnership took an honest and bold step of admitting that the variables that will impact ongoing costs—particularly those related to scoring—make it impossible at this point to make a specific state-by-state estimate for ongoing administration costs. The range of costs presented appears realistic, but the suggestion that teacher-scoring would lower overall and per student costs is only sound if the intent is for teachers not to receive any stipend for scoring, something that runs against the experience and precedents set by at least some of the Partner states. In fact, in some states, hiring of teachers to conduct scoring has been more expensive than using outside readers hired by a scoring contractor. Thus, the assertion that for the majority of states in the Partnership, the estimated costs of the new assessment would be less than or equal to their current assessment systems seems insufficiently justified. Indeed, it is evident from their proposal that the Partnership “takes the financial sustainability of the proposed assessment system very seriously”; regardless of best intentions, there is simply a great deal that is still unknown that may impact sustainability.

The Partnership’s budget clearly identifies Level 1 and level 2 budget modules; aside from the concerns noted that relate to scoring (both costs of live scoring and costs associated with programming/“training” needed to prepare for AI scoring), the budget appears adequate to support the development of an assessment system that meets the requirements of the absolute priority. However, there remain clear issues related to sustainability over time due to uncertain funding. Overall, costs appear reasonable in relation to the objectives, design and significance of the project, as well as the large number of students who will be served. Questions/concerns may be raised about the fact that there appears to be no allowance for any expense other than travel for such individuals as invited experts and speakers (under Governance), invited speakers or technical partners (under Assessment Design and Development) and guest experts at the Technical Advisory Committee meetings (Research and Evaluation). Without evidence that leading experts will donate their time, it seems appropriate to have budgeted for at least a modest honorarium in addition to travel expenses. Stipends for these individuals may be covered—at least in part—by the funds earmarked for expert consultants in the Project Management module, in which case per diem figures would have been helpful. Based on the description in this module of the duties in which the expert consultants will be engaged, the line item does not appear to refer to the outside experts mentioned in other modules.

Although the management plan is likely to result in implementation on time, there remains some question about whether this will be within budget and in manner that is financially sustainable over time; for this reason and based on other observations, this response was scored in the high “medium” range.

Suggestions to DOE:

Day-to-day project management of PARCC is to be the direct responsibility of a director who will be hired only after the grant is awarded. Given the importance of this role, the Department would be well advised to seek confirmation of that individual’s qualifications (just as the qualifications of the project team as anticipated at this point—presented in the Summary Table for (A)(8) are subject to scrutiny).

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>PARCC provides strong evidence of commitment on the part of IHEs in their member states. The Partnership far surpassed the minimum requirement that at least 10% of the direct matriculating students in public IHEs in member states be represented, securing commitments that represent 90% of direct matriculation students in those states. Although not required, the Partnership also received letters of support from a number of private IHEs from four Partner states wishing to participate in this collaborative venture. In addition, their supporting materials include a very strong letter of support from SHEEO (State Higher Education Executive Officers), the executive committee of which is comprised of officers from non-Partnership as well as Partnership states.</p> <p>Careful thinking about the important roles that IHEs must play in the Partnership is evident in their strategies for building engagement among higher education stakeholders in the design and development of the assessment system (including, critically, setting college-ready standards and validating CCR scores as a reliable indicator of students' readiness for college). Particularly striking is the intended effort to involve a broad constituency through activities that engage faculty within and across IHEs, not just leadership (e.g., chancellors and college/university presidents) or a select group of faculty representatives.</p> <p>There is also merit in the Partnership's intent to build a "sustainable platform for collaboration" that can support other instructional improvement efforts and facilitate meaningful policy changes.</p> <p>It would have been helpful to have some clarification of why the direct matriculation students in participating IHEs as a percentage of the state total was so low in comparison to all other states in the Partnership in the case of Mississippi and New Jersey (e.g., which public institutions are not included, and for what reason[s], although one may surmise, based on review of the appended letters of intent, that often various community colleges did not choose to participate). The absence such detail is not critical, however, since support from the vast majority of Partnership states' IHEs is considerable and the applicant has provided letters of intent that represent far more than the minimum 30% of direct matriculation students in public IHEs across the consortium as a whole. Therefore, the applicant received full competitive preference points under this priority.</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>PARCC has presented clear and fairly detailed plans for the development of a new assessment system to be used by multiple states to ascertain that students are on track to demonstrate or have demonstrated readiness to enter college or the workforce with the skills and knowledge needed to succeed in post-secondary education or in a 21st century career.</p> <p>Based on their application, the assessment system proposed by the Partnership should be able to cover the full range of college and career-ready standards, through items and tasks that elicit complex demonstrations of knowledge and skills. The potential obstacles—which they have generally considered but still leave in some doubt the ability to overcome—relate to some of their intended uses of technology</p>		

and also the capacity of states, through existing funding, new funds, and leveraging of shared resources to sustain this program once it is operational.

Furthermore, while the proposed system appears to be able to support accurate measurement of student growth over the course of a full academic year or course, information in the proposal that addresses measurement across the full performance continuum is somewhat general; it would have been helpful to learn more about any additional strategies supported by this system that might be implemented towards this end (e.g., use of off-grade assessment). This is particularly critical since PARCC proposes a fixed form model.

The proposed system provides for both summative and through-course assessment of English language arts and mathematics administered at grades 3 through 11; in addition, the Partnership seeks support, through a Budget Module 2, for the development of formative assessment resources for grades K-2, which has the capacity to initiate thinking about students' "on track" status in meaningful and appropriate ways and create more cohesiveness in instructional communities across K-12.

Much of the focus in this application on assessing all students focuses on plans to establish common definitions, policies and procedures for English learners and students with disabilities. The Partnership provided much more detail regarding some specific strategies possible (including reference to support their thinking about accessibility) related to students with disabilities. It would have been helpful for them to demonstrate in more detail their thinking about English learners.

The Partnership's proposal provides sufficient detail to instill confidence that this assessment system will produce data that can be used to inform determinations of school effectiveness (accountability data) and in support of teaching, learning, and school and program improvement. In addition, it can help identify and inform professional development for teachers and instructional leaders. While the proposal does not provide sufficient information to say with confidence that the data produced can be used to make determinations of individual teacher and principal effectiveness, the Governing States in the Partnership have committed to using results from the system in this way and assurances are expressed in the proposal that the data will allow for this use. The Partnership demonstrated some thought about the issues and implications associated with this anticipated use of data, manifest in remarks about the benefits and drawbacks of teachers serving as raters, even by means of a distributive mechanism whereby they did not score their own students' work.

Overall, the proposal meets the Absolute Priority.

Suggestions to DOE:

As evident in comments across all sections of this review, there are genuine and serious issues related to the extent to which technology will be able to support all aspects of the proposed system, and if not, what fallback position the Partnership is prepared to take to ensure that the assessments become operational within the intended timeframe. If a cooperative agreement is entered into with this entity, this is a topic about which there should be discussion between the Department of Education and representatives of PARCC.

Grand Total	220	166
-------------	-----	-----

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

As with virtually all collaborative efforts, the Partnership expects the combined efforts and resources of Partner states will lead to economy of scale and support the sustainability of the assessment system over time. At present, however, there remain many unknowns—particularly in regard to the true costs of

technology—to instill confidence that as claimed, states will for the most part wind up spending the same or less on assessment than they do right now.

In addition, there do not appear to be funds earmarked for a number of services that will be required in the design, development, and implementation of the assessment system. It is not clear whether or not the various experts anticipated to be involved throughout the term of the grant would receive any sort of remuneration. It would appear from the narrative throughout that invitees to various meetings/working groups may not always be included among the expert consultants identified in the budget for project management. In any event, some detail to serve as rationale for the anticipated cost (listed under Contractual) would have been useful.

The true costs of having a portion of student responses scored by teachers or outside raters (as anticipated) are also not sufficiently clear, and indeed may not be able to be accurately ascertained until a determination is made of policies that impact the amount of training, scoring, recalibration/retraining, and rescoring (e.g., policies governing qualifying scores for raters, accuracy and productivity targets for raters, etc.).

The budget section does contain a number of sound ideas and plans, including the intent for States to have discretion in the use of funds provided to them, and plans to share existing assessment tools and technology.

The major item in the budget is approximately \$80,000,000 to support one or more vendors in the design and development of the assessment system. It is difficult to confirm or challenge this figure, since the figure is based on a proprietary assessment cost model provided by an organization, Assessment Solutions Group, hired to assist in the estimation of the price of the proposed assessment system. While undoubtedly extensive data is available on costs related to more conventional item/task development and scoring, one may question how much testing industry experience has informed projections of costs of developing innovative items using new or yet-to-be determined formats and technological features or the costs of not only operating AI scoring but conducting all the preparation necessary to program a system to address varied and often complex performance tasks. Successful uses of AI scoring for which information is more readily available involve scoring of writing samples generated in response to fairly formulaic prompts developed in anticipation of use of AI scoring. Large scale applications of AI scoring in place today such as scoring of the GMAT were not without considerable glitches that had to be ironed out. There is no detail in the proposal to indicate whether the budget is sufficient to address glitches in scoring of components of this assessment system.

Throughout, estimates for different costs are made in consideration of such factors as changes in the number of States (and the number of those in different roles), fall-out rates for work produced, cost-of-living increases, and other variables.

It would have been helpful to have more information about scoring costs, including a more detailed breakdown of specific costs and a comparison of the cost of hand-scoring using outside raters versus teachers, and using a scoring site based model or remote scoring (or some combination of both).

Level 2 Budgets

Name: Formative Assessment Tools for K-2

The development and dissemination of formative assessment tools for K-2 is certainly a worthwhile project deserving of funding. The Partnership identifies various developmentally appropriate measures generally recognized as being appropriate tools for assessing learning in early childhood. However, the budget breakdown addresses under Content Development only the cost of developing "the items," leaving unclear what consideration was given to developing observation tools, checklists, and on-demand performance events or specifically those components of a formative assessment toolkit for K-2 that incorporate use of technology. Curiously absent, as well, considering that this project is intended to support ongoing formative assessment of student learning, is any description of—or identification in the budget for—resources such as a bank of sample annotated responses (graphic, written, and/or filmed records of products and performances). The scoring of "items" for field test purposes will produce a wealth of material that can be turned to

meaningful uses in the classroom; it would have been good to see those accounted for. As presented, this budget module supports only partial development of formative assessment tools for K-2.

Name: Technology Enhancements

The technology enhancements proposed are certainly worthwhile, although there is no detail provided that would demonstrate that the funds requested for this purpose are a sufficient means to the ends described in the explanation. It is curious that, given the more immediate importance of ensuring that system load is addressed and that technology-based scoring methodologies are up to the task of yielding valid, reliable results, the Partnership has made PARCC technology enhancements Priority 2 rather than Priority 1. Of concern is the fact that issues with the "current state of the art" of AI scoring are somewhat obscured by making note of them only at this point in the proposal (and it should be noted that the current specific limitations are far more in number and type than described therein). Finally, one might challenge the reasonableness of aiming to enable 100% of through-course tasks to be AI-scored, since there has of yet been no detailed analysis of which standards lend themselves to item/task types amenable to AI-scoring and which, if any, can best--or only--be measured through students' creation of products and/or performances that require human raters apply criteria to make scoring decisions.

Name:



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b) (6)

(A)(1) Consortium Governance

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

(A)(1) Reviewer Comments:

Evaluation:

PARCC's proposed governance structure provides a significant level of confidence that it will enable the successful design, development, and implementation of an assessment system aligned to the CCSS standards and consistent with the NIA's goal of improving student outcomes and its accountability, interstate comparability and instructional-improvement goals. In the MOU, all States acknowledge the multiple purposes of the common assessment system, including the use of its results for accountability purposes (including common standards and measures of student growth, "on track" status, and college/career-ready graduation) and, more generally, to improve student outcomes. Governing States commit to use the results from the proposed assessment system in their state accountability systems, including for determinations of school effectiveness and teacher and school leader evaluations. Those States also commit to provide all of the decision making and staff support required to ensure the success of the Partnership's work plan. Important concerns arise, however, including that the extensive decision making role and responsibilities placed on the Governing Board, which meets only quarterly, could delay crucial work on the project, and that there is no clear leadership or coordination mechanism for the design and work groups that (as the application states) "will do the heavy lifting necessary to design and develop a sophisticated common assessment system." Overlapping this concern is a question whether the decision making division of labor between the Governing Board and the design groups, which is premised on a difference between "major" and other decisions, will succeed in practice, given the literally thousands of operational decisions that will arise along the way. Recommendations are made below to ease these concerns.

Explanation:

* The Partnership's membership (26 States, 11 Governing, 15 Participating) is broadly representative of the US in geographic distribution and urban/rural demographics and includes both of the initial Race to the Top States (DE and TN).

* The Partnership's vision, goals, role, and key deliverables are clear and concise, focus on and (including via the MOU) clearly commit Governing States to make both accountability and instructional-improvement uses of the proposed new assessment system, and are consistent with the Partnership's theory of action. The accountability uses include common achievement (including growth) standards and mechanisms for evaluating schools, principals and teachers based on assessment outcomes. The vision and MOU describe a package of common standards (CCSS), common "through course" summative assessments, common achievement standards, technology (for delivering, scoring, moderating, and reporting results of assessments), and the building of educators' capacity to use the data from the assessments to improve instruction. That package is consistent with the NIA's goals and holds out a prospect of substantial improvement of existing systems and, as a result, improvement in student outcomes, most notably, students' completion of high school college- and career-ready (CCR). The assessments are designed to cut

across the full range of CCSS standards, and the Partnership commits to meet the difficult challenge of effectively assessing students across the full range of achievement levels and to provide complex demonstrations of knowledge and skills -- a challenge that is especially difficult because of the Partnership's decision to use approaches other than computer-adaptive technology.

* The Partnership's package includes no interim or formative assessment components, although its through-course structure and the diversity of different test components intentionally provides many of the benefits of those other kinds of assessments and may do so in a more thoroughly aligned manner than those assessment typically accomplish.

* The MOU is detailed and comprehensive, and it repeats (indeed, in many cases it exceeds) the detail provided in the proposal itself.

* As is confirmed by the MOU, the organizational structure of the Partnership and the differentiated roles, rights and responsibilities of member States are clear. As is set out in the proposal and MOU, there is a logic behind the differentiation of roles: States that agree to be Governing States make significant commitments in leadership and staff time and, in return, have decision making authority. Participating States are provided regular opportunities for input by, and the provision of information to, their leadership (one meeting a year for political leaders; one for Chiefs only) are encouraged to participate fully in all work processes, and are encouraged and given incentives to become Governing States and increase their commitment of staff.

* Governing States commit to provide the staff necessary to complete the work of the Partnership. Also of note are the explicit commitment of all States to the time lines set out in the proposal; the commitment of all States to participate in piloting and field testing; and the commitment of Governing States to identify what amounts to a state project manager responsible for coordinating all of the work and staff commitments and district and school contributions (e.g., towards piloting and field testing) that are generated by the work of the Partnership. (PARCC's (A)(1) section reveals no intention to use member State educators to write or score items, which passes up an opportunity for educator input and embedded training.)

* Entrance and exit protocols are clear and simple. Exit occurs automatically following the passage of a set period of time after a State notifies the Partnership of its intent to exit and provides a reason. A recommitment process occurs for any State that undergoes a leadership change at the Gubernatorial or Chief level, providing an important mechanism for ensuring the proper prioritization of the Partnership's work as leadership changes occur.

* The Partnership's method and process for making different types of decisions are clear and in many cases are likely to be effective. The Governing Board is composed of a Chief-level representative from each Governing State is led by a chair elected for one year and already in place. The Governing Board meets quarterly and makes decisions by consensus if possible and, if not, by a majority-plus-one vote. This super-majority requirement can be changed at any time by a vote that satisfies the existing voting rule. The Governing Board creates and charters "design" groups made up of staff from member (probably, mostly Governing) States. Design groups' decision making is governed by bylaws and charters and is generally divided into two kinds of decisions. "Major" decisions are in the form of consensus proposals to the Governing Board; if no consensus is reached, the committee provides the Board with options. Other committee decisions are by consensus or, if none is reached, by a vote of a majority of Governing States plus one. Decisions are memorialized (an important step, given the vast number of decisions and overlapping work of the various committees).

* Although clear and well thought out, the proposed decision making structure does raise some significant concerns.

(A) The proposed decision-making process contemplates a series of policy or design decisions. (Indeed, the committees are called "design" committees.) This structure likely will prove less workable when the work switches from design to actual development, at which point the number of activities that could be called "decisions" and thus would be subject to the consensus/majority-plus-one requirement will rise precipitously. At that point, there may have to be deference to a committee leaders or to some other designated person or intermediate coordinating committee in order to enable the Partnership to make the

decisions needed to push the work along. Currently, no such leaders or intermediate leadership structure is provided.

(B) The Governing Board meets only quarterly (for two days). As the quantity of "major" issues increases (or as questions arise as to whether issues are or are not "major"), and given the speed with which the Partnership hopes to complete its work, it could be disruptive to have to wait as long as three months for major decisions to be reviewed by the Board. Often, the development of policies or designs will require a sequence of major decisions, each one dependent upon the resolution of the ones that occur before it. For example, the (A)(1) timeline identifies a large amount of work that must be completed between Fall 2010 and February 2011. There likely will be only one Governing Board meeting during the bulk of that period, and if a number of intermediate decisions get held up pending Board action, it is unlikely that the crucial February 2011 RFP deadline will be met.

(C) Even more challenging than the division of labor between the Governing Board and the committees will be the inevitable division and overlap in labor between different committees: technology often will depend on assessment policy, which will depend on accommodation requirements, which will depend on research findings, etc. There is no obvious mechanism described in the proposal for resolving day-to-day disagreements and removing dependencies and obstacles arising between committees. The same may occur between committees and work taking place in States.

(D) For these reasons, the absence of some kind of leadership structure for each of the committees and the absence of a full-time coordinating body below the level of the Governing Board with a designated leader responsible to the Governing Board (e.g., a full-time Coordinating Committee with representatives from the various design committees and with an operationally responsible chair of the committee) is a source of significant risk. Likewise the absence of a coordinating body to interact with the State-level coordinators who are contemplated in the MOU is a source of governance risk.

* The Partnership's plan and timeline (no "process" is specified) for setting key policies and definitions for the proposed assessment system is sensible, and States explicitly commit to it in the MOU. The idea is to complete many of the policy decisions affecting assessment design and administration, including accommodations policy, during Fall 2010 and early winter 2011 in time to be included in one or more RFP to be issued then.

* Subject to the concern above about sequences of important decisions required so quickly over a period when there will likely be only one Governing Board meeting, the Partnership's goal of making these decisions early is laudable. The same goes for the goal of phasing, and getting started quickly on, the processes for setting common performance level descriptors (prior to field testing) and for setting common achievement standards.

* Waiting several months before launching the key RFP has some risks in terms of meeting timelines thereafter, making it important to monitor the development of that document closely to be sure it is released on time.

* The Partnership's plan for managing funds is solid, and clarifies the primacy of the Partnership over the Fiscal State (Florida) when it comes to major financial decisions.

* The Partnership's procurement process, and evidence of each member State's commitment to that process, is solid as well and is laid out with good detail in a separate addendum in the MOU. The commitment to "best value" procurement should allow the Partnership maximum capacity to balance quality against price, and the absence of any evident constraints (subject to undisclosed aspects of Florida law) that, e.g., forbid the use of penalties, incentives or up-front (as opposed to upon-delivery) payments also provides needed flexibility.

* Each State commits to identify and remove barriers in state law to deploying the proposed assessment system prior to full implementation of the summative assessment system. The MOU provides States with an opportunity in Addendum 2 to go beyond that commitment and to describe their plan for transitioning from current standards and summative assessments to new ones, including the removal of barriers to doing so, and all but four States (CA, IL, OH, PA) took the opportunity to do so, most with considerable care and

thought (see, e.g. AZ, FL, KY, OK and TN). One State (MA) expressed reservations about the CCSS and the common assessments, but on the understandable ground that it does not want to commit to adopting standards or using assessments less rigorous than those currently in place in the State. The current chair of the Board of Governors is from that State, dispelling any indication of a lack of commitment. No steps are indicated here, however, through which the Partnership will monitor the timeliness of, or assist, States' efforts to remove barriers.

* The Partnership also recognizes the possibility of, and leaves to States to flag, the need for US/ED waivers in order to effectuate the new accountability scheme. This is an important matter because of the possible need for waivers to permit off-grade testing.

* The RFP clarifies Intellectual Property rights among the States and the Partnership.

Recommendations:

* Consider creating a Coordinating Committee between the design committees and Governing Board, that (1) is led by a single individual responsible to the Governing Board and member States for meeting the Partnership's deliverables, (2) has representatives from the work groups, and (3) has a capacity to make day-to-day (especially operational) decisions.

* Consider creating a leadership structure on the committees to assure that decisions can be made swiftly as necessary to keep the work going, and to assure proper coordination and removal of dependencies between different committees.

* Especially if the first two recommendations are not adopted, monitor carefully the division of decision making authority between the Governing Board and the committees, especially in the crucial period between Fall 2010 and February 2011, to be sure that uncertainty in regard to what decisions are and are not "major", and the infrequent meetings of the Governing Board, are not holding up the work.

* Periodically survey States' progress in removing barriers, and provide a clearinghouse mechanism through which different States can learn from the obstacles identified and barrier-removal strategies used in other States.

(A)(2) Theory of Action

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

(A)(2) Reviewer Comments:

PARCC's theory of action is logical, coherent, credible and comprehensive and has a good chance of improving student academic outcomes.

The theory of action attempts to maximize the extent to which the summative assessment system is formative in two ways. First, it gives educators information multiple times a year and uses assessment tasks to model good instruction, so that "teaching to the test" means superior and improved teaching that develops student learning across the full range of rigorous standards and develops content knowledge and higher-order thinking. Second, textured information is frequently provided to educators, students and parents about where instruction is working for particular children and where it needs to be improved.

A combination of through-course assessment and end-of-year testing is designed to maximize the capacity to adequately measure student work across the full CCSS spectrum, and to provide enough ways to demonstrate different levels of proficiency to measure differences and longitudinal changes in student attainment across the full spectrum of students' initial achievement levels. Meeting the latter goal is a challenge in the absence of computer-adaptive techniques, and may require off-grade testing and/or the lengthening of the end-of-year assessment, but is consistent with the structure of what is proposed.

The overall approach -- and the emphasis it places on common and rigorous standards and on common, easily compared measures of year-to-year achievement, growth, and whether students are on track to graduate -- also recognize the key contribution accountability makes to improved instruction and to the public's willingness to support public education, and the ways in which better assessment information can improve accountability systems focused on States, districts and schools and evaluation systems focused on principals and teachers.

PARCC is sensitive to the needs of parents (e.g., in promising a timely end-of-year indication of how each child is doing), policy makers and the public. The role of higher education is not an add-on to the theory or action. Instead, it is fully integrated both from the standpoint of designing the right assessments and accountability measures and from the standpoint of measuring the effectiveness of K-12 schools by the extent to which they generate graduates who college admissions officers and faculty consider to be prepared for college.

Technology is properly incorporated as a tool, not an end in itself. PARCC also notes the role of enhanced curricular supports and training, but is modulated in its description of the extent to which this project and grant program can satisfy the new needs in those domains.

Finally, the theory of action acknowledges the key places where it is pushing out the envelope the most, and provides a good explanation of why the risk associated with that kind of innovation is warranted and what fall-back measures will be used to manage those design risks.

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	46

(A)(3) Reviewer Comments:

Evaluation:

The design of PARCC's assessment is innovative and designed in a rigorous and ambitious manner that takes the Partnership's theory of action, and the promise of the CCSS generally, to their logical conclusion. Where adopted, it would not only substantially increase assessment quality and accountability but also substantially upgrade classroom practice and have a strong chance of significantly improving student outcomes. PARCC's goal is to go beyond the idea of "aligning" standards, assessment, curriculum and classroom practice to the point of making them more or less the same thing. The processes of preparing for, taking, and evaluating the results of assessments would be rich curricular exercises -- occurring throughout the year -- in higher-order, content- as well as skills-focused, learning for students and educators. The system thus intends to embed substantial "formative" and "interim/predictive" capacity directly into its summative assessments. The system is designed not only to cover the full range of CCSS standards but to focus carefully on the proper balance among tested skills and knowledge. Also commendable is the commitment to provide annual assessments throughout the high school years and to determine whether computer testing for young children is feasible without distortions in capacity to measure what they know.

A key challenge for the Partnership is meet its commitment to measure very high- and low-performing students without using computer-adaptive techniques. The proposed system has a number of options for meeting this commitment -- e.g., off-grade test forms or an extended end-of-year assessment with many off-grade items -- but does not fully identify the strategy it will use in this regard. Additionally, the system's very ambition raises feasibility questions relating both to (1) the successful completion of the RTTT-A-funded project and to (2) the administration of the completed system in schools. The first of these concerns underscores the need for a fast and efficient decision making process (see (A)(1) comments) and highly experienced and professional project management (see (A)(8) comments). The latter of the two concerns

puts a premium on professional development and public outreach (see (A)(6) comments) and on serious consideration of steps to reduce substantially operational cost and complexity. Such steps could include, for example, extending computer administration and scoring to grades 3-5 and making some adjustments in favor of items that are relatively less expensive and less difficult to administer and score.

Explanation:

* Beyond being aligned to the CCSS standards, PARCC's assessment system seeks to infuse the standards into the ongoing curricular practices and learning outcomes of schools. The scope and sequence of the assessments is explicitly designed to influence and improve the scope and sequence of the curriculum, all with reference to achievement of the CCSS standards. Test items aim to model good curriculum and instructional practice. The frequent release of items and the transparency of the item prompts have a similar ambition. Comprehensive coverage of the standards is designed into the system.

* Because the system is simultaneously formative, interim, predictive and summative, its cost and complexity is properly compared not to existing summative testing, but to comprehensive packages (where they currently exist) of formative, interim and summative assessments.

* The system will permit a comprehensive set of outcome measures, frequently reported to different constituencies during the school year, on student attainment and growth, which the Partnership States intend to embed in their school-focused accountability systems and use – or at least have the capacity to use – in evaluating, assigning and developing principals and teachers.

* The system of common performance level descriptors, cut scores, achievement standards for college- and career readiness (CCR) and for being on track to CCR, and growth measures is impressive and consistent with the NIA. Of particular note is the goal of designing the summative assessment's weighted score based on scale scores with equal-interval properties. On the issue of growth measures, the formative logic of the PARCC's system – that educators should derive diagnostic information throughout the year from all data generated by the system – strongly suggests the need to adopt metrics that educators can themselves "reverse engineer" from available data (e.g., the underlying data in the interactive data tool) using simple arithmetic, rather than utilizing opaque regression analyses for this purpose. The Partnership appears to be considering both approaches, however. The Partnership's evident desire to facilitate cross-State, district and school comparisons that are not confounded by differences in the socio-economic status of students at each level can also be achieved by using growth measures (e.g., growth percentiles) that are insensitive to students' opening achievement level. The Partnership risks limiting the usefulness of growth measures by omitting them, and explanations of them, from its Periodic Feedback Reports.

* The impressive array of reports the Partnership envisions include tools enabling all categories of users to compare a variety of achievement and growth outcomes by Partnership, State, district and school (and in some cases classroom), and to slice and dice the data, including by relevant "NCLB" categories of students. Also impressive is the decision to provide data down to the item level, with interpretive information to help educators derive rich diagnostic information from patterns of wrong answers. The ambition of the Partnership's reporting strategy, however, enhances the concern expressed elsewhere in these comments about whether the Partnership has left enough time and budgeted enough money to translate the reports it wants into code, and from there into attractive and useable computer views and drill-down and customization tools.

* The Partnership's approach to accommodations – and to designing systems in the first instance to reduce the need for accommodations – is strong.

Some limitations arise from PARCC's system design itself or from particular decisions the Partnership has made or failed, as yet, to make:

(A) Because the proposed summative assessments are not themselves computer adaptive (although the Partnership contemplates a formative tool that is), the system faces a substantial challenge in measuring students at all performance levels. The diverse nature of the tasks called for by the constructed response, computer-enhanced and performance task items will help alleviate some of this problem, especially for high-performing students, but the problem will remain. The risks are increased by the

Partnership's evident focus on fixed-form assessment tools (as opposed, for example, to "testlets" to determine students' approximate performance levels and/or the use of alternate assessments for some students). On the other hand, the structure of the system the Partnership proposes encompasses the use of off-grade (fixed-form) assessments for some students, and the proposal explicitly contemplates the possible need for member States to seek waivers of the sort that would be necessary, given existing restrictions on off-grade assessment. The proposal's end-of-year assessments also may be sufficiently robust to include the range of items needed to measure achievement and progress among students well above- and below-grade. The proposal is not, however, as detailed as it could be in describing the mix of strategies it will use to meet its commitment to accurately measure achievement and growth for students who are very low- and high-performers.

(B) Combining the scores of the four math portions and four scored ELA portions into a psychometrically valid composite score will be a challenge.

(C) The system requires a substantial amount of distributed scoring – much more than summative assessments, even ones with the highest proportions of constructed response items, currently entail. This contributes to the relatively high per-student cost of these assessments, the large management burden of moving so many physical answer sheets around, and the time that will have to be spent scoring. The high amount of hand scoring makes it unlikely that the Partnership will accomplish its goal of having each through-course component reported within one to two weeks. Given that there will be multiple through-course assessments each year per student (several in math, even more in ELA), a decision to use teacher-scoring could require teachers to be out of the classroom for significant amounts of time. The Partnership plans to use its through-course system to help teachers gain valuable information while minimizing the burden of administration, "so teachers can focus on learning rather than testing." The concern here is that, in fact, teachers may spend a lot of time scoring papers from other schools. The extent of the distributed scoring is magnified by (1) the decision to commence computer administration and thus partial computer scoring only in grade 6; (2) a set of cascading decisions about item types – e.g., to make no use of selected response items in any but the end-of-year test and to use substantial numbers of performance tasks, rather than somewhat less costly and easier-to-administer constructed response and computer-enhanced items at various points; and (3) the Partnership's failure to reach agreement on a single human scoring approach (e.g., by teachers or by vendors), which makes it harder both to obtain full economies of scale if vendors are used, and out-of-state scoring if teacher scoring is used (out-of-state scoring is strongly preferred, to enhance reliability and comparability). Absent early and expeditious efforts to study and rethink some of these decisions based on pilots and research, significant feasibility issues may arise. Estimates of scoring time per test subject per year should be provided.

(D) The apparent decision not to make the Annual Stakeholder Reports available to the public and press is sharply inconsistent with the Partnership's theory of action.

Recommendations:

* Provide more detail on the mix of available strategies that the Partnership will use -- e.g., off-grade test forms, extended end-of-year assessment -- to meet its commitment to measure the achievement and growth of very high- and low-performing students.

* Consider early pilots or simulations, permitting expeditious studies to determine (1) whether computer administration is possible in grades 3-5; (2) the correlation in the learning information provided by selected- and constructed-response items, computer-enhanced items, and performance tasks, to reveal whether the current set of decisions in favor of more expensive and difficult-to-administer items has achieved the best mix of items; and (3) whether a higher proportion of items can be scored using Artificial Intelligence than is currently contemplated. Early information on these questions could save substantial amounts of money in development and scoring work and enhance the feasibility of the system.

* Encourage Partnership States to reach an early decision on a common approach to distributed scoring, which would simplify the development process, foster economies of scale if vendor scoring is adopted, and facilitate out-of-state scoring if teacher scoring is used.

- * Provide estimates of human scoring time per test, subject, grade level and year to inform decisions among item types and scoring strategies.
- * Release the Annual Stakeholder Reports to the public.
- * Report changes from year to year with standard error bands to notify users when changes are large enough to provide confidence that they are not random.

(A)(4) Assessment System Development

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

(A)(4) Reviewer Comments:

Evaluation:

PARCC has a detailed and complex plan for developing the proposed assessments and assuring they will be equally sensitive to the needs and actual performance levels of all, including special needs, children. In substance the plan is thoughtful and professional and has the components needed to ensure that the assessment content will adhere to PARCC's design and be ready for wide-scale administration on time in 2014-15 following pilot and field testing to allow the system to improve based on feedback. However, the plan's complexity, aggressive time line, and only modest attention to the difficulties of vendor management underscore concerns expressed in comments to (A)(8) regarding the need for more experienced project management. The plan for developing the technology components needed to administer, score and report on the assessments and for integrating the assessment content into the technology system is not well-specified. The same is true of the plans for developing scoring, reporting and quality control (QC) mechanisms (apart from item review, where quality assurance is appropriately specified). In all four of these latter cases, the proposal states in some detail what the new systems and reports will do, but not how they will be developed or (for QC) how they will be implemented. The process for continuous improvement over time is not specified in any detail.

Explanation:

* Using a combination of evidence-centered design at the system level, and universal design applied to particular components and items, PARCC has laid out an intricate and comprehensive plan, involving many task-oriented meetings of both content and technology experts and educators, design committees and review committees, for designing, producing, and rigorously reviewing assessment content for the various types of assessments and items the Partnership has in mind. The Partnership hopes to develop a relatively small number of task-design templates based on which item writers can develop many similar but varied tasks that students will be asked to undertake to demonstrate proficiency. The proposed process moves from unpacking the CCSS standards into granular pieces and linking each to different item types, to developing a test blueprint (including item types and maximum score points linked to each "piece" of the unpacked standards), and then to item acquisition and writing, item review, pilot testing, field testing, psychometric analysis of field test results, and alignment review to determine coverage of CCSS, hard to measure topics, and student performance at the low and high ends of the performance range). Simultaneously with the early stages of this process, an RFP or RFPs will issue and content vendors will be selected to provide existing items or to generate new ones. The goal is to produce eight parallel test forms at each grade level in ELA/literacy and math.

* The intention to take advantage of experts, including higher-education experts, who helped to develop the CCSS is a plus.

* Experts in special needs populations will be integrated into the design process to maximize the extent to which initial item designs are accessible to and effectively measure the proficiency of all populations and

minimize the need for actual accommodations. Review committees, pilot and field testing, statistical analysis and other mechanisms will be used to determine whether the goals in this respect have been met. Accommodations that continue to be required after this process will be developed and memorialized in an Accommodations Manual.

* Pilot testing of initial ideas (2011-12) and field testing (2012-13, 2013-14) will occur using stand-alone methodology with a large enough pool of carefully sampled students to compensate for the degradation of information that stand-alone testing invites (absent stakes, students tend to skip the harder questions).

* The proposal describes in some detail the Partnership's system requirements for technology, scoring (where less detail is provided because of the desire of members States for flexibility), reporting and quality control (QC). But the proposal does not set out, comparably to the processes described for assessment and accommodations design/development/review, the processes that are to be used to design and develop the technology, scoring and reporting systems to that will be used to implement the QC requirements.

* Concerns include the following:

(A) The intricate processes contemplated for assessment and accommodations design and review require a lot of committees – composed of individuals more or less evenly divided among numerous States (it isn't clear whether this means Governing States or all Partnership States) and among grade bands and various independent experts from a variety of sources including higher education – to convene, commit themselves to an intensive, task-oriented, time-constrained effort and complete work on short deadlines. The Partnership's stated intention to involve multiple competing vendors in developing items increases the number of moving parts. The initial design, development, review and pilot testing are to occur over 21 months from October 2010 to June 2012. The realistically available time period is considerably shorter, however, because item-development vendors won't be selected until the end of June 2011, leaving only 12 months for the great bulk of the work. The Partnership's plan to issue RFIs, in addition to RFPs, to obtain advice on item development, text complexity, and accommodations/accessibility adds time risk. Although the Partnership hopes to purchase some items, its strong preference for Performance Tasks and for complex Constructed Response and Computer-Enhanced items, none of which are in wide use today, means that most items will have to be developed, not acquired, and that item review will be demanding and difficult. (The Partnership sensibly plans to have as many as five committees review every item.) Organizing a large number of schools to engage in piloting, then field testing, will also be a difficult task, especially given the desire for a carefully stratified sample of piloting and testing schools and the reluctance of most schools to devote time to no-stakes developmental projects of this sort. All of this is to say that it is imperative that the Partnership have proven, experienced project management. This includes a strong capacity (from a governance perspective) to commit many individuals from States and school systems to complete quality work on time and experience assuring that multiple vendors (among whom spotty quality control is not unusual) do the same.

(B) The development plan is silent on whether the field testing is assumed to take place using the new technology system that is being developed for test administration, scoring and reporting purposes. Because the initial item development and the technology development processes are both slated to end on June 30, 2012, with field testing starting shortly thereafter, it is unclear whether there is time to integrate the two systems. Additionally, as is discussed in (A)(7), the one-year, post-contract timeline for developing the technology system is probably too short. A plan to permit field testing on systems other than the one being developed for this purpose may have to be developed.

(C) Concerns about the Partnership's failure to reach agreement on a common scoring approach and about the cost of so much human scoring are set out in comments to (A)(3). The proposal does not provide adequate information to evaluate the cost of 100% vendor scoring and 100% double reads, and the cost and time demands (both classroom time and overtime) posed by 100% scoring, read behinds and double reads by in-service teachers – all of which at least some States seem to believe is necessary to their adoption of the system. The impact of different methods on test security, reliability and comparability also is not developed. The proposal rightly notes that scoring "flexibility will need to be carefully monitored to ensure it does not negatively impact comparability across states," but no plan or research agenda for this purpose is set out. More generally, there is no description of the process to be used to develop, monitor,

and implement different scoring plans. Instead, basic requirements are stated in terms of the baseline amounts of human and AI scoring, double scoring and read behinds and the QC approach, but no development process is indicated.

(D) Likewise, for the reporting and technology streams, the what is described but not the how and when. The Partnership recognizes the need to access data from each State on students, teachers, principals, classrooms, schools and districts in order to provide the relevant reports, but does not describe the process for doing so. Based on experience, this process will be one of the most difficult tasks the Partnership faces, given varying data formats and given the failure of most States to have, e.g., unique identifiers for teachers statewide, even if they have them for students. Consequently, the Consortium will likely have to go through a process, not fully acknowledged in its proposal, for selecting data formats that are most compatible with those used by the Member States, and for each State to work with its districts and schools to conform their data collection to the new format. In the absence of information in the proposal, it has been assumed in these comments that the RFP(s) for the reporting and technology plans will be completed on the same June 30, 2011 schedule as the assessment RFP(s). Whether that schedule is consistent with the plan to consult national experts to provide guidance on platform design is not clear.

(E) The QC discussion also is focused on the what, not the how, and only covers assessment content and scoring, not reports and technology.

* The proposal has a thoughtful plan for pilot and field testing, and provides ample time for the latter, dependent upon its success meeting very aggressive timelines for procuring vendors for and developing the assessment content and technology. There is no stated plan, however, for ongoing collection of user feedback as a mechanism for continuing to improve system design over time.

Recommendations:

- * As a very early deliverable, provide ED with a detailed project plan with deadlines indicating that the development and review process can occur on time.
- * Provide ED with a detailed project plan with deadlines indicating that the reports and technology system can be developed on time.
- * Provide a plan for developing a common data format, and for implementing Partnership data collection from States and State data collection from districts and schools.
- * Devise a plan for ongoing collection of user feedback as a mechanism for continuing to improve system design over time.
- * See comments to (A)(3) for recommendations on scoring.

(A)(5) Research and Evaluation

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

(A)(5) Reviewer Comments:

As laid out in this section and in Appendices (A)(5)-A and (A)(5)(B), PARCC's "research and design" and "research and evaluation" plans are structured to ensure that the assessments developed are valid, reliable, and fair for their intended purposes and for all student subgroups, including with regard to construct, consequential, and predictive validity; external validity; reliability; fairness; precision across the full performance continuum; comparability within and across grade level, and also in regard to whether intended effects on individuals, institutions, and outcomes are being achieved. The research plan is so ambitious, in fact, that it seems unlikely that the \$8.5 million set aside for it (a figure that seems on the high end, relative to the other project demands) will suffice. As PARCC recognizes explicitly, some prioritizing

will be necessary, and steps (suggested below) to limit some of the initial Request for Information (RFI) and committee wind-up might also be warranted.

One major research component that is missing is a set of studies of the feasibility and efficiency of the balance of components in PARCC's overall design. The Partnership adheres to what might be called a "more, better" strategy -- more test administration occasions spread across the year and better (more complex and sophisticated) item types. Looming large in regard to the the feasibility of PARCC's vision are two questions: (1) whether the projected number of test occasions is optimal, given administration, scoring and other costs, and (2) whether the mix of Performance Tasks and Computer-Enhanced, Constructed-Response and Selected-Response items is optimal, given the Partnership's preference for item types that are more costly from both a development and scoring standpoint. "Return on investment" studies of the major components of the system (test occasions, test items) are necessary to enable the Partnership to assess whether the system could be implemented less expensively with little or no loss in acuity. The Partnership currently plans to assess whether its system does what it says it will do vis-a-vis, e.g., measuring proficiency across the full span of CCSS standards and across the full range of students. What is missing is a plan to determine whether a different mix of test occasions and item components could also do so, but in a less costly fashion.

Also missing are studies geared towards developing valid "on track" standards. A comprehensive process is described for determining whether high school students are college- and career-ready, but not whether elementary and middle school students are on track to reach that status.

Operationally, the plan favors an intensive, belts-and-suspenders commitment to gathering available information and expertise up front. This approach combines multiple RFIs with a variety of research committees and task forces. This approach makes good sense for some of the more difficult issues that will be addressed later in the development process, such as growth modeling, measuring teacher effectiveness and evaluation, standard setting, and vertical scaling. Because of the compressed nature of the development cycle, however, and the extent of the decision making that needs to be completed by July 1, 2011, when the item-developers come on board (e.g., item development, text complexity, accessibility/accommodations), it appears that the RFIs and expert panels slated to be use early on in the process are excessive and will delay the decisions that need to be made by the crucial July 1, 2011 take-off point. For these matters, consideration should be given to composing the expert groups so that they include the expertise that could otherwise be provided through RFIs, if time permitted.

A related concern has to do with phasing the research to be sure results are available when needed in the development process. The research proposal is careful, for example, to consider phasing in regard to the question of when and how to research the questions posed by common achievement standards. But phasing is not taken into consideration elsewhere. For example, given the contribution to project costs of the decision not to use computer-administration or computer-scoring for children in grades 3-5, and given the contribution to project risk and cost of uncertainties associated with Artificial Intelligence scoring, it would make sense to begin the planned studies of these issues immediately, rather than going through the slower RFI/expert-committee process currently envisioned.

Also related to the above two points, early resolution of the key question of who will do the research (a single vendor, multiple vendors, or the PMP) is warranted. Not having the actual research team on board from the beginning deprives the Partnership of that team's expertise (which in some cases could substitute for the RFI and committee process, and in other cases could inform that process) and also risks misalignment between the research plan committee experts develop and the one the procured experts would consider optimal.

The Research Strategy Group is made up of experts, not managers, but is given operational responsibilities, namely, managing the work of the task forces to a timeline. This management strategy adds operational risk.

The Partnership acknowledges the need to feed its research and evaluation findings back into the design and development process but has no real plan for how that goal will be met.

Recommendations:

- * Quickly resolve the question of who will do the research studies, so research vendor(s) can contribute to, and possibly reduce the complication and cost of, the up-front process of RFIs and committee considerations.
- * Conduct return-on-investment (ROI) studies on having more rather than fewer test occasions, to determine if the mix of 5 annually for ELA and 4 annually for math is optimal.
- * Conduct ROI studies on Performance Event vs. Computer-Enhanced items vs. Constructed Response Items vs. Selected Response items to determine the comparative value of each relative to cost to facilitate proper balancing of item types.
- * Conduct early studies of AI and computer testing of children in grades 3-5, to see if findings can help decrease cost and risk by allowing computer testing of early grades, and can justify a smaller amount of human scoring and influence Partnership policy on scoring generally.
- * Consider whether there is a need for both RFIs and expert committee deliberations on research topics, especially where answers are needed early in the process in order to facilitate the development work.

(A)(6) Professional Capacity and Outreach

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

(A)(6) Reviewer Comments:

Evaluation:

PARCC's 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 instructional practice is modest, leaving the vast bulk of the capacity-building responsibility to member States. Given this, the capacity-building budget appears to be too high, and some of the resources devoted to this component are not well-justified. Additionally, PARCC has not developed a mechanism for monitoring and benchmarking the all-important data-training efforts of member States or for States to share effective practices in this regard.

PARCC's strategy for informing the public and key stakeholders about the assessment system and for building support for the system is strong, although it may be under-funded. It was difficult to tease out the funding for this activity from the Project Management budget, where only \$400,000 in dedicated funding was identified. Of concern, the outreach focus seems to be more on building support for the the high school exit standards than for the implementation of the assessments in the lower grades, even though the latter implementation is likely to create the greatest need for public outreach and explanation to counteract the concerns the Partnership acknowledges in regard to public perceptions and potential misperceptions about the impact of the new assessment system on the use of classroom time.

Explanation:

* The Partnership does not plan to conduct any direct training of teachers in member States, instead leaving that responsibility with the member States. This is reasonable given the need for State flexibility and choice and given resource limitations.

* The Partnership lists four capacity-building strategies it will pursue. Two strategies are to build a series of online video training modules in order (1) to assist educators in administering the assessment, including by developing certified experts in test administration, and (2) to assist educators in using assessment results to improve instruction. Although successful completion of the latter set of courses can aspire to be only a small part of what it takes to train teachers and principals to use the rich new data effectively to diagnose and cure instructional failure, the funds devoted to these efforts (\$2 million) seem likely to be effective and well-spent relative to their amount and goals and relative to other project components.

* Another capacity-building strategy is for the Consortium to provide assistance to States in understanding how to transition from existing standards and assessment systems to the CCSS standards and the PARCC assessment system -- using gap analysis, curriculum maps, realignment of existing curricula, etc. This seems to be valuable assistance, but there evidently is no budget for it. Perhaps this service will be available to States that choose to purchase it from the Consortium or its PMP.

* The remaining capacity-building strategy, to which \$4.5 million of the \$6.5 million budget for capacity building is devoted, is to create a cadre of content experts. As also is described in comments to the budget in section (A)(8), the entire expense here is for 90 educators to travel to, and take part in, 32 meetings focused on assessment design and review. The total number of individuals involved is unclear, because it is not said whether each educator will attend only one of the 32 90-person meetings (in which case there will be 2880 total participants) or whether each person will attend all 16 meetings (meaning there are 180 total participants), or something in between. The goal is for these individuals to acquire expertise through their participation in the assessment-development work, then for them to "help model good instruction and support professional development training for other educators in their schools and districts." This quoted passage is the only description of any plan for how these educators will contribute to professional capacity building in their States. The proposal doesn't indicate how the value of these individuals will effectively or systematically scale statewide. The more of these individuals there are (up to 2880, about 90 per State), the thinner their training will be; the fewer of them there are (as few as 180, or 6 per State), the thinner their ability will be to provide professional development support at scale in their home States. Even for any educators who take part in all 16 meetings, it is unlikely, given the meetings' focus on test design and development, that sufficient professional development will occur to assure their expertise in the use of the data for formative and diagnostic purposes. If these individuals are crucial to the assessment development process, then their modest contribution to professional capacity building may be a worthy side benefit. But if the main goal is for capacity building support, the strategy is not convincing as currently explained, given its cost.

Given (1) how crucial it is for educators in the States to become proficient in using the rich data the new system will generate to improve instruction, (2) how demanding that task will be precisely because of the richness and complexity of the data being generated by the distributed tests, and (3) the inability of the Partnership to fund or conduct significant professional training in member States, there is a significant risk that a substantial value of the new system will be lost for lack of the necessary, intensive, embedded professional training. The Partnership takes no step to reduce this risk by providing a clearinghouse function for state training strategies in much the same way as it plans to do for state public-outreach strategies. Nor does the Partnership ask States to inform it of their capacity-building plans, contemplate meetings among States to facilitate improvement of those plans, or have any other mechanism for monitoring and benchmarking the plans against each other in a way that could create a race to the top among them.

Recommendations:

* Provide more of a justification for the \$4.5 million devoted to the teacher meetings. If the expenditures are mainly in service of assessment development, allocate them to that budget component, and consider whether they are justified for that purpose. If the expenditures are mainly in service of professional capacity-building, provide an explanation of how the expertise provided will be useful at scale in the membership States.

* Provide a clearinghouse, monitoring and benchmarking function for State efforts to enable educators statewide to become proficient in the use of the rich new assessment data in diagnosing and curing instructional weakness.

* Consider whether the public outreach activity is sufficiently funded and properly balanced between public outreach in regard to the standards and assessments for high school students versus those for students in the lower grades.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	5

(A)(7) Reviewer Comments:

PARCC has worthy aspirations to use technology effectively to improve the quality and accessibility of its proposed assessment system. The Partnership's strategy is to use the inducement provided by the Race to the Top/Assessment funding to prompt industry to create open-source and cloud technology that permits seamless two-way links between each member State's existing data system and the Partnership's assessment database and reporting engine.

Aside from a single notional and unelaborated diagram, however, the Partnership provides no description of the architecture it has in mind; provides no reference to an existing system with the technology it imagines; appears to underestimate substantially the difficulty of identifying or developing technology that will enable data to flow easily between its system and the various systems of different vintages in use in the Partnership's 30-plus States; and almost certainly will not be able to realize its aspirations within the \$22 million its budget has allocated to this component.

Indeed constructing the contemplated (1) back-end data store, (2) identity-management and security systems, (3) real-time population of and access to data, (4) four-second response time, (5) content management capability (e.g., the PRC item bank with functionality for educators to design their own formative/interim assessments and online libraries of other resources), and the (6) front-end reporting engine (e.g., the Interactive data Tool), (7) the reports themselves, (8) supporting analytic and data-customization tools, (9) incremental backups, (9) core metrics, and (10) field testing will likely eat up most or all of the budgeted amount.

That will leave insufficient resources, by a wide margin, to establish the (11) intended links to the various State systems and, beyond that, to provide the various assessment-related tools that are contemplated as part of this budget component, including (12) technology to support innovative assessment items (video, drag and drop, highlighting etc), (13) training for members of assessment design committees, (14) assessment administration functionality (e.g., equation editor, pull-down menus, calculator tools); (15) online tutorials and practice functionality, (16) technological accommodations for special populations, (17) scoring capability (e.g., capacity to recognize highlighted passages and graphs within a tolerable margin of error and leading-edge Artificial Intelligence scoring capacity), and (18) web (PDF?) distribution of papers for human-scoring with an intelligent feature that limits who can and cannot score a particular student's paper.

Additionally, the current specification calling for only 1 million users at a time is insufficient in a Partnership with 21.4 million test takers. This demonstrates the necessity of providing the additional back-end capacity that the Partnership has made a portion of its Level 2, Priority 2 back-up request.

The Partnership correctly looks forward to the day when industry will provide the K-12 domain with modular, inter-operable, open-source tools that permit the many one-off state systems to communicate with each other. It suggests that RTTT-A resources can be used as leverage to hasten the coming of that day. In that way it hopes to assure that the technology it develops for this project can be reused by other States or organizations in a cost effective manner. However, unless additional funds are available – including, for example, if a major portion of the technology funds from competing RTTT-A projects were merged into a common fund to create a single inter-operable system for multiple consortia – it is unlikely that the inducement the Partnership hopes for will arise. A \$22 million contract is not unusually large by current IT standards, considering the private as well as the public sector. And the lack of leverage is magnified by the Partnership's worthy desire to require the creation of open-source technology. Doing so may diminish the value of the contract to many vendors because they have no hope of reselling the code and tools they develop and will have to obtain a profit based entirely on development fees, requiring them to raise the

price they charge the Partnership for programming and other development work. Additionally, because existing open-source technology (e.g., for content management) has not been widely used in the K-12 domain, there are not many consulting-support organizations that are experienced in customizing that technology to K-12 requirements.

The Partnership appears to recognize the risk here. It qualifies its commitment to "leverag[e] advances in technology needed to bring our sophisticated assessment design system to scale by 2014-15" by acknowledging that it "will need to closely monitor the feasibility of its technology plans during the design and development stages of the work to ensure that the Partnership and our partners develop state-of-the-art technology solutions . . ." The problem with this approach – and the Partnership's hope that it can rely on other technology that may materialize (e.g., the system the Council of Chief State School Officers has proposed) – is timing. The timeline gives the Partnership about a year after putting out an RFP for and contracting with vendors to produce the system in time for field testing. Even with sufficient resources, that is an aggressive timeline for the large amount of functionality described above. There is no time available to wait for other technology to appear.

Feasibility, therefore, depends on the Partnership's ability to shift additional resources to this component.

PARCC recognizes the equipment and infrastructure requirements imposed on States, districts and schools by an all-computerized assessment system, and the data requirements the system imposes on States (including a capacity to link students, teachers and administrators). The Partnership has developed a credible strategy for assisting States in finding ways to understand and meet these challenges.

Recommendation:

- * Direct additional resources to this budget component.
- * If there are multiple RTTT-A grantees, consider combining the resources available to all of them for technology and requiring them to develop a single, inter-operable system. To the extent possible, consider leveraging systems newly created or being created by particular States or LEAs.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	14

(A)(8) Reviewer Comments:

Evaluation

PARCC's project management plan has some strong features but does not provide full confidence that it will result in implementation of the proposed assessment system on time, within budget, and in a manner that is financially sustainable over time. The Partnership's Project Management Partner (PMP) does not have significant experience with a management, integration and implementation/operational task of this scale. The Partnership proposes to place chief management responsibility, including the Director role, in 8 FTEs who have not yet been hired, whose qualifications are unknown; and who have never before worked together on any project, much less one of this scope. Without elaboration, the proposal mentions in passing the possibility of donated assistance from an organization (Battelle) with more of the required experience. But a project of this scale requires a qualified PMP under a contractual arrangement that gives it full responsibility for management success or failure – which is unlikely to be the case for an occasional, unpaid advisor. The proposed PMP has an impressive – indeed, heroic – track record of conducting carefully crafted sequences of meetings and interim drafting and reporting processes that eventually generate a policy consensus around educational standards where none had existed before or seemed possible. In one case, the PMP used that process to obtain consensus around end-of-course standards for Algebra II and then Algebra I, after which the PMP contracted with and managed a vendor that constructed

two exams aligned to the standards. Those exams, however, were relatively traditional in design and scoring methodology and are administered to only about 100,000 students a year. The PARCC proposal that the PMP has helped craft is built in this same image, devoting very substantial amounts of funds to travel and meetings, on top of even more substantial amounts of fund devoted to project management itself and separate additional funds for the fiscal agent budget (which is devoted almost exclusively to contract management), while spending comparatively little time discussing the process of actually building and integrating PARCC's many proposed, complex and novel products. Those products are on a much larger scale than the two end-of-course tests for 200,000 students annually that the PMP has built in the past and include, e.g., at least 18 summative tests (considering the through-course testing modules, it is more like 81 different tests) and some formative assessments, for 8 grade levels, in 2 subjects, for 21 million students annually, using items (e.g., performance tasks) and scoring techniques (e.g., Artificial Intelligence) that have never before been applied in this manner in the context of K-12 summative tests) and requiring a large-scale technology platform of at sort the PMP has never built, with data collection and storage, test-administration, scoring, reporting, business intelligence, content management and social networking functionality, and also requiring training exercises and public outreach efforts on a national scale. All of those components have to be designed, developed, seamlessly integrated, and delivered on a compressed, grant-disciplined timeline. The proposal does not make sufficiently clear how the Partnership intends to make the large number of small incremental policy decisions that will have to be made each day. Managing that kind of process through many periodic meetings of experts convened from around nation is not a proven approach. Too much of the budget appears to be devoted to administration, leaving other project components (especially building and implementing the technology components) seriously under-funded.

Explanation:

* The qualifications of the PMP, as evidenced by its mission, size, experience (including past success in implementing similar projects), and key personnel assigned to this project are impressive in regard to some (e.g., policy making, standards-test alignment, and research) aspects of the project but are not sufficient for the crucial management, integration, and operational aspects of the project:

(1) The PMP has extensive experience and an unmatched track record of assisting groups of States, higher education representatives, other dispersed actors in reaching a general or national policy consensus, and building sufficient consensus and will within each State to enable it to take effective steps to implement the policy behind the consensus. The PMP has advised, supported, and provided rigorous institutes for States that have undertaken to implement those agreements by adopting educational standards and aligning summative and formative assessments and curriculum and public outreach to the standards. In the case of Algebra II and I standards, it helped develop specifications for summative assessments to measure attainment of the standards and managed a contract with a vendor with experience building similar assessments in the creation of those tests. Many of its employees have been K-12 educators themselves, primarily in math but in at least one case in ELA as well. It has substantial research experience and capacity. (Until recently, the PMP's experience is weighted heavily towards high school standards and assessments.)

(2) The PMP has an \$8 million annual budget (up from \$4 million in 2003). As currently designed, this Project Management effort would triple its budget. If the entire grant amount is considered, this project would increase its budget 20-fold. Its limited experience building assessments is confined to end-of-course exams in math for the upper grades. It has no experience managing a project of this scope, measured by either the project budget; the number of components, employees, consultants, and meetings; the number of contracts and small and large vendors to be managed; the task of building and integrating large-scale technology components; the need to integrate often novel content, scoring methodology, technology, research, training and public outreach; and an aggressive timeline tied to a funding mechanism with no "give" in the case of delay or cost overruns. The PMP's self-described mission, strengths and qualifications do not encompass these project management, systems integration and operational disciplines and activities. None of its current employees is a trained or certified project manager.

(3) The PMP contemplates a management team of 23 people, 12 of whom are current employees. Among those 12, however, 8 would spend 50% or less of their time on the project. Only 4 of the 15 people

who will work full time on project management are currently known and employed by the PMP; the other 11 remain to be hired, including the overall Project Director, Finance and Federal Reporting Director, Assessment Director, Education Community Outreach Director and Postsecondary Engagement Director. This is a large project team – perhaps larger than it needs to be – especially taken together with the four FTEs included in the separate Fiscal Agent budget component. A large amount of time is slated to be spent convening and conducting meetings of experts costing close to \$4 million (exclusive of travel) including a full time Meeting Planner. Pulling together a new project team composed mainly of full-time staff who have to be hired for this purpose, have never worked together, and have no common project management approach, protocols and tools imposes significant risks.

(4) Risk is increased because the proposal contemplates numerous separate contracts and RFPs, the largest of which (for assessments and technology) evidently will not result in the identification and contracting of vendors until at least nine months into the project (June 30, 2011). Focusing only on item-content development, the proposal says: "The Partnership is likely to release multiple RFPs for various components of the system to take advantage of and encourage innovation across multiple sectors, rather than issuing a single RFP." PARCC thus has decided to do the contracting piecemeal, rather than at least inviting, if not requiring, consortia of vendors – with a single project-management, integration, and fiscal structure – to bid on collections of functionality and to assemble a group of subcontractors to provide desired diversity and inter-vendor competition. This approach expands the management cost (multiple contracts, billing protocols, invoices, governance structures, and management teams) and risk. Even just the Partnership's technology plan, although composed of six or more large components and distinct tools, provides for no integrator or integration funding. Rather than a single set of integrator-to-integrator interactions between the Partnership and the vendor group, the Partnership contemplates many interactions taking place between and among different vendors, different management staff, and different design/work teams and teams of convened experts, expanding risk.

* The project work plan and timeline, including, the major milestones, deadlines, and entities responsible for execution, and the approach to identifying, managing, and mitigating risks associated with the project are not fully adequate to the project:

(1) The Project Management plan is clear, well-written, and in most cases well-documented. It commendably facilitates review and evaluation of what is intended. The management scheme in Figure 1 of the management section along with the supporting narrative, and the timeline in Figure 2, reveal that the Partnership has a good overall picture of the major management task it faces.

(2) The proposed management and decision-making process, involving the convening and achievement of agreement among diverse and representative groups of experts and policy makers, seems fairly well-suited to the initial policy decisions that are slated to be made in the Fall 2010 to Winter 2011 period and for test design. One concern, discussed in the (A)(1) comments above, is that the quarterly meetings of the Governing Board will be insufficient to provide the necessary resolution of disagreements and final decisions in a timely fashion. Another concern is that the planned use of both RFIs and expert panels will slow down these initial decisions.

(3) The proposed management and decision-making process is not, however, as well-suited to the development, integration, and operational processes that will commence as of about June 2011. For that part of the project, the Partnership needs to delegate, but has not delegated, significant day-to-day decision-making authority to some executive or set of executives with the project, with the Governing Board focusing on major decisions and resolving occasional disputes. Whether this executive is the Project Director contemplated in the PMP structure, or needs to be something more like the coordinating committee and its chair that is contemplated in the recommendations to (A)(1) above, is a matter for the Partnership to decide. But a sufficient process is not currently set out in the application.

(4) In regard to the timeline:

(A) Many key policy decisions about the scoring blueprint, common policies and procedures (e.g., for defining EL students, special-needs accommodations, test security and item-release policies) are set to occur between October 2010 and February 2011. Reaching these kinds of agreements are the forte of the

PMP. Under these circumstances, the aggressive timeline for the getting major foundation decisions made is commendable and will reduce risk for the RFP process and for vendors.

(B) The timeline has a sensible overall logic: Foundational policy decisions are made between October 2010 and February 2011; RFPs generate vendors by June 30, 2011. Initial development of key components occurs between then and June 2012, along with pilot testing. Full field testing occurs thereafter, followed by further development steps in reaction to the testing. Development of scoring and reporting strategies follows the initial development period as well, given the need to design the tests before deciding how to score and report on them. In regard to all of these steps, there is a much clearer and more detailed plan and timeline for the assessment work streams than for the reporting and technology ones – consistent with the areas of relative strength and weakness in the proposal as a whole. Given the ambition of both the new test content (which amounts to content for 81 different tests each year, spread over 9 grades, 2 subjects, and 9 distributed testing points (5 for ELA, 4 for math) – and in many cases imagines quite novel material) and the new technology system, there is a serious question whether one year of initial development (July 2011 to June 2012) will suffice. It might be possible to begin scoring and reporting development during the later stages of the assessment and technology design processes in order to get a sufficiently early start on those decisions, both of which have significant implications for the technology system.

(C) More generally, extracting, transforming and loading data from the member States is one of the most difficult tasks presented by this project, and little consideration is given to that function in either timeline or the budget.

(D) It is unclear whether sufficient time has been built in for integrating the various technology pieces (the technology timeline, refers to six separate development projects for different tools in the system, but has no reference to their integration), nor for integrating assessment formats for administration, scoring and data into either the reports or the technology system. Initial item development, and development of the technology platform, are both slated to be completed about the same time (6/30/12), and field testing begins two months later. It is unclear whether that is enough time for the products of those two work streams to be integrated.

(E) Although crucial, the process and timeline for developing performance level descriptors, a common growth model and common achievement standards are not fully set out in the timeline and narrative, creating a risk that the Partnership's admirable goals in this regard – a key strength of the proposal – will not come to fruition on time.

(5) Adding to project risk, one or more significant matters of baseline policy remain to be determined by PARCC. Evidently, the Partnership has not reached agreement on scoring policy – i.e., the appropriate mix of human scoring (whether by teachers or a third-party vendor) and Artificial Intelligence scoring. The possibility of having to conform the system to different approaches taken by different States adds risk to the development process and cost estimates.

(6) The Partnership devotes a single paragraph to risk management, which consists largely of a statement that the Governing board "with support from [the PMP]" will monitor the PMP. This paragraph makes passing reference to the possibility of donated assistance from a contract research and development organization (Battelle). The Partnership is to be commended for recognizing the importance of project management support from an organization like Battelle. But the informal, unpaid, and unspecified arrangement that is suggested will almost certainly not provide the necessary amount of risk management.

* The Partnership's Level 1 and Level 2 budget modules are clear, but in the case of the Level 1 budget, there appears to be some imbalance between components that in some cases seem to be over-funded in relation to project objectives and in other cases seem to be inadequately funded to support the development of an assessment system that meets the requirements of the absolute priority:

(1) The overall administrative component of the budget is too high relative to the other needs. This includes fiscal and product management (\$18 million), support for contributing states (\$7.3 million), and significant parts of the overall budget for travel and meetings (~\$12 million). Although some of these parts of the budget are more properly allocated to development and training than to administration, the 25% of

the budget devoted to them seems excessive. One modest, but significant contributing factor is the decision to hold all meetings in a relatively expensive location, Washington, D.C. Another item of concern is the money allocated to pay for support from Governing States (\$7.3 million). This is for FTEs who will perform a project coordination role on a state-by-state bases. Having a dedicated and responsible state coordinator is a valid expense, above the administrative costs normally required for projects of this sort, which is necessitated by the interstate nature of the endeavor. Cost containment strategies ought to be considered, however.

(2) The Level-1 Technology budget is too low by a significant factor. It reaches a level that is closer to reasonable in order to assure basic system integrity and functionality only when combined with the Level-2, Priority 2 budget (\$10 million). In particular, \$2.5 million for the Reporting, Data Warehousing, and Interactive Data Tool seems low by perhaps a factor of two. The \$3 million for the Content Management system and Partnership Resource Center also seems low. The cloud computing infrastructure is a new technology in the K-12 domain, providing significant risk in regard to the \$5 million infrastructure estimate (as is indicated in the Level 2, Priority 2 request for an additional \$5 million).

(3) The \$80 million for assessment design and development may be somewhat high. One contributing factor is the high cost of Performance Tasks (PT), which, for example, are over six times more expensive than Technology-Enhanced constructed-response items (TE). In fact, PTs are far more comparatively expensive than that, given that PT items can only be used once, while TE items can be used more than once, and given that a smaller percentage of developed PT's survive the review process than do TEs. It appears that something like 14 PTs per year, per subject, per grade are contemplated, which may be more than is necessary. More generally, given the increasing sophistication of TE's, it is unclear whether the value added by PTs is in all cases worth the expense added – a matter that, at the least, is worthy of careful study but is not included in the Partnership's research plan. Likewise, the extent to which the frequency of through-testing adds to operation expense is not set out, but deserves similar return-on-investment analysis.

(4) A small amount is set aside for paper-and-pencil assessment for a "small number" of students. A concern arises whether the absence of computing equipment and infrastructure in some member States will create more pressure for additional paper-and-pencil capacity.

(5) A major cost in item development, review and human scoring is the per-hour cost of the experts used to conduct these functions. It would be useful to know what that cost is, including in order to allow comparison to the cost of using in-service teachers to do the work. Although the Partnership recognizes the value of participation in item design as a basis for creating expertise among its projected training specialists, it does not establish a policy that would take advantage of the same dynamic by involving teachers from all member States in assessment design, review and scoring.

(6) There is some uncertainty in regard to the online delivery expenses, in part because they seem to have been estimated separately from the technology work stream. There is a sentence in the proposal stating that: "The costs for an online assessment test delivery system including scoring of selected-response and technology-enabled items, hosting, maintenance, and other associated system costs has been estimated at \$3 per student per year based on the nature of the assessment system for a total of \$400,000." The amount stated only makes sense if it assumes that 133,333 students will be taking the assessment, a number of test takers that is unexplained. (The Partnership has 21.4 million tested students.)

(7) There is substantial uncertainty in the actual cost of Artificial Intelligence scoring, adding risk, as again is suggested by the additional \$5 million requested for this purpose in the Level 2 budget.

(8) The amount devoted to Quality Control seems low.

(9) The \$400,000 allocated for public outreach seems low given the magnitude of the task of avoiding the possible misinterpretations of the impact of the new system on the use of classtime.

(10) Approximately \$4.5 million of the budget for Professional Capacity Building – i.e., over 75% of the \$6.5 million budget for that component – is for travel and the cost of 32 meetings for 90 educators each

from the partnership States. The purpose of these meetings is not entirely clear from the budget narrative or the discussion in (A)(6). Evidently, the desire is to have a number of educators from each State become expert in the new system by working with the design and review teams, who then would be available as resources for the State when the system rolls out. It is unclear whether these individuals are necessary to the item-development/review process. If so, their expense should be budgeted to that domain. But if these individuals are mainly being used for capacity-building, then it is unclear how effective they will be, given the small numbers involved relative to the entire body of teachers in a given State and given their uncertain distribution within States and lack of any obligation to make themselves available as trainers or developer when they return to their home State. Part of the problem is that it is not clear whether each participating educator will attend only one of the 16 meetings (in which case there would be 2880 such people whose expertise would be based on participation in a single meeting) or whether, on the other hand, the same 180 people would take part in all 16 meetings (meaning 6 highly trained experts to cover each of the member States). Neither of these models has obvious power as a capacity-building device.

* In its Level 2 Budget Modules, the Partnership gives a lower priority to additional technology support than to a formative assessment tool for K-2 students. As is discussed above, additional funding for the technology system is necessary to assure its basic viability. However, it appears that cuts in other parts of the Level 1 budget could make provide space within the \$150 million cap to accommodate those additional technology needs, without having to take advantage of Level 2 spending.

* The K-2 formative assessment tools contemplated in the Priority 1 portion of the Level 2 Budget Module is appropriately addressed as a Budget 2 priority. Devoting \$10 million to pushing forward the know-how and capacity associated with K-2 assessment would be a welcome development given the current under-development of K-2 assessment, its importance, and also given the sensitivity to the needs and also to the substantial challenges that is exhibited in the Partnership's Level 2 proposal. Based on experience, the current under-development of formative K-2 assessment tools is a substantial barrier to improved student learning at this crucial early age, and this proposed Level 2 item would go a long way towards filling this gap. A further specification of details about what is intended would facilitate evaluation.

* As of now, neither the Partnership nor its individual members has a clear plan to enable States to bear the estimated costs for the ongoing administration, maintenance, and enhancement of operational assessments going forward. The per-pupil cost of operating the system (\$32.68) seems high. There is no information about how this figure (which does not include equipment and infrastructural costs) compares to the per-pupil funding the member States currently devote to summative assessments. Nor is there a well-developed process for determining the equipment and infrastructural needs of each State and providing assistance in meeting them.

* The Partnership does not examine the cost to States of contributing the data needed for the new Data Warehouse and its Interactive Data Tool, including developing data formats that align with the Partnership's tools and capturing data (e.g., linking students to teachers and classrooms) that most States do not currently collect centrally. The Partnership is to be commended, however, for acknowledging the importance of student and teacher identifiers.

Recommendations:

*Reconfigure the management structure (including for policy decision-making and project and fiscal management) by:

- Encouraging the existing PMP, in lieu of hiring new unconnected individuals, to subcontract with a seasoned, financially committed project management and integration specialist with substantial experience integrating and implementing components of this diversity and magnitude at this scale.

- Reconsider the decision-making and management structure to provide more of a capacity for day-to-day decision-making, especially on questions of implementation that places less reliance on panels of experts convened intermittently whose proposals then have to be approved at quarterly meetings of the Governing Board. The coordinating committee suggested above and described in the comments to (A)(1) is one possibility.

- To retain, but streamline, the state-by-state project coordinator function, offer States the option of empowering a member of the Project Management office to serve as their state-by-state project coordinator, or detailing an employee to the PMP, or teaming up with other States to hire or designate a coordinator. In this way, there could be less than one FTE per Governing state assigned to this crucial but expensive line item.

* Consider combining the major RFPs (e.g., for assessment design, scoring, technology, reports) into a single vehicle, and inviting major industry integrators to organize consortia of vendors with the full range of technical, K-12 domain, fiscal and project management expertise. If diversity among, e.g., item-writing vendors is desired, this can be accomplished by requiring a diversity of sub-contractors, and creating contractual mechanisms through which their content is competitively compared. By retaining freedom at the vendor-identification stage to select between, or mix and match among, different bidder teams and individual bidders, the Partnership could assemble the right overall group of vendor partners, while moderating the expense of running different RFPs and having to deal with and coordinate among many different vendors, multiple billing and invoice practices and schedules, etc. The most important gain here would be creating a single, strong project-management and integration operation on the vendor side to interact with the Partnership's own project management office.

* After making changes of this sort, redraft the budget, aiming to decrease the overall amount devoted to administrative costs and increase the amount available for technology. Consideration of cuts in the Professional Capacity and Research and Evaluation budgets is also warranted.

* If travel and meeting time remain high, consider establishing the Project Management Office in a more centrally located hub location to which travel, and where meeting space, can be obtained less expensively than the District of Columbia.

* Given the Performance Tasks' high financial and class-time cost, inability to be reused, scoring and other risks, psychometric challenges, and potential overlap in functionality with the computer-enhanced constructed response items, consider decreasing the number or cost of performance tasks, or conducting an early pilot comparison of them to the technology-enhanced items (including administration, scoring, psychometric analysis and reporting) to determine whether the return on investment is sufficient.

* Increase the Level 1 Technology budget by at least the amount contemplated in the Level 2, second-priority budget. Consider teaming up with other consortia or use of new systems being developed by States or LEAs to decrease technology costs.

* Survey computing and infrastructure capacity across member States to provide information about the costs of implementing the new system apart from operational costs.

* Survey member State's data collection formats and warehousing policies to determine the best design for the Partnership's data formats and to determine the cost to States of providing the necessary data.

* Incorporate the process of designing Common Achievement Standards into the timeline and budget.

* Consider extending the one-year initial development period for item development and for developing the new technology platform.

* Consider whether the budget and timeline devote sufficient attention to the large-scale psychometric task of combining four (math) and five (ELA) different tests each year, using three or four different types of test items into a single, reliable, comparable score.

* Provide more detail on how the Partnership proposes to develop K-2 formative assessments in its Level 2 proposal.

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
--	-----------	-------

Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	17
---	-----------	-----------

Competitive Reviewer Comments:

Universities that, according to the States, represent 90% of the direct matriculation public students in the member States, and some private universities, have sent letters to the Secretary committing to participate with the Partnership in the design and development of final high school summative assessments in mathematics and English language arts in order to ensure that the assessments measure college readiness. Among the Partnership States, only two had fewer than a majority of their direct matriculation public students represented, and both had over 20%. The Partnership has developed a rigorous process in which IHEs will be intensively involved in this process.

Additionally, the MOUs signed by each State commit the State, on a schedule not specified, to obtain "commitments from public IHEs or IHE systems to use the assessment in all partnership states' postsecondary institutions, along with any other placement requirement established by the IHE or IHE system, as an indicator of students' readiness for placement in non-remedial, credit-bearing college-level coursework."

The IHE letters to the Secretary that have been obtained thus far suggest some weakness in the commitment of IHEs or IHE systems 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 and any other placement requirement established by the IHE or IHE system. This weakness is visible in the letters' use of language of the following sort:

* IHE "pledge[s] to work collaboratively with our K-12 state counterparts and our higher education colleagues across the partnership states to develop high school assessments that can serve as an indicator of readiness for non-remedial, credit-bearing, college-level coursework in mathematics and English."

* IHE is "committed to working with the Partnership to ensure that students who score college ready on their high school assessments can enter credit-bearing coursework without remediation at this institution."

* IHE is "prepared to participate with the Partnership in the following next steps: . . . Use of the assessment in all Partnership states' postsecondary institutions as an indicator of students' readiness for placement in non-remedial, credit-bearing, college-level coursework."

* IHE is "prepared to participate in the design, development and standard setting process of the Partnership as follows: . . . Review how the new measure(s) may be used in conjunction with existing placement measures established by the University of Arkansas System campuses, as an indicator of students' readiness for placement in non-remedial, credit-bearing, college-level coursework."

* IHE agrees to pursue the "the goal of using the new measure(s) as part of our course placement system once the Partnership has set the college readiness standards for the assessment(s)."

Nonetheless, a sufficient level of commitment to meet the requirement was determined to be present.

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:

PARCC'S application demonstrates its intention to develop and implement an assessment system that comprehensively measures student knowledge and skills against a rigorous set of college- and career-ready standards. Its system aspires to: measure proficiency across the full range of those standards, including standards that traditionally have been difficult to subject to measurement; to elicit complex student demonstrations of knowledge and skills across the full range of CCSS standards and student abilities; and to provide an accurate measure of student growth over a full academic year. Its system consists of multiple through-course assessment components and sophisticated item types in ELA and mathematics that produce student achievement and growth data the Partnership States will use to: determine whether individual students are college- and career-ready or on track to being college- and career-ready; effectively and accessibly assess all students, including English learners and students with disabilities; and produce student achievement and growth data that can be used to inform determinations of school effectiveness, individual principal and teacher effectiveness; profession development needs and instructional improvement.

Two concerns are noted, but are sufficiently addressed by the proposed assessment system and development plan to address the concerns and warrant an affirmative answer to this absolute priority. First, are significant concerns about management capacity and the distribution of resources in the budget, as is developed in comments to (A)(8) above. Without proper management, it is impossible to conclude that the Partnership "will develop and implement" a qualifying assessment system. Given PARCC's recognition of the issues here, however, and given its ability to solve the problem by enhancing its PMP's project management capacity through a more formalized version of the arrangement it proposes to create with Battelle, this concern is adequately addressed.

A second concern is whether the proposal will develop and implement an assessment system that provides an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students. Absent computer-adaptive techniques, and particularly in the context of the fixed forms the Partnership appears to prefer, meeting this requirement poses a challenge. This concern also is allayed, however, by the availability of techniques that can enable to the Partnership to fulfill its clearly stated commitment to develop and implement a system with this capacity. Those techniques include (1) the use of off-grade test forms (along with the waivers that the proposal contemplates that its member States may request from ED), (2) broad use of off-grade items in the contemplated end-of-year exam, and (3) adjustment of the through-course assessments earlier in the year to include some number of selected response or constructed response items focused on students performing well below-grade.

In prior comments, specific recommendations are made in regard to both of these concerns.

Grand Total	220	151
-------------	-----	-----

Budgets

Level 1 Budget
Name: Level 1 Budget(s) See comments in (A)(8).
Level 2 Budgets
Name: Formative Assessment Tools for K-2 See comments in (A)(8).
Name: Technology Enhancements

See comments in (A)(8).

Name:



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b)
(6)

(A)(1) Consortium Governance

	Available	Score
(A)(1) Consortium Governance	20	16

(A)(1) Reviewer Comments:

- The detail of rights and responsibilities signed by participating states is a significant strength of the partnership’s submission. All of the required signatures have been secured.
- The major components of the assessment system outlined in the vision are consistent with the expectations required in the grant and outlined in the Theory of Action (summative assessment, formative assessment, technology platform, scoring and moderation, reporting platform, and tools for professional development and capacity building in formative assessment).
- Membership in the partnership exceeds the minimum requirements for the grant and is strengthened by the involvement of states with very large populations. That being said, Governing states have limited representation from the western half of the country, which has a relatively minor negative effect on the submission.
- PARCC requires the chief state school officers to serve on the Partnership’s Governing Board, which constitutes a significant strength in the commitment of the states involved as well as in their ability to move issues along.
- Participating states must commit to piloting and field testing, which is an advantage to the PARCC submission efforts in acquiring the psychometric data necessary for implementation of the assessment program.
- Florida is identified as the fiscal agent, which is a strength because one state is identified for the fiscal accountability of grant funding. Unfortunately the submission provides no information as to the reason for this selection and, specifically, information pertaining to their track record in managing grants of this nature.
- The partnership has already concluded the competitive process to contract the project management partner and, therefore, will be ready to launch if this submission is successful.
- The submission recognizes potential political risks which can influence commitment to a project with as many political overtones as is evident in the Race To The Top. In the event that there is a change in governor or chief state school officer, the state is required to affirm the state’s ongoing commitment to the agreements previously in existence.
- Membership opportunity outlined in the submission is a strength because non-partnership states may apply to join at any time, and because member states may withdraw at anytime with only needing to provide an explanation for the withdrawal. The latter aspect improves the partnership’s accountability toward its members.
- Since the Governing Board makes decisions by consensus, the implication is that unanimity is valued; yet, a simple majority (referenced inaccurately as “supermajority”) is deemed adequate when consensus is not achieved. The submission is weakened by not emphasizing consensus sufficiently when total consensus is lacking. Many issues in assessment are philosophical in nature and benefit from agreement of more than just a simple majority when dealing with pushback from the education system.

- In attempting to provide effective governance, the submission intends to establish "optimal size" working groups. This is a laudable objective but could become a threat to the collaborative nature of the partnership if inconsistent application of optimal size becomes evident.
- The submission identifies a dozen policies that require resolution and, therefore, become starting points for the partnership's activities. Most of these policies have a resolution date prior to the end of August 2011, and requiring so many decisions early in the process is considered a strength because working committees will have critical information to guide their thinking.
- This governance section of the submission is evaluated in the high range and at the low level.
- Recommendation: That consideration be given toward identifying a higher standard of agreement than 50% plus 1.

(A)(2) Theory of Action

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

(A)(2) Reviewer Comments:

- The Theory of Action is built on a continuous process of assessments based on through course summative assessments supplemented by optional formative assessment tools. This plan features a considerable amount of testing which has a potential for negative reaction by teachers and parents. That said, the through course component provides more opportunity to gather summative data on student achievement. Therefore, the potential for acquiring meaningful data on student achievement from a summative aspect is a considerable strength even though the time that might be necessary to accommodate both types of assessment somewhat weakens the submission.
- The computer based testing model leads to a standardized test approach that requires more items to zero-in on a student's level of achievement and, therefore, more time to conduct the assessments. The submission is weakened by not providing sufficient clarity as to the reasons why computer based assessment is the chosen methodology over other types of approaches.
- The Education Department's selection criteria regarding testing, reporting, evaluating teachers and principals, building capacity, and moderation are all addressed within the Theory of Action.
- The commitment to ensure 90% of the school year is completed prior to commencing the final summative assessment is a strength, especially in light of the additional commitment to have the scoring completed for use in the students' year-end report cards.
- The submission makes it clear that student success is contingent on several factors including the quality of service provided by the school (teachers and principals). This strengthens the submission because there is an acknowledgement that improvement in student achievement is more than designing programs.
- Teacher professional development is an important requirement in making a submission, and the partnership sees this element being addressed by the ongoing use of through-course assessments. The partnership's commitment for rapid turnaround of results is a positive feature because professional development needs should be readily evident for each teacher and school during the same school year that has student learning being assessed.
- The assessment plan includes students from grades 3 through 11, which exceeds minimum grant requirements. Attribution for student achievement will be more easily identified for senior high programs, which benefits the teacher evaluation aspect in the area of accountability. Further, students and their parents will benefit from an ongoing assessment of progress relative to College and Career Readiness without experiencing a gap between grades 8 and 11. The submission is accurate in identifying this added

feature as providing increased motivation for students and, therefore, is considered a positive factor in the submission.

-- This component of the submission is evaluated at the low level of the high range.

--Recommendation: The partnership indicates a commitment in their application to connect with the Smarter Balanced Application Consortium (SBAC) which proposes to utilize computer adaptive assessment, and it is recommended that a discussion take place regarding the benefits and drawbacks associated with this submission's computer based and SBAC's computer adaptive assessments. Specifically, to what degree do these assessment technologies measure growth in student achievement over the course of one year , and provide an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students.

(A)(3) Assessment System Design

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

(A)(3) Reviewer Comments:

-- The assessment proposal is consistent with the grant's selection criterion of being through-course summative, end of year summative, as well as providing teachers with optional interim assessments for formative purposes. This feature constitutes a positive. That said, the through-course assessments appear to be aligned with periodic points (e.g. quarterly) through the school year with no indication whether the curriculum taught early in the grade will be re-assessed at later points. Ambiguity on this issue constitutes a negative.

-- The assessment system outlined in the narrative is consistent with the theory of action, and the summary table contained at the end of this section is an excellent and succinct portrayal of the narrative.

-- Feasibility is a concern in this design because so much time during the year is set aside for testing. Backlash from parents and teachers is a distinct possibility.

-- The partnership intends to assess outcomes that have been traditionally more difficult to assess by incorporating performance components for speaking, listening and writing in the ELA summative assessment. These components will receive more classroom emphasis because they are being assessed in a summative fashion. That said, speaking and listening assessments will not be incorporated into the summative component but only be available for the report card. These also are important skills for College- and Career- Readiness (CCR), and may not experience the desired emphasis in classroom discussion because they are not included in the summative assessment, which is the basis for staff evaluations. Recognizing that "what gets measured gets done", this aspect then weakens the submission.

-- It is intended that prompts will be released so that teachers and students can better understand the performance expectations of Common Core State Standards (CCSS). This is a necessary feature and strengthens the design because prompts are easily recalled from year to year, and attempting to keep this information from the public opens the door to some teachers and students not having equal access to such important information.

-- The commitment to a short timeframe for the turnaround of results is a significant plus in the submission made even stronger by the lateness in the school year for beginning the final summative assessment window. This combination of quick turnaround and late testing provides opportunity for teachers to have the assessment results in time for year-end report cards.

-- Testing in grades 9 through 11 exceeds requirements and is a significant strength in the submission because of the continuity of useable data already evident in the annual assessment program for grades 3 through 8.

- This section contains frequent use of examples to demonstrate what is meant by concepts being introduced thereby making interpretation of the submission's intentions relatively easy. It also indicates that the partnership has already thought through many issues making this a strong aspect of the submission.
- The assessment design is strengthened by indicating that weighting of the components will be applied to the various summative assessments, thereby indicating that this is an issue which eventually needs to be resolved.
- The submission recognizes that the high school standards will lead to course structures and, eventually, to assessment design for mathematics. This positive aspect is strengthened by committing to meet with higher education before final decisions are made, so that they have input into how courses are structured and aligned with their entry-level courses.
- A plan to develop K-2 optional performance tasks is included as a level 2 budget priority. In essence, this proposal substantially embellishes on the minimum grant requirements.
- A cost-saving measure is indicated in the proposal as a result of an arrangement with Massachusetts to release items aligned with CCSS standards to the PRC item bank.
- The test scoring portion of the plan is evaluated as a weakness because an inconsistent approach across the partnership is employed. Use of teacher scorers is optional which introduces unfairness in the marking process. Further, not including teachers reduces the professional development potential for building assessment capacity in teachers.
- Table 1 provides a rationale that speaks of an inconsistency. Double scoring is projected for the high school assessments because these are high stakes in some states; yet, double scoring is not an option for the other grades. It appears that the potential for teacher and principal evaluations to be considered as high stakes is overlooked or, at the very least, is not regarded as being high stakes.
- Tables 2 and 3 commit the partnership to an extensive distribution of data reports for various stakeholders. It is also indicated that parents will be given access to reports on schools in the area or schools of similar make-up. This type of information demonstrates a higher level of commitment to accountability than merely providing evidence to parents on the performance of their child's school alone. This feature is a positive enhanced by schools also having access to this type of information, which provides opportunity to identify and contact exemplar schools for improvement ideas.
- The assessment plan is strengthened by allowing teachers to determine timing of the formative assessments so that these occur soon after instruction has occurred.
- The proposal speaks of the need for a high percentage of double reads in scoring where high stakes accountability is an issue. It appears as though the grant requirement to have student achievement data inform determinations of principal and teacher effectiveness for purposes of evaluation is either forgotten to be high stakes or is not considered high stakes. In fact, the introduction of these evaluations being based partly on summative assessments means that high stakes are now universal.
- The submission is strengthened by recognizing and stating the important role accountability plays in motivating improvement. Both support and pressure strategies are necessary elements for improvement but, frequently, the pressure aspect is downplayed.
- Accommodations are a significant aspect of an assessment program and the submission is enhanced by indicating its commitment to consulting with associations who are particularly focused on fair assessment practices for all students.
- The assessment design is clearly articulated in this section and in sufficient detail indicating that a high degree of preliminary planning has already occurred.
- The section on system design is evaluated at the high end of the medium range.
- Recommendation: That the partnership consider summative through-course assessment to occur less frequently during the school year to allow more opportunity for formative testing.

-- Recommendation: That the partnership consult with SBAC on their experience and philosophy regarding grade 3-5 students' readiness to undertake computerized assessments.

-- Recommendation: The proposal recognizes interpretation issues when there is evidence of 'low scores with high growth' or 'low growth with high scores'. It is recommended that the partnership consider using a gap reduction strategy for scoring performance proposed by the Center for Policy Research in Education (CPRE) rather than the traditional tactic of measuring improvement.

-- Recommendation: That the partnership be more innovative by considering computer adaptive testing in the summative assessment program in addition to its planned use in the formative assessment program. Computer adaptive testing is more innovative because it adjusts for the level of difficulty after each student response. In this way, it is an individualized assessment and has two significant advantages. First, less testing time is required because fewer items are utilized to pinpoint student achievement. Second, there is greater accuracy at the tails (higher or lower achievement) because the test adapts with more questions for students whose level of achievement is at the tails. Computer based testing introduces a greater margin for assessment error for measuring growth in students achieving at the high and low ends of the achievement scale. This recommendation is limited to ELA because the partnership believes this issue is resolved in the Math assessment design; however, it may be a worthy consideration in Math as well.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	27

(A)(4) Reviewer Comments:

-- The partnership intends to infuse rather than retro-fit universal design into the assessment program, which meets the requirements stipulated for the grant application. This application is particularly strong in its commitment for ensuring that universal design is evident.

-- The process for developing the assessments is clearly enunciated as are the qualifications of the personnel working on the various committees. This process is enhanced by a plan to unpack the CCSS and determine the optimal assessment strategy for each standard. Unpacking these standards will be accompanied by a very important activity of using samples of student work as evidence of what students can actually do.

-- Production of a test blueprint will be informed and refined by discussions about weightings to be applied in the assessment design, as well as ensuring that there is a natural progression of tasks so that assessments are comparable from one year to the next.

-- The PARCC's commitment to using authentic texts as much as possible is costly, but they understand how students will benefit from writing assessments that are more like the real world.

-- The process incorporates a good quality control strategy by using a Bias and Sensitivity Committee to review all assessment items.

-- The grant requires completion of an assessment strategy by 2014/15 which requires aggressive staging of activities. To this end, the partnership enhances its submission by setting a goal to pilot some test items in 2011/12 and begin formal field testing in 2012/13. This schedule will provide ongoing feedback necessary to achieve the project's target. Further, the proposal provides a recognized sampling strategy as their way of ensuring test reliability.

-- Conducting field tests at the same time of the school year that the assessments will actually be taken facilitates comparability from year to year, which is a significantly positive feature in a well-designed assessment program.

- The plan to have end of year tests scored entirely by technology is likely necessary for the quick turnaround feature of this proposal, but its plan is weakened by introducing a risk to accurate scoring because double scoring or read behinds in the through-course assessments is not being utilized.
- The submission is strengthened by articulating a full range of marker moderation strategies for enhanced quality control.
- There are many aspects of this submission which indicate that development processes are well-in-hand and, therefore, this section is evaluated in the high range. The concerns raised in this evaluation place the overall section at the low level of the high range.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	26

(A)(5) Reviewer Comments:

- Grant requirements specify the need for quality control mechanisms to achieve fairness, reliability and validity in the assessments. The partnership has substantially enhanced its submission by designing a series of committees comprised of qualified personnel to oversee psychometric issues which, if not handled properly, can be an impediment to constructing quality tests.
- The submission is enhanced by proposing to undertake a wide range of research regarding assessment issues as well as on how the use of assessment information can be improved. Analysis of these topics indicates this to be a strong research agenda, and is particularly significant because the topics include vertical scaling, which has ramifications for measuring improvement in student achievement.
- The research design is strengthened by the partnership's focus on the relationships regarding student achievement and post-secondary assessments as well as international benchmarking assessments.
- This section acknowledges the legitimate concern that their assessment design will increase substantially the amount of external testing teachers may have experienced previously. Therefore, their intent to study the social impact of the testing program further strengthens their submission.
- The partnership recognizes that their testing program is jeopardized by budget increases for this assessment program and, therefore, their research design is enhanced by including a focus on how economies of scale might be leveraged.
- The submission responds to the selection criterion that consequential studies be conducted. Not only are systemic issues emerging from the expanded use of summative assessments being examined, but the submission is enhanced by a plan to study the degree to which formative tools are actually accessed by teachers.
- The submission's research and evaluation component addresses all of the requirements of the grant and 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:

- PARCC's professional development plan is consistent with grant expectations. A positive feature of the partnership's plan is about how they endeavor to achieve budget efficiency by not supplying the training

directly but, rather, leveraging resources already being spent on professional development in each state. Since PARCC refers to this project as having the "potential to profoundly change what goes on between educators and students in classrooms", this approach to have states re-direct expenditures to this project and, thereby, increase ownership is a significant strength.

-- Re-directing professional development funding at the state level does not mean a 'hands-off' approach but, rather, a 'handing-on' notion whereby the partnership commits to developing experts as well as training tools/courses. This approach is enhanced by the partnership's commitment to maintain electronic records of course completions, affidavits and assurances. Therefore, the potential for accountability within the workforce is being strengthened.

-- The submission recognizes the significance of scoring student work accurately, and commits to providing training for developing teacher capacity in this critical fairness activity. Participation in this training is voluntary and, even though teachers can obtain certification, the approach outlined in the submission lacks accountability because the marks scored by teachers on student tests will not be used. The submission is weakened by not outlining some strategy for following-up on teachers uninvolved in training, yet still scoring student work for marks on the student report cards. That being said, it is unclear that this professional development fully satisfies the criterion that the professional development is intended to inform instructional practices to then improve instruction.

-- The partnership anticipates a negative reaction to the increased number of testing days in a school year and strengthens its proposal by outlining a communications plan for ameliorating these concerns. Particularly impressive is the commitment to focus on parents and how the new assessment process will benefit them personally (e.g. What should I do if my child is not on track).

-- The proposal is strengthened by its commitment to build coalitions with key stakeholder organizations. Support from the post-secondary level is critical and the submission is enhanced by its commitment to identify a position entitled, Director of Post-Secondary Engagement.

-- A concern in the outreach aspect pertains to the preponderance of one-way communication techniques without providing opportunity for feedback.

-- The submission in this category is evaluated in the high range and at the low level.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	5

(A)(7) Reviewer Comments:

-- This section lacks some of the specifics evident in other sections of this submission. Frequently the proposal responds to grant requirements for this section with phrases such as "will provide", will need to closely monitor", will carefully study", etc., and does not provide lists of examples to clarify their intentions for the reader.

-- Most of the technologies identified in this section demonstrate little innovation as much is already in use. The control mechanisms for accommodations during testing is innovative.

-- There is an intention to "push the field" to develop an improved AI scoring mechanism which can be identified as a positive. That said, AI is a significant assumption in this submission and a suitable back-up plan is not articulated if this technology does not meet expectations.

-- Continuing to utilize paper-pencil assessments in grades 3-5 is a weakness in the submission.

-- The submission is strengthened by a plan to develop a school site certification process so that states can determine which sites need upgrading.

- This section of the submission is evaluated in the medium range.
- Recommendation: That the partnership consult with other testing organizations regarding the use of computers when testing students in grade 3 through 5.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	22

(A)(8) Reviewer Comments:

- The submission is strengthened by the early identification of a project management firm, which has already engaged in numerous projects in the education sector and across many states. Also, there has already been a connection with most of the states in the partnership because of these previous projects, including work in assessment.
- The firm is located in Washington and, as a positive, states that it is an "independent, bipartisan, nonprofit education reform organization"; however, it is relatively small in size and will need to supplement its organization with new staff. That said, the base of the organization has already established itself as a reputable firm in the eyes of 'Education Week' as one of the most influential groups in the nation.
- Qualifications of the firm's senior staff are included as per grant requirements, and the submission is enhanced because the firm's president has significant experience in working with education standards and assessment.
- The project management will benefit from the organization's commitment to ongoing communications with its members through a members-only web-site.
- The project is strengthened significantly by being aligned, at no cost to the partnership, with a large independent contract research and development organization that will assist in project management.
- A weakness in the plan emerges by the lack of consistency in the scoring process whereby states will have flexibility to choose their own mechanism for scoring student responses. This notion introduces unfairness and reduces potential for economies of scale. Hence there is a large differential in potential assessment costs indicated depending on the model chosen by each state.
- The partnership is of the opinion that test administration costs will be lower than current costs, which would be a positive feature given that the new assessment scope would be considerably larger than what has been in place.
- A project timeline listing major events is contained provides a quick summary regarding priorities, and information about how states will fund and maintain the assessment system over time is also included.
- This section is evaluated in the high range but at the low 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:

- The partnership meets selection criterion in its interactions with higher education by securing their commitment to use assessment results as an indicator of student readiness for general education programs. Further, the selection criterion for eliminating participation in remedial courses is also met.
- The commitments represent 90% of direct matriculation students in public IHE's in all partnership states, which significantly exceeds the threshold identified in the selection criterion.
- The proposal outlines a detailed strategy on the collaborative work between the project and higher education.
- Commitment letters from the IHE's are signed as per the selection criterion.

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 submission is consistent with, and in some cases exceeds, all of the main requirements identified for grant consideration. It covers the full range of standards, including standards against which student achievement has traditionally been difficult to measure. An accurate measure of student growth over a full academic year or course is provided. It consists of assessment components in Mathematics and in English language arts that include, for each subject, one or more summative assessment components that are administered at least once during the academic year in grades 3 through 8 and at least once in high school, and intends to produce student achievement data and student growth data that can be used to determine whether individual students are college- and career-ready or on track to being college- and career-ready.
- The submission has a plan to assess all students, including English learners and students with disabilities.
- The proposal seeks to provide student achievement data and student growth data that can be used to inform determinations of school effectiveness for purposes of accountability as well as determinations of individual principal and teacher effectiveness for purposes of evaluation.
- The proposal meets formative requirements by incorporating determinations of principal and teacher professional development for improving teaching and learning as well as the program.
- Extensive detail of issues that need to be considered in the process leading to implementation is provided, and the submission is significantly strengthened by the vast amount of examples used in explaining intentions more fully.
- As previously stated, the proposal addresses the key issues of accuracy in measuring student achievement across the full performance continuum, including for high- and low-achieving students, and plans to provide an accurate measure of student growth over a full academic year or course. While the 'Absolute Priority' elements are technically met, feasibility in satisfactorily addressing these issues requires greater clarification. They may intend to build these features into the end-of-year test, utilize an approach with multiple forms of tests or, even, adopt off-grade testing, but the methodology is not articulated. Without specifying and incorporating these kinds of approaches it is difficult to determine how these key issues will actually be achieved without utilizing a test that is extremely time-consuming.

Grand Total	220	166
--------------------	------------	------------

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

- The strength of the submission is enhanced by external commitments of support which are provided at no cost -- i.e. New York city's contribution of research, Battelle's potential contribution of in-kind services, and Achieve's donation of time by their president.
- The requested amount fits within the grant requirement.
- The budget's accuracy is enhanced by pegging travel costs to GSA rates for DC.
- The budget rationale is strengthened substantially by the significant amount of detail contained which allows for easy monitoring and adjusting as the project progresses through its various stages.

Level 2 Budgets

Name: Formative Assessment Tools for K-2

- The dollar amount of this level 2 budget request is within the grant requirements.
- This funding request is for creating assessment tools beyond the grant requirements contained in the level 1 grant submission -- i.e. K-2 formative tasks. Having knowledge about student achievement success in these early grades is important to schools in an area where there are relatively few assessment materials. This proposal recognizes that learning problems resulting during the initial grades in school might be ameliorated early in a student's career by having formative assessment results available before minor problems become major.

Name: Technology Enhancements

- The dollar amount of this level 2 budget request is within the grant requirements.
- This funding request is for activities which also fit within the parameters of the level 1 budget. Certainly there are going to be technology needs to accommodate having students take computerized assessments. Evergreening technology is an ongoing need and might be addressed within yearly school budgets supplied by the states. Further, this particular budget request could be subsumed within the partnership proposal if insufficient funding is available for a level 2 budget request.
- The aspect of pursuing improvements in AI scoring is a high priority for this proposal given the potential for accuracy and fiduciary efficiency in scoring student responses.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b)
(6)

(A)(1) Consortium Governance

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

(A)(1) Reviewer Comments:

The Partnership has clearly defined the primary goal of the organization—to create a common assessment system for Grades 3 to 11 that can annually report on students' progress toward college and career readiness. To this end, the Partnership members have established a governance framework and set of guidelines to achieve the goal. Each member state, which includes 11 governing states and 15 participating states, have completed a Memorandum of Understanding (MOU) to formalize their commitment to the development and implementation of the proposed assessment system.

Decision-making authority, including significant expenditures and disbursements of grant funds, rests on a Governing Board. The fact that Governing Board is comprised of the chief state school officer from each governing state is another indication of the level of commitment of member states. The governance structure provides opportunities for differentiated participation and engagement levels. Clear roles and responsibilities have been defined for those states in governing and participating membership, as well as advisory and working groups who will provide key support to the design and development process, those who will review and validate the process, and those who will use the information to produce informed decisions about students' readiness for college and careers. Consensus is the goal for all Consortium decisions. There are also clear procedures for joining and leaving the partnership and for changing roles.

Each MOU also includes assurances regarding participation in the Consortium procurement process. A Procurement RFP is expected to be issued in fall 2010. Florida is the designated fiscal state agent.

There are some areas that Partnership states identified as possible barriers that could negatively impact their participation in the project. While all member states stated in their MOU their commitment to redirecting current state assessment budget allocations to the implementation of the proposed assessment system, some are concerned about the potential for state budgets getting cut by their legislatures (AL, GA, KY). Others singled out the need to adopt the Common Core State Standards (AL, KY, SC, TN), imbalances in technology infrastructure (GA, SC), and resolution of conflicts with state laws/regulations (CA, DC, DE, GA, IL, LA, NH, NJ, OH, PA, SC, TN).

(A)(2) Theory of Action

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

(A)(2) Reviewer Comments:

The Partnership's Theory of Action provides realistic and collaborative goals and objectives consistent with the intent of the Race to the Top program in a defined time frame and within the Partnership' capacity for implementation. It also actively brings together the collective capacity of local teams, higher education and other expertise to produce results that can inform decisions about readiness for college and careers and

increase the rates at which students graduate from high school prepared for success in college and the workplace.

Reliance on a combination of interim or benchmark assessments distributed over the academic year after roughly 25 percent, 50 percent and 75 percent of instructional time and an end-of-year assessment focusing on higher order knowledge skills to produce weighted "annual combined scores" in each subject provides a base from which progress can be measured. The assessments will function as an integrated element in a larger system made up of a wide range of item types, including multi-session extended performance tasks, achievement scores, annual student growth measures, evidence-based benchmarks to monitor whether students are on track or have achieved the targets, and ongoing collaborative, professional work linked to curriculum, administrative support, and parental involvement. Analyses of school effectiveness across multiple dimensions (e.g., in terms of both "achievement status" and "student growth") and comparison of students' performance across Partnership states and against a widely shared benchmark of post-secondary readiness will help regulate school-based incentives and interventions.

Finally the Assessment Design Committee includes content leads, assessment experts and teachers from each member state, and nationally recognized expertise to ensure the quality and fairness of the assessments.

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	50

(A)(3) Reviewer Comments:

The Partnership's proposed assessment system is designed to measure the extent to which students have acquired the knowledge and skills needed for post-secondary education or the world of work. To this end, college and career readiness benchmarks will serve as an anchor to the system to identify skill deficiencies before the end of the 11th grade—and thus providing timely opportunities for teachers to intervene before students enter post-secondary classrooms or the workforce. The assessment system includes four through-course assessments in English/Language Arts (including a speaking and listening assessment), three in mathematics, and one end-of-year assessment. The through-course assessments will be distributed throughout the year to ensure that measurement of key skills and critical knowledge occurs near in time to when they are taught. Performance tasks will include an appropriate range of cognitive skills to provide measures of achievement across the ability spectrum, and will apply principles of universal design to make the assessments accessible to students with disabilities and English language learners. Results from these administrations will be combined in each subject to calculate weighted "annual combined scores" to monitor students' progress on college and career readiness.

The proposed assessment system makes use of advances in technology to administer and score students' ability to apply what they know to real-life tasks. All of the assessments in English/Language Arts will be online, grades 6-11, as will the end-of-year assessments in Grades 3 through 11. Recognizing that keyboarding and other factors could negatively impact test performance, the Partnership is opting for paper-and-pencil tests in grades 3-5 until it can be determined that students in these grades have sufficient computer skills to respond online within the test time limits. The assessments comprise a mix of constructed-response items, performance tasks, and computer-enhanced and scored items. The speaking and listening component is a "live" classroom performance tasks administered by the teacher.

Examination of the proposed content to be measured indicates that the assessments are targeting the full range of skills and genres that are considered essential for success in college. These include those that are difficult to measure, such as ability to synthesize information from a range of text formats or demonstrate conceptual understanding and ability to do mathematical modeling.

The assessment design envisions the generation of various types of information about student performance, including achievement scores, growth measures, college and career readiness indicators, and item analyses. Periodic data feedback to support instruction, professional development and accountability decisions will be available at the individual student level, aggregated at various levels (classroom, school, district, state), and disaggregated by NCLB and other student groupings. Additional support includes built-in diagnostic and various activities to support teachers' formative needs, such as embedded assessments and ready to use formative assessment tools. Finally, the design plans to develop online, digital resources to model curriculum frameworks that teachers can use to plan instruction and gain a deep understanding of the Common Core State Standards, and released items and tasks.

There are some areas in the design that may impact the psychometric soundness of the assessment system. The use of technological innovations such as video and audio clips may impact some psychometric standards related to unidimensionality (maintaining the measurement focus of an item on a single construct) and local independence (ensuring that items do not give clues to other items). Likewise, use of artificial intelligence routines to score constructed-response performance task items in both English/Language Arts and mathematics could be problematic. Research has shown that AI current state of the art is not particularly effective scoring math items, nor at analyzing stylistic elements of writing, such as hyperbole.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	34

(A)(4) Reviewer Comments:

The Partnership's plan for developing the proposed assessment system includes at least two important elements to help ensure wide-scale administration of the system in a timely, cost effective and consistent manner. First, it uses an evidence-centered design approach to developing the system of assessments while using universal design methods within each component of the system. This combination allows for the consideration of a full construct using multiple measures, and ensures comparability of assessments and preservation of construct validity when modifying presentation and response factors for students with disabilities and English language learners. Secondly, it uses task design templates to quickly and cost-effectively develop similar but varied tasks that validly measure the same construct in multiple ways.

The Partnership has explicitly described the key processes involved in the development of the assessments, such as content analysis, blueprinting, performance level descriptors, appropriate distribution of items or tasks across the full breadth of the standards for a particular grade level or subject area, item development and alignment review. The use of cognitive labs through the item selection and review process to evaluate student response strategies to different items across all types of students will help ensure that the assessments are measuring the content and performance expectations called in the Common Core State Standards across the ability spectrum.

The assessment system development plan considers accessibility from the beginning. Not only does it include the active participation of expert reviews from the National Center on Educational Outcomes (NCEO), the Center for Applied Special technology (CAST) and others, but also envisions common procedures for identifying eligible students, selecting and accepting accommodations, and implementing accommodations. Furthermore, the plan anticipates the use of a NAEP-like stratified matrix sampling approach to investigate the impact of accommodations and special populations on construct validity. Piloting of some assessment components will take place in 2011-12, field test in 2012-13 and 2013-14 and implementation in 2014-15.

The plan also has built a quality assurance system to ensure the reliability of scoring. In this sense, the plan will include double scoring and read behinds. Artificial intelligence scoring will be validated by human scoring. Rangefinding will take place to identify key characteristics of a representative sample of responses

across all score points and lead to the production of a set of “anchor” responses and annotations from the most common types of responses at each score point, scoring guides, and training materials. Engagement of teachers, assessment experts, higher education, state content teams, and other expertise will further contribute to the score quality assurance system.

While computer-delivery of the assessments can be much more efficient than traditional paper-and-pencil test, hardware limitations may have a minor impact on the types of items that can be administered by computer. Items involving detailed art work and graphs or extensive reading passages, for example, may be hard to present.

(A)(5) Research and Evaluation

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

(A)(5) Reviewer Comments:

The Partnership’s research agenda includes two phases designed to ensure that the assessments are valid, reliable and fair for the intended purposes and for all students. The first phase will involve issue-focused task forces to provide guidance and expertise on college readiness benchmarks, setting performance standards, and ensuring comparability of results when testing students with special needs (accessibility and accommodations), age and grade appropriateness of text complexity, and the pros and cons of vertical scaling and best way to implement it.

The Research Strategy Group advisory group will provide advice and guidance during the design and development of the assessment system. The Bill and Melinda Gates Foundation will host “resource roundtables” on pooling resources, technology platforms, professional development validation and research methodology, and comparability issues.

The second phase will focus on the impact of the assessment system on teaching learning and student readiness. Several topics will guide this phase, such as the psychometric and other measurement properties of the assessment components, including the validity of achievement levels cut scores, criterion-related studies, and the consequential impact of the assessments. The Research Strategy Group, for example, is recommending a set of consequential validity studies to evaluate whether the system has a positive impact on classroom instruction for all student populations. Critical questions to be explored include the extent to which the common assessment items and tasks and the formative tools provide teachers valid and useful guidance for planning units and lessons, and the degree to which teachers and administrators use the data provided by the assessment system to make adjustments to the curriculum to improve instruction. Another set of predictive validity studies will focus on the impact the system has had on student achievement in general and on rates of student readiness for college and careers in particular. One basic question relates to whether high school students who meet the college and career readiness standards succeed in entry-level, credit-bearing college courses or in apprenticeship and training programs.

Finally, and as an indication of the thoroughness of the research agenda, the Partnership plans to conduct a study on the fiscal impact of the assessments on member states in order to ascertain costs, hidden and unintended, that may impact the implementation of the project after the four-year grant period and how leveraging economies of scale might help drive down costs.

(A)(6) Professional Capacity and Outreach

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

(A)(6) Reviewer Comments:

The Partnership recognizes that effective implementation of the proposed assessment system rests in large part on sound professional capacity-building. In line with this idea, the Partnership's professional development plan includes a training and certification program to support teachers and administrators in implementing the system in ways that ensure reliable and valid use of results. A central feature of the plan is the building of a leadership cadre of content experts with an in-depth understanding of the essential elements of the assessment system and their connection to the Common Core State Standards to provide a framework for local training and support efforts. The plan also envisions the development of effective training tools to help educators implement the system, score student responses, interpret and use assessment results, and understand and address the curricular and instructional implications of the Common Core State Standards and the assessments. Appropriate deliverables include the development of informational materials such as FAQs, brochures, presentations, sample assessment items and tasks with accompanying annotations, and examples of student work from the field tests. Partnership members will have access to these resources on the Partnership's web site along with detailed recommendation for how to use the materials.

The Partnership's capacity-building plan includes collaborative strategies among its members to take advantage of existing funding streams, such as professional development dollars provided under Title II and ESEA and state resources devoted to teacher training related to standards and assessments. Considering the funding challenges that states and districts face to secure sources of funding for their educational initiatives, especially in the context of current and anticipated budget pressures, the formulation of such strategies is an effective tool to sustain the delivery of professional development supports and services.

Given the amount of testing that the proposed assessment system holds for students and teachers, Partnership members are also proactively working on a Public Outreach and Stakeholder Engagement Plan to build public support and higher education engagement. This is a phased, multi-tiered effort to ensure that parents, K-12 and post-secondary educators, policymakers and members of the public understand what they system intends to accomplish and its rationale. It is also an effective strategy to address expected parental and community fears of "more test days" and "over testing."

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	9

(A)(7) Reviewer Comments:

The Partnership's use of technology to facilitate the implementation of the assessment system is extensive. Assessment administration, scoring and reporting will be done by computer for all students in grades 6 through high school starting in 2014-15. Online tutorial sessions and practice tests will be available to familiarize students and educators with the online assessment delivery platform, interfaces and tools of the assessment. The Partnership also plans to use Web conferencing to reduce the need for large, costly in-person meetings. All of these elements create the conditions necessary for the implementation of the assessment system in an efficient and secure manner.

The Partnership's use of open source solutions to build a technology platform for assessment delivery for field testing and for full operational administration in 2014-15 is a cost-efficient strategy. By using programs whose source code is made available for use or modification through public collaboration and as users see fit, member states are likely to save on those costs associated with purchasing, maintaining and upgrading computer programs. Data from summative assessments will be housed in the Interactive Data Tool, which will be open source, compliant with platform aligned data standards for easy accessibility and use and FERPA regulations. Access to this tool will be based on user roles and passwords.

The Partnership will also benefit from the use of cloud-based, distributed technology architecture as the framework for data mining, data access and exchange of data. This should help mitigate the difficulty of local systems to integrate, extend or change as quickly as the requirements for assessment delivery, scoring and reporting.

The Partnership claims that the only local technology requirements will be an Internet connection and a standard Web browser. While this may be a true statement, it omits the fact that online administration is feasible if a school has enough computers for a large number of students and the students are at least partially computer-literate. The absence of one or both these requirements at any level in the Partnership's membership continuum from state to district to a school building can be a significant limitation on the Partnership's ability to successfully implement the proposed system.

(A)(8) Project Management

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

(A)(8) Reviewer Comments:

The Partnership has identified Achieve, a Washington DC firm, to serve as the project manager. Examination of the firm's credentials shows that it has the expertise and the resources to carefully plan, manage and coordinate the Partnership's proposed assessment system not only in terms of the overall project but also in terms of coordinating the work among the states to ensure quality control and coherence across the work of the various experts and contractors that the Partnership plans to employ.

Achieve's project work plan includes the use of timelines and milestones to guide the work and ensure that deadlines are met and important decisions necessary to keep the work on track are made. Multiple levels of information appropriate for different audiences with the Partnership are part of the work plan as well. This approach represents an effective communication plan to disseminate information about the plan and expectations to enough people, including external stakeholders.

A positive element in Achieve's plan is the inclusion of risk management as a way to monitor progress, evaluate mid-course corrections and introduce adjustments to the design, the development process or other significant areas of the project.

While the proposed budget supports the costs associated with staffing, assessment design and development, research and evaluation, and others, not all costs for the Technology item have been estimated. Given the extensive use of computer technology to administer, score, report, communicate, and train, it is recommended that the plan should include conducting a needs assessment among the member states and associated cost to define the current status of technology staff, equipment and infrastructure, and to establish a built-in mechanism for periodic review, evaluation, and revision for projecting future needs and affordability.

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:

Through formal letters of intent, participating IHEs have expressed strong support for the Partnership and pledge to work collaboratively with their K-12 state counterparts and higher education colleagues across

the Partnership states to develop high school assessments that can serve as an indicator of readiness for non-remedial, credit-bearing, college-level courses in mathematics and English.

The total direct matriculation students in participating IHEs as a percent of state total is 90 percent.

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 Partnership for Assessment of Readiness for College and Careers (PARCC) provides a strong vision and theory of action of what an assessment system ought to be in the pursuit of preparing students for college and career readiness. The system, as designed, is equitable and accessible, and may reflect the fact that it can tap the collective experience and expertise of member states that have played leading roles in advancing standards-based reform and building sophisticated assessment systems.</p>		
Grand Total	220	206

Budgets

Level 1 Budget
<p>Name: Level 1 Budget(s)</p> <p>May be adequate to defray the development phase of the assessment system. May not be sufficient to address the costs, particularly those associated with the maintenance and potential upgrade of technology infrastructure, during the implementation years.</p>



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b)
(6)

(A)(1) Consortium Governance

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

(A)(1) Reviewer Comments:

Consistency of the consortium's vision, goals, role, and key deliverables with the theory of action. The consortium's assessment system consists of the following: (1) A summative assessment system for grades 3-11 comprised of four components distributed throughout the year in English language, three components distributed throughout the year in mathematics and one end-of-year component in each subject area; (2) A technology platform for delivering and scoring all assessment components; (3) A scoring platform and moderation system; (4) A reporting system that provides data from the assessments to key stakeholders shortly after each through-course component as well as annually and that allows users to access data on their own through an interactive online platform; (5) Tools for professional development and capacity-building to deliver the assessments and use results to inform instructional decisions. Overall, these are very closely aligned with the Theory of Action proposed by the consortium. Each of these deliverables constitute aspects of an assessment system envisioned in the Theory of Action for measuring whether students are on track towards college and career readiness throughout the year as well as an end of year summative assessment.

Missing. The vision presented by the consortium does not include specific strategies for improvement of learning. In particular, how the proposed assessment system will enable teachers to improve instruction and learning are not described or discussed. Consistent with this, there is no emphasis on teacher professional and capacity development. This is a major weakness in the proposed plan in terms of its effectiveness for improving learning. Improvement in learning can only be expected if targeted changes are made to instructional practices and student learning processes. The consortium's vision does not include strategies for changing teaching and learning in classrooms across the participating states.

(Theory of action components are: (1) using college and career readiness as an anchor, (2) measuring rigorous content and ability to apply that content, (3) measuring learning and providing information throughout the school year using through course assessments, (4) through course components, and (5) end of year assessment.).

Structure and Operations. Twenty-six states are participating in this consortium. The consortium has put together a detailed and well thought-out governance structure. The 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 application process for new members, and how to opt out are clearly identified.

The Governing and Advisory States are clearly defined and identified. The governing structure consists of a Fiscal Agent, a Governing Board, a College-Ready Advisory Board, an Assessment Design Team, a Technical Advisory Committee and a Research Strategy group. In addition, a Project Management Partner has been identified through a competitive process. These committees cover the key functions of the Governance Structure. The members, roles and functions of each of these committees are clearly

identified. The decision-making processes within and across committees are described clearly (e.g., recommendations by consensus, if consensus cannot be reached alternatives to be proposed to the Governing Board). The funds will be managed by one of the Governing States who will serve as the Fiscal Agent.

The consortium's purposes and uses for the assessment system results are clearly identified in the Memorandum of Understanding (MOU). By signing the MOUs, each participating state makes a commitment:

(1) To measure and document students' college and career readiness by the end of high school and progress toward this target. Students meeting the college and career readiness standards will be eligible for placement into entry-level credit-bearing, rather than remedial, courses in public 2- and 4-year post secondary institutions in all participating states.

(2) To provide assessments and results that:

- Are comparable across states at the student level;
- Meet internationally rigorous benchmarks;
- Allow valid measures of student longitudinal growth; and
- Serve as a signal for good instructional practices.

(3) To support multiple levels and forms of accountability including:

- Decisions about promotion and graduation for individual students;
- Teacher and leader evaluations;
- School accountability determinations;
- Determinations of principal and teacher professional development and support needs; and
- Teaching, learning, and program improvement.

(4) To assess all students, including English learners and students with disabilities.

These commitments reflect common goals and vision for the participating states in the consortium that are necessary for the successful implementation of the assessment system.

(A)(2) Theory of Action

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

(A)(2) Reviewer Comments:

The primary goal for the consortium is identified as "to dramatically increase the rates at which students graduate from high school prepared for success in college and the workplace" and four components of the assessment system are identified as (1) using college and career readiness as an anchor, (2) measuring rigorous content and ability to apply that content, (3) measuring learning and providing information throughout the school year using through course assessments, and (4) leveraging technology for innovation, cost efficiency and speed.

There is a poor alignment between the primary goal of the consortium that is improving learning and these four components. An essential element of education improvement is engagement and professional development of teachers. This element is missing from the theory of action. No details are provided about what the expectation is regarding how improvement in teaching and learning will be established without

consideration of how teachers and learners will be engaged in using assessment information. It is unrealistic to expect real improvement in education simply by setting high standards and sending messages to educators about expected learning outcomes. This approach ignores the importance of teacher agency and preparedness in using assessment information for achieving improvement in education.

Component 3 addresses distributing through-course assessments so that assessment of learning can take place closer in time to when skills and concepts are taught. However, it does not describe any provisions for variations across districts and states. This raises questions about opportunity to learn for all students across the states in the consortium as well as how the assessment program will provide good measurement accuracy for examinees at different ability/competency levels.

(A)(3) Assessment System Design

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

(A)(3) Reviewer Comments:

The consortium presents a very detailed description of most aspects of the assessment system. This demonstrates the consortium’s progress in working together and capitalizing on existing expertise and resources to identify many elements of the assessment system.

Innovativeness. The assessment system is innovative in its use of combination of through-course and end of year assessments to support both formative and summative purposes. The ELA/Literacy assessments will consist of two focused literacy assessments, an extended research/writing assessment, a speaking and listening assessment, and an end-of-year assessment. The mathematics assessments will have three through-course and an end of year component and together will contribute to an annual combined score. The through course assessments will give opportunities to teachers to time them closely to instruction as well as obtain scores/feedback soon after the administration of these assessments. The multiple test administrations will give multiple measurements of student learning as well as provide opportunities for teachers to evaluate teaching/learning over the course of the year. These assessments will still be “external”, in other words will not be created by teachers and have the potential of not being well aligned with teaching and learning in the classroom. Therefore, their success depends on the extent to which they will be aligned with instruction in the classroom and the extent to which they promote valued learning outcomes. These assessments will also make it possible to assess the full range of standards (over multiple test forms and administrations over the year).

Samples of innovative items are presented in the application. In these items computer use facilitates asking questions in ways that are different than they are used in paper and pencil tests. However, these uses do not constitute innovative ways of assessing ELA or Math. Rather, they are versions of similar items/tasks commonly used in paper and pencil tests using computers. For example, one ELA item asks students to read two passages and think about similarities and differences between the opinions expressed. Students are then asked to click and drag one sentence from each paragraph that shows how the two opinions are similar (into a Similar box) or different (into a Different box). Another item asks the students: “Highlight the word in the passage that refers to a famous temple. Use the highlighting tool to click on the word and highlight it.”

Similar examples could be found in the Math assessment. One item presents the students with a trapezoid and asks students: “Using the drawing tool found in the lower left hand corner of your screen, connect the two points that would connect the line of symmetry for this figure”.

These examples demonstrate that the items that are presented as “innovative” by the consortium are in fact very ordinary items that assess relatively routine knowledge but enable students to do on computers what they normally do on paper and pencil tests.

Measurement accuracy across ability levels. The consortium is proposing to measure all ability levels with similar measurement accuracy by including tasks that will be targeted to assess competency at the low end of the scale as well as at the high end of the scale. The assessments will be computerized (grades 6-11) but not computer adaptive. Administration of single forms of assessments cannot be expected to provide similar levels of measurement accuracy for all ability levels even when tasks with different difficulty levels are included in the assessment. Flexibility of administering different test forms with different difficulty levels is not included in the plan. Good measurement accuracy for all ability levels is essential in order to measure growth for all ability levels. Tests that are too difficult or too easy do not provide scores with sufficient measurement accuracy to be able to measure growth accurately. In such a scenario, examinees do not get an opportunity to demonstrate improvement in their knowledge, skills and competencies. This can seriously impact the extent to which assessment data can be used to inform accountability, evaluate teacher, school and principal effectiveness.

Formative assessment. Formative assessment tools will be in the form of ready to use assessments. The consortium will support development of resources for teachers and leverage current and planned efforts in member states to develop tools. The resources will include curriculum frameworks, released items and tasks (including performance tasks) and a text complexity diagnostic tool. These will be useful for allowing teachers to familiarize students with test items. However, this strategy focuses on testing driving instruction without giving teachers relevant professional development to improve learning. This can result in classroom teaching that is very heavily focused on testing and re-testing to get ready for the through-course and summative assessments.

ELs and SWDs. The consortium has developed clear plans that will help them develop assessments that are valid for ELs and SWDs. They will establish two committees, the Committee on Accessibility and Accommodations and the Research Strategy Group. These committees will provide guidance and technical assistance in making assessments accessible to ELs and SWDs. The consortium will explore innovative technologies as well as guidelines from previous research and operational experience to develop assessments for ELs and SWDs.

Extensive detail about data and its uses. The consortium provides detailed information about the data that can be used to support instruction, professional development and accountability decisions. Four kinds of data will be provided about student performance. These are data on student achievement levels, student growth measures, college and career readiness, and item analysis of released items. In addition, the consortium presents detailed information about how data can be used to inform teaching and learning, identify professional development needs and additional program supports, for evaluating teacher and principal effectiveness and inform student placement into entry level, credit bearing college courses in English and mathematics.

In addition to providing detailed information about the assessment system in text, the application presents extensive tables of information about assessment components, type of data that will be produced and how it can be used, number and types of items and distribution of item types, administration mode (online, paper based), scoring method and estimated turn-around time, score reports and their targeted audiences.

Recommendation. A major revision to the assessment design is required in order for assessments to provide meaningful and accurate data about student learning and growth for all students across the performance spectrum. This revision should include the flexibility of administering different test forms with different difficulty levels that will be targeted to student performance levels. This flexibility can be in the form of alternate forms, multi-stage branching tests or computerized adaptive testing.

(A)(4) Assessment System Development

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

(A)(4) Reviewer Comments:

The applicant's plan for developing the proposed assessment system is 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 has thorough piloting, field testing, piloting and validity investigations. Comprehensiveness of the consortium's plan is evident in their descriptions of all aspects of the assessment development. For example, with regards to validating test items, the validity investigations address the following questions:

- Does the item or task measure what it intends to measure?
- Does the item or task respect the diversity of the assessment population?
- Does the item or task material have a clear format for text?
- Does the item or task material have clear visuals (when essential to the item)?
- Does the item or task material have concise and readable text?

The approaches for developing assessment items. Several committees will be established to ensure quality and fairness of the assessments. In particular, an Assessment Development Committee that consist of content leads, assessment experts and teachers from each member state will be established to oversee the assessment development process. The consortium provides a clearly explained rationale for using evidence centered assessment design to develop summative assessments and universal design within each component of the system. The consortium sees benefits to the explicit articulation of claims, evidence and data from assessments in evidence centered assessment design. Universal design approach will be used to consider accessibility of all students from the beginning of the assessment design.

(A)(5) Research and Evaluation

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

(A)(5) Reviewer Comments:

The consortium has an extensive list of research studies that are targeted to verifying and documenting properties of the assessment system, as well as identifying areas for revising and improving the system. It has two phases: (1) research and design and (2) future studies to evaluate the theory of action, including to measure the impact of the assessment system on teaching, learning and student readiness for college.

Research and Design. The proposed research and design plan includes psychometric analyses of the assessments, research on accessibility of the assessment for all students, timing of test administrations and opportunity to learn, testing time, growth models, inter-rater reliability, efficacy of the computerized assessments for younger students (in grades 3rd through 5th), test data structure (dimensionality, etc.), construct-irrelevant variance, standardization and flexibility, content related validity evidence, reliability and generalizability across assessment components, and validity of inferences made based on test scores by different stake-holders including classroom teachers.

Some excellent elements of the research plan include validation of the readiness scores by using external criteria. These criteria include statistical relationship between performance on the consortium's assessments and other examinations such as NAEP, ACT, SAT, ASVAB, Workkeys, among many others, as well as performance on college courses. In addition, the content of the high school exams will be compared to the content measured by college admissions tests, placement tests and entry-level college courses. This research is essential for gathering validity evidence for scores that can be compared to

college and career readiness empirically. It provides evidence that the scores indeed reflect college- and career readiness and is a good starting place for extrapolations for tracking progress. All these efforts are necessary for credibility and utility of the scores by the public in general and higher education institutions. Trustworthiness of the scores from the assessment system and these external criteria can be empowering to maintain students on a believable college readiness track, and in enabling them to enter and be successful in college.

In addition, the consortium plans to conduct validity studies to examine the extent to which the assessments have been internationally benchmarked by comparing performance on the consortium assessments and those on TIMSS, PISA and PIRLS for the age groups in the consortium who are taking these tests. It will also investigate the assessment content on these tests with those of the assessments in the consortium. These two sets of validation studies, (1) linking to college admission and work readiness tests and (2) linking to international assessments are essential for validity of college- and career readiness scores and one of the key elements of the consortium's excellent research and evaluation plan.

Research and Evaluation. The consortium plans to conduct research and evaluation studies after the 4-year granting period once the assessments have been implemented. The plan includes researching and evaluation of the impact of the theory of action such as whether the system is improving instruction as intended and the impact the system is having on student achievement in general and on rates of student readiness for college and careers in particular.

Missing. Three issues are missing from the research and evaluation plan. These are (1) ongoing quality assurance procedures for examining and verifying the quality of processes and deliverables; (2) cognitive studies for examining what the items are measuring; and (3) research for examining measurement accuracy for all performance levels.

No specific research studies or quality control procedures are described for examining the alignment of the assessments with CCSS. Alignment of assessments with curriculum and instruction is essential for formative use and broader interpretations of assessment results. In addition, such alignment information is necessary for determining the types of changes needed in curriculum and instruction to improve learning.

One aspect of the validation that is missing is gathering evidence to demonstrate that the tests are capturing the intended cognitive processes. This type of validation research would utilize cognitive labs and or think-aloud protocols. In particular, it is essential to examine what kinds of cognitive demands "innovative" items are tapping in order to evaluate their contribution to the assessment. The proposed plan does not include any cognitive studies to investigate what the items are measuring.

Measurement accuracy is closely tied to the match between the ability levels of students and the difficulty level of the assessment. The proposed plan states that the assessments will provide similar measurement accuracy for students at different performance levels. However, obtaining similar levels of measurement accuracy with a single test form is unlikely. The consortium has not presented a plan for research and evaluation to examine the degree to which the assessments will meet the requirement of good measurement accuracy for all performance levels.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	5
(A)(6) Reviewer Comments:		

The consortium's plan for professional capacity development and outreach includes four main components: (1) building a leadership cadre of content experts, (2) developing training tools to help educators implement the assessment system, (3) developing training tools for them to interpret and use assessment results, and (4) sharing advice to help educators and administrators understand and address the curricular and instructional implications of the CCSS and the Partnership's aligned assessments. All of these components are essential for proper implementation of the assessment system. However, they lack explicit efforts to provide professional development for educators to help improve learning. The plan's **Training tools for interpreting and using assessment results** are described as

- Conduct gap analyses between the state standards and the CCSS, with a focus on curricular and instructional planning necessary to address the gaps and changes;
- Create model curriculum frameworks to serve as a bridge between the CCSS and the new assessment system; and
- Create and share exemplar lesson plans and student work nominated and evaluated by teaching experts and subject-matter leaders in the Partnership.

Provision of these tools electronically is an indirect way of engaging teachers with the assessment data. No information is provided about how the teachers are expected to make to use of these materials to improve learning.

Possible effects of the lack of teacher agency and choice in the proposed assessment plan is evident to the consortium: "the Partnership's distributed approach to summative assessment might be interpreted as an attempt to introduce "more test days" into the school calendar and raise fears about "over-testing" of students."

As a preventive strategy " Partnership will develop materials and messages that help prevent such misunderstandings while emphasizing the benefits of the new system of common assessments." The proposed plan's strategy is to explain and illustrate that learning will not stop while assessments are being administered, tests are far beyond "bubble test", assessments provide rich and frequent information, data from assessments is much more relevant and useful.

These strategies are to remedy problems that have been associated with large-scale assessments for decades. In assessment systems with useful information for teachers and students, which give teachers and students agency about their teaching and learning, the benefits of the assessment system should not require such justification. The proposed assessment system is overall well-designed and psychometrically-sound but gives only cursory consideration to teacher choice, agency and involvement on an ongoing basis. Without full teacher involvement, assessments can drive instruction as a way to improve performance on tests rather than student learning, in the best case scenario.

Outreach. The consortium has four components to its communication and outreach activities.

Public Outreach that focuses on developing tools to help states inform the public about the assessment system.

Targeted Coalition Building will build coalitions of supporters among key stake-holder groups by engaging them in different aspects of the system.

College-Ready Outreach Strategy for Students and Families by informing students and parents about how information from the assessment system can be used to ensure that students are college and career ready.

The three strategies for outreach to the public and students and families are essential for public accountability of the assessment system. However, these plans lack details about how states will develop communication and outreach strategies that are targeted to the needs of particular states. They also assume a one-way communication strategy can ascertain development of support for the assessment system, instead of a more interactive communication strategy with focus groups or other mechanisms that allow stakeholders and the public to provide feedback and suggestions for the system which is more likely to develop support for the system.

Higher Education Engagement Strategy by involving higher education leaders and faculty in design and development of the assessment system. The consortium has identified strategies to deepen understanding and commitment between the IHEs and the consortium. The College Ready Advisory committee will (1) convene several times a year to plan for critical junctures in the Partnership's work plan; (2) will reach out to higher education associations to seek regular feedback from their members and promote alignment with the Partnership's work; (3) will collaborate with the Partnership's design committees to design and develop the high school assessments; (4) will engage colleagues to review the CCSS, discuss how the standards will address the gaps that students' bring to their classrooms and provide input on how the Partnership's high school assessments can measure the CCSS so results can be used for placement decisions; (5) will lead research on the implementation and implications of using results from the high school assessments for placement decisions. The regular engagement of IHE members through the College Ready Advisory Committee in the assessment design and development and implementation of the assessment system are effective strategies for enabling the assessment system to function as a bridge between grades 3-12 and higher education and for preparing students for college and career readiness.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	3
<p>(A)(7) Reviewer Comments:</p> <p>The consortium is planning to use technology to support all aspects of the assessment system. These include item development, item and assessment management, test administration, scoring, data management, reporting, data availability and professional development. No details are provided about the centralization of these functions and processes at a single portal or how systems for different operations will be connected to create an efficient system that can be easily monitored for quality control.</p> <p>An open-source integrated system will be created to host, manage and supply data according to the interests and needs of stake-holders. The consortium will work with the partner states to develop online system specifications and minimum school and district technology requirements. However, since the consortium has not yet 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 will develop a site certification process so that states can ensure all schools meet minimum technology requirements for implementation of the assessment system. This is a necessary step for determining preparedness for all school sites for the assessment system. However, no information is provided about systematically identifying and taking into account existing barriers for schools within the participating states and how the barriers will be addressed so that all schools will be able to fully participate in the assessment system. Identification of the barriers for states in the application development phase in collaboration with the participating states would have enabled the consortium to propose plans that would take into account at least some of these barriers in the initial planning stage.</p> <p>Inadequate use of technology. In the proposed plan, the assessment design does not include effective use of technology for creating tests that can be adapted for different performance levels. For example, in a computerized testing environment, students can be routed to different test forms depending on their</p>		

performance level on the routing test. This type of multi-stage testing is expected to provide better measurement accuracy with shorter tests targeted to student performance levels. With the wide ranging student performance levels across the states in the consortium, such use of technology is not adequately explored in the proposed plan.

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	22

(A)(8) Reviewer Comments:

Project management partner. The consortium identified Achieve as their Management Partner through a competitive process. Achieve presents information about their involvement in other projects and lists extensive experience in many areas that are relevant to the successful management of the consortium's activities. These include multi-state consortium management, multi-state assessment development, development of educational standards, including CCSS, quality reviews of standards and assessments; establishment of college-ready assessments and cut scores, facilitation of K-12 and higher education collaboration on college-ready agenda nationally and in states, strategic advice on the development and implementation of state CCR policies, including standards, assessments, graduation requirements, data and accountability policies, outreach and communications support to states around CCR policies; and support to states in building sustainability plans for large-scale reforms.

Achieve's role is to manage and coordinate the work of relevant design committees and contractors, including those responsible for developing the assessments. Successful coordination and management of the work across committees and contractors is essential for the assessment system. Achieve's experiences and capacities in assessment and multi-state consortia will be a great strength for the consortium.

Project work plan and Timeline. The proposed work plan presented a detailed timeline for all 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.

Sustainability over time. The consortium estimates cost per student as \$32.68. This estimate will constitute savings for most states in the consortium. The states are expected to use funds currently allocated to existing state assessments that will be replaced by the Partnership assessments, and, if necessary, secure additional funding. As part of their MOUs, states indicate their plan for securing the funds necessary to implement the assessment system statewide. Overall, savings in most states will make the sustainability of the assessment system feasible.

The consortium's consideration of the range from \$17-\$18 to \$45-\$50 per student, as well as their identification of factors that may affect the cost the most (as scoring procedures) demonstrates their thoughtful approach to the cost estimation. However, it is not clear what these estimates are based on. 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.

The consortium has allocated too little resources to quality control and quality assurance processes. This is evident in the budget allocation of less than \$200,000 funds for four years for quality control procedures.

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:

The consortium has secured commitment from 188 public institutions and systems of higher education, representing 90% of the direct matriculation students in public IHEs in all participating states. The letters of commitment indicate participation by states in

- (1) A collaborative and comprehensive effort by K-12 and higher education faculty and leaders across the [Partnership] on test design and development.
- (2) A coordinated effort across the consortium to design and participate in validity studies and comparisons with current placement instruments to ensure that the assessments developed are an accurate measure of college readiness.
- (3) A thorough, research-based process to establish common achievement standards on the new assessments that signal students' preparation for entry level, credit-bearing coursework.
- (4) Use of the assessment in all partnership states' post secondary institutions as an indicator of students' readiness for placement in non-remedial, credit-bearing college-level coursework.

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-four of the participating states obtained commitments from IHEs that represent large proportions (well over 50%) of the direct matriculation students in the respective states. The rest of the two states (Mississippi and New Jersey) obtained commitment from IHEs that represent a smaller proportion of direct matriculation students in these states (20%, 21% respectively). All states signed letters that indicated their strong commitment to the consortium's goals.

The consortium will involve higher education stakeholders in the ongoing design and development of the assessment system. This collaboration has great potential to have an impact on college- and career readiness in over half of the states in the country.

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 has demonstrated in its application that the proposed assessment system will measure student knowledge and skills against a common set of college- and career-ready standards in mathematics

and English language arts. These assessments will be targeted to cover the full range of those standards, including standards against which student achievement has traditionally been difficult to measure.

The consortium plans to use a variety of item types that will be designed to assess complex higher order thinking. The current plans do not include verification and validation of measurement of complex thinking processes using cognitive labs or interviews.

The assessment design includes the use of both easy and difficult items in order for assessments to provide similar levels of measurement accuracy across the full performance continuum, including for high- and low-achieving students. Whether the assessment system will provide accurate measures of student growth over a full academic year or course will depend on the degree to which the assessments have appropriate difficulty for students. Administration of single test forms (for a given administration, not including tests forms with accommodations), is not sufficient to ascertain similar measurement accuracy across the performance spectrum. The great diversity in the performance spectrum across the states and the limited number of test items each student can take create key challenges to meeting this requirement. The proposed plan does not indicate whether multiple test form administrations will be considered to establish similar measurement accuracy for all performance levels.

The assessment system includes ELA/Literacy assessments which consist of two focused literacy assessments, an extended research/writing assessment, a speaking and listening assessment, and an end-of-year assessment. The mathematics assessments will have three through-course and an end-of-year component and together will contribute to an annual combined score. These assessments will be used to produce student achievement and growth data that can be used to determine whether individual students are college- and career-ready.

All students, including English learners and students with disabilities, will be assessed using the proposed assessment system.

The proposed assessment system will produce data, including student achievement and growth data that can be used to inform (1) evaluations of school, principal and teacher effectiveness; (2) for identifying needs for support and professional development for school principals and teachers; and (3) teaching, learning, and program improvement.

Key recommendation.

In the current plan, details are not provided about how similar measurement accuracy will be established for all performance levels. A single test form for each grade is unlikely to ascertain this type of measurement accuracy. If the current plan is based on a single test form, a revision to the assessment design is recommended. This revision should include the flexibility of administering different test forms with different difficulty levels that will be targeted to student performance levels. This flexibility can be in the form of alternate forms, multi-stage branching tests or computerized adaptive testing.

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

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

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 **Level 1 Budget** is presented with great detail organized by each of the selection criteria: Governance (\$2.7 million); Fiscal Agent Budget (\$2 million); Support for Governing States (\$7.3 million); Assessment Design and Development (\$80.5 million); Research and Evaluation (\$8.5 million); Professional Capacity and Outreach (\$6.5 million); Technology (\$22.5 million); Higher Education Engagement (\$4 million); and Project Management (\$16 million).The allocation of costs to

these components is proportional to the emphasis indicated in the Theory Action, with the majority of the budget (70%) being allocated to the design, development and technology associated with the assessment system.

The budget is presented by each of the modules above as well as by the different budget categories (personnel, equipment, etc.) by year for the four years of the project. A close review of estimated salaries, travel, supplies and other budget categories indicate a reasonable estimation for each of these categories and for the great majority of the modules.

The budget for quality assurance for assessment design and development is estimated to be \$184,690. It is indicated that this amount will cover the labor to review test decks, scoring systems, and other programs. This amount seems very low given the size of the project. Over a 4-year project, the estimated cost for quality assurance will be less than \$50,000 annually. This amount is less than one full time content or technology expert's salary which is not sufficient for reviewing test items, test decks, scoring systems and other programs.

Level 2 Budgets

Name: Formative Assessment Tools for K-2

The Level 2 budget includes requests for the following two Level 2 modules: Formative Assessment Tools for K-2 (Priority 1) and Technology Enhancements (Priority 2). The consortium is estimating approximately \$10 million for each of these modules. Formative assessment tools are essential for supporting teachers to improve learning, as are the technological capabilities of states for successful implementation of the system. However, more information is needed to judge the reasonableness of the estimated costs for these efforts.

Formative assessment tools for K - 2 Literacy and Mathematics. These formative assessment tools can indeed be useful for teachers in K - 2 to tailor their teaching to student needs. However, it is difficult to judge the reasonableness of the cost estimate of \$10,000,000 for these assessments. Details are not provided about why the cost for content development would add up to \$7.2 million. Also, the consortium does not provide information about exploring cost efficiencies that may be gained from using existing K - 2 tests (in some states) for formative purposes.

Name: Technology Enhancements

Technology enhancements. The technology enhancements are intended to support infrastructure to minimize the risk of system overload during test administration and accelerate the development of test delivery and scoring technologies. Too little information is provided about the necessity for such an enhancement, in other words, what the risks are associated with not having this enhancement. In addition, cost estimates are given in two broad categories (Technology Infrastructure and Test Delivery and Scoring) with no details about these categories. For example, no details are provided about how the cost is divided between equipment, software and labor.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b) (6)

(A)(1) Consortium Governance

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

(A)(1) Reviewer Comments:

The organizational structure of the consortium and the differentiated roles for states seem clear. Rights and responsibilities for each role are nicely described. The consortium's decision-making processes seem thoughtful and representative of the roles that states are taking. These include that each Governing State chief state school officers will serve on the Partnership's Governing Board and the governing states will directly nominate chair candidates. Decision-making is specified by consensus, with a clear protocol for super majority decision procedures when complete consensus is not present. The Assessment Design Committee is described and reports appropriately to the Governing Board. Policies for how day-to-day decisions will be made without calling on the governing board more frequently than may be possible is unclear, and could lead to some delays given the infrequent meetings of the Governing Board.

Protocols for joining and leaving the consortium seem clear, with both an application process for new members clearly stated including requirements for approval of the application, and procedures to be carried out for membership opt-out processes. Participation requirements include a recommitment requirement policy if the state leadership of governor or chief state school officer changes during the four-year project period, which should help keep states focused on both achieving the partnership goals and educating/informing during leadership changes.

The plan for managing funds through the fiduciary state seems reasonable. With one state serving as fiscal agent, procurement issues are likely to be streamlined or at least possible to negotiate within more of a single context of state regulations. The contract with Achieve can also be managed by the fiduciary state. However, the streamlining of the financial process does not compromise decision-making authority, as decisions regarding expenditures and disbursements remain the jurisdiction of the Governing Board of all Governing States.

As described in the Theory of Action review section, the consortium's vision does seem somewhat limited or incompletely described in terms of use for learning outcomes. Strategies for how teachers and schools should or could use the data to make differences in instruction and change learning outcomes directly for students are minimally described in the proposal. However, for what is described in terms of design, development, deployment and data collection of the operational system, the governance structure seems reasonably robust.

Most of the consortium's goals, roles and key deliverables seem consistent with the theory of action (though the theory of action itself is limited as described above). The professional development activities seem less coherently described, and it is unclear what deliverables will result for teacher professional development, particularly how teachers and administrators will be instructed to use the assessment system to make actionable instructional decisions for students.

Processes and timelines for setting key policies and definitions are established. The preliminary standard setting process is helpful.

The state diversity represented here is commendable. Plans for identifying any state barriers may need strengthening and more detailed policy and role assignment. Timelines seem as if they should allow more sufficient opportunities for impact dialogues and resolution.

Recommendation: It would be helpful for the Governing Board to generate some more specific bylaws for delegation of decision-making, so that subcommittees and working groups do not experience too many delays and miss key milestones on timelines.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	3
(A)(2) Reviewer Comments:		
<p>The applicant's theory of action seems logical and coherent in terms of data collection and display. The degree of innovation seems not as extensive as might be employed to support the theory of action's intent to substantially enhance learning gains. The instructional uses seem insufficiently described to fully support the theory of action as credible in instructional decision-making for teachers and administrators. The system is described as primarily summative, but the theory of action does not adequately describe how summative uses of the system will bring about the gains described.</p> <p>The anchoring to the common core standards, the quarterly opportunities for assessment, and the opportunity for inclusion of some richer tasks should be helpful to students. It is helpful for support of the theory of action that all states remaining in the partnership have agreed to engage in all the time points once the system is operational. The components of through-course tasks and end-of-year assessment seem well-related to each other. All will be framed around the common standards, and seem to include sufficient commonality but also unique roles and structures between and among the components.</p> <p>More clarity and vision on how the results will inform teacher practice and will be used instructionally and/or for accountability to change learning outcomes would support a more compelling theory of action. There is insufficient information on how the assessment results produced by each component will be used as compared to simply collected and displayed. This also brings into some question exactly how the educational system as a whole will improve student achievement and college/career readiness; in essence, having information is not the same as showing how use of the system will influence learning outcomes in effective ways.</p>		

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	25
(A)(3) Reviewer Comments:		
<p>Designs of the end-of-the-year summative assessments and the through-year interim/summative assessments show some elements of being innovative and are somewhat feasible and consistent with the theory of action, but substantial issues and concerns do arise. First, a basic overview of the specific constructs and scores to be reported is not sufficiently detailed in the proposal. Such information is needed to help determine if it is feasible to generate accurate scores within the design and constraints of the system specified. While constructs and scores are described as reflecting coverage of the common core, important information such as how many separate scores are expected to be reported at the quarterly timepoints is needed to evaluate some aspects of feasibility. How PARCC intends to translate the domain (Common Core State Standards) into operational measurement constructs and what at least a high</p>		

level overview of these measurement constructs would be essential to evaluate this portion of the proposal. This makes it difficult to assess the adequacy of the assessment system design.

The number and types of components for formative assessment are not well described, except to the extent that the proposal does acknowledge this is not a major focus of the initiative. This is an important concern in evaluating this proposal, since formative assessments are closely linked in many studies in the educational research literature with large effect sizes in improving student learning outcomes. Yet formative assessments are not employed to much effect to help students and teachers in this proposal.

At the item level, the assessment system does describe how innovative assessments will be used to consider student knowledge and skills across the full range of standards, and in areas traditionally difficult to measure. Concerns remain as described above that insufficient detail is presented on the constructs and aggregation to determine if students will actually be well measured across the standards, as called for in the NIA. The longer performance tasks are commendable, and will be helpful for college readiness measures if valid scores can be reported.

It is unclear in this proposal how or whether the system will be able to provide an accurate measure for student performance for high- and low-performing students. Without knowing what and how many constructs will be reported, and over what amount of test-taking time, it is hard to gauge this for any students, at any proficiency level. The challenges become even more complex for students more at the extremes. These appear to be fixed-form instruments, except for possibly selection of reading passages. For fixed-form instruments it would be important to know much more about the test characteristics to form a judgment on this. However, the key personnel associated with this project do have substantial experience in these areas, at least for more well-defined and previously assessed constructs. Research studies are specified and there is good attention to concerns of universal design.

The intended timing, types of items, and mode of administration for the summative components are clear. However, regarding timing, for through-course assessments to be feasible and fair, and to be meaningful and useful for educators, students should have all had the opportunity to learn the material on which they are being assessed. Since the states will not have adopted common curricular materials, nor agreed to time points at which standards are taught according to this proposal, it is unclear how in this proposal the sequence of instruction and timing of instruction relative to the common core standards will be established in each school and classroom quarterly, such that through-course assessments can deliver meaningful information for students, teachers and schools.

The methods for scoring seem reasonable except in the case of where AI (Artificial Intelligence) is specified. The presentation in the proposal does not provide sufficient detail to know what AI will be used, how and whether concerns could be addressed. Also, the proposal does mention at one point that states may choose to do a lower percentage of read-behinds or overlap scoring, which could undermine the scoring arguments made in this section of the proposal, since a sufficient number of read-behinds and overlaps can help identify problems and issues with automated scoring.

(A)(4) Assessment System Development

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

(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 approaches for developing assessment items are evidence-centered and the intention is to incorporate universal design practices. The rationale for using these approaches is well-supported. Development phases and processes as well as personnel to be involved seem robust, in most cases, though a few

critical concerns do present themselves. The main concerns here are (i) the date for standard setting is late and is past the duration of the proposal, (ii) the building of test forms and all the stages of data aggregation into constructs and scores and test assembly seem either extremely underdefined for the intents, purposes, and needs of this proposal, or are specified far too late in the process, (iii) the procedures for growth modeling and sufficient consensus among the states in such areas as vertical scaling do not seem present for the methodology described in tracking and reporting, and for comparability requirements of the proposal, and (iv) the empirical validation process with the higher education institutions is likely under scoped.

In relation to concern (i) above, it is helpful that there is preliminary standard setting stage specified earlier in the process. This may allow some concerns to surface and be addressed. A primary issue is likely to be that the standards needed to actually ensure college/career readiness may be far higher than state achievement levels, leading to substantial impact concerns.

The approach and strategy for designing and developing accommodations involves asking appropriate questions and seeking appropriate expertise. Designing and developing accommodations and associated policies and methods seems possible within this plan. However, as described previously it is not entirely clear that the assessment design can accommodate all levels of learners, providing accuracy of measurement across the full continuum, within a fixed form format as seems to be specified in most areas.

Although there is no specific quality control plan fully described in the proposal, a number of approaches for addressing quality control have been included. In most areas these seem reasonably robust, including field testing, research studies and the expertise of the personnel involved. However, a caveat in the research and evaluation plan, as mentioned in that portion of this review, is that not all studies will be implemented, and there is no prioritization of the studies. Additionally, a key aspect of quality control for development involves how accurate measures for all students will be generated from fixed forms. This is not thoroughly addressed, and may require out-of-grade test forms and alternate assessments, which are not well described in the proposal.

Recommendation: An external evaluator to provide formative evaluation and to review the accuracy of student measures across the distribution seems helpful to the implementation of this project. Suggestions and ideas from an external viewpoint would help to provide perspective.

(A)(5) Research and Evaluation

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

(A)(5) Reviewer Comments:

This proposal includes an extensive array of associated research studies. However, it should be noted that the proposal also states in one brief concluding section that not all the studies will be completed, and priority and range of studies are yet to be determined. This suggests that some unspecified subset of the studies will be completed, which makes it difficult to determine to what degree information will be available for research and evaluation.

If all the studies were to be completed, and in robust ways, it does appear that sufficient *information* would be available to assess the appropriateness of the summative components (not the formative components though), and to come to some conclusions on the validity, reliability, fairness, precision and comparability. The information may reveal that some aspects of this system are inadequate in certain of these areas, as described previously. How the information will be addressed by the consortium so that the end result ensures that the assessments have these characteristics, as required by the NIA, is not clear.

In particular, the plan for determining if the assessments are being implemented as designed suggests the need for clarity from the consortium on external evaluation. Establishing whether the intended effects on individuals are being achieved seems to call for a strengthening of the empirical validity studies at the post-secondary level with the identified IHEs, and also the establishment of some studies at initial career for

those not pursuing higher education goals. In RFI-1 on pg. 197, the consortium acknowledges that the system will address "constructs that rarely, if ever, have been measured by large-scale assessments in the United States." This seems to support the need for more extensive empirical studies of post-secondary validation of the theoretical constructs. One study is mentioned on p. 204 ("analyzing the relationship between scores on the Partnership's assessments and actual performance in college courses). This is good but the extent, range and length of the study; whether it will actually happen given the caveat mentioned above; and the degree of involvement and commitment of the many IHEs that would be integral to such a study contribute to concerns of whether the measures indeed are well associated with actual subsequent career and college readiness.

Other key questions are raised but not addressed, such as teacher effectiveness and evaluation. The proposal discusses as a concern that teachers gain valuable and useful information, but the consortium's proposal specifies little and designates only a small percentage of discussion and funding toward how teachers will operationally make the information valuable and useful in their teaching and decision-making processes.

The intended involvement of the Bill & Melinda Gates Foundation as specified in the proposal to host resource roundtables is helpful, due to the empirical work the foundation is doing in post-secondary college readiness studies.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	8

(A)(6) Reviewer Comments:

The consortium's plan for supporting teachers and administrators in the stage of *implementing* the assessment system seems fairly robust. The described courses and workshops for implementation of the system's data collection processes, along with the experience of the personnel, seem well suited to these needs. It is not clear how many and to what extent teachers will take part in these activities, for instance whether they will be required or at the discretion of states or other local entities that may have to compete for teacher time and availability with other school processes.

Despite some of these strengths, the proposal seems weakly described in developing the professional capacity of educators to use the assessments and results to inform and improve instructional practice. Strategies are not well described for this. It seems as if this trend can be seen systematically throughout the proposal: practices for data collection and summative use of the system to help judge student performance tend to do a good job of following current accepted practices in large-scale assessments. However, often large-scale assessments in the U.S. have been less used for the second target, of directly informing classroom and school decision-making. Here is where the theory of action, proposal outlines, goals, objectives, components, funding and the experience of the professional team for this consortium seem somewhat limited to fully address the needs, scope and opportunity of the NIA.

The communications strategy and plan for informing the public and key stakeholders, including legislators, seem adequate in most respects. However, a challenge for this proposal will be gaining support from schools and communities for what may be seen as an increase in testing and assessment, especially if the utility of the data use within schools and for students to directly improve learning outcomes is not well described or well understood by schools. Having the theory of action more robustly address direct data utility in schools would improve the ability to communicate to stakeholders.

Recommendation: It might be important to more robustly inform legislators sooner of the need to investigate any barriers to adoption of the system. This is mentioned briefly as a role of the fiduciary state on pg. 22 of the proposal but seems to have limited focus in the proposal. Additionally, it is important to provide state

leadership and policy makers with sufficient standard setting and impact data sooner so that they can engage in any dialogues that may result in needs for adjustment in the systems.

(A)(7) Technology Approach

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

(A)(7) Reviewer Comments:

The technology approach described has some nice strengths. These include the desire to establish interoperability standards and to establish a marketplace for technology partners to provide services.

Additionally, the integrated modular approach described is also very much to be appreciated, if it can be achieved, especially the desire to integrate data systems. This is a challenging but very important goal, to well-position schools for use of data. If numerous data systems are in play, with a vast variety of interfaces, tools and reports that schools need to master, and this proposal's assessment system becomes an additional add-on system, the burden on schools, teachers, and administrators becomes large, and effective evidence in decision-making processes is less likely to happen. Not only consolidating systems but allowing integration where other data sources will need to continue to exist is a helpful priority of this proposal. However, systems architecture, back end and database development are not yet well described for this project, and need considerable attention to meet the proposal's specifications of what will be accomplished.

Technology use in item development and item analysis is well described and seems promising, except for (i) the scoring approaches as discussed previously, where more clarity on the AI involved would be helpful, and (ii) examples of innovative item types in this proposal tend to be more migratory from paper-and-pencil and less transformative in full use of the computer platform. The various technology tools described for data management, reporting, data availability, and professional development seem helpful.

The Partnership does state that there are large challenges in ensuring that schools and districts have the necessary technology infrastructure to support schoolwide CBT (computer-based testing) several times per year. How these challenges will be addressed is not sufficiently discussed. It should be noted that depending on existing infrastructure to contribute much of the digital resource will likely mean taking important technology resources that are operational in schools now for teaching and learning away from their current uses. This can mean monopolizing the technology resource in schools and using them nearly exclusively for the purpose of assessment, which would indeed be unfortunate for students.

(A)(8) Project Management

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

(A)(8) Reviewer Comments:

The quality, qualifications and role of the project management partner seem reasonably matched to the Partnership goals. The experience of the personnel involved seems to especially match the summative aspects of the proposal guidelines, and the Partnership acknowledges primarily pursuing a summative set of goals and objectives.

The project timeline, major milestones, deadlines and division of duties seem well conceived, except where discussed previously in the construct explication, test aggregation, and standard setting areas. These previous concerns are described more at length in the Assessment System Design and Assessment System Development section, but are mentioned here again because important deliverables that seem

missing or problematic are clear operational definitions of the constructs, more information on how tests will be aggregated to provide valid and reliable assessments at extremes of the scale without alternate assessments, and earlier milestones for the standard setting final impacts. The experience of the project management partner does help to mitigate risk of the many unknowns in the proposal, as prior experience seems suitable for the summative aspects of the focus, although assessment development at this scale has not been done before based on information provided in this proposal.

Some concerns overall for the funding priorities are discussed elsewhere in this review and they include: (i) the extensive body of research studies seems not possible to complete under the funding allocation, and the proposal should more clearly describe which studies will actually be engaged in rather than presenting a catch-all that suggests more than will be done, (ii) the partnership with Higher Education should better support empirical validation of constructs and outcomes, as well as establish initial career validity studies, (iii) the estimated costs to the State should include more information on addressing the digital infrastructure and professional capacity building costs, which are likely to be substantial under this proposal. Also, caveats about state decisions on lower rates of read-behind scoring and other investments that support the complexity of the system should be more clearly detailed as either required or optional in the proposal, and the validity and use arguments adjusted accordingly.

Competitive Preference Priority: Collaboration and Alignment with Higher Education

	Available	Score
Competitive Preference Priority: Collaboration and Alignment with Higher Education	20	12

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 more limited 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 concern for this proposal and the commitment of the IHE's to the Partnership to ensure that the assessments assess college readiness (priority portion a) is that only one validity study empirically evaluating the relationship of the assessments to the intended outcome is described. This seems insufficient commitment by the IHEs to the needs for establishing (a) as a foundation for this section. This is a critical foundation so a large limitation in the commitment. There also is a caveat in the proposal stating that only a subset of the research studies will be completed, further undermining the commitment.

Recommendation: Encourage each of the IHEs to include empirical work in addition to the specified policy position and expert opinion in their commitment to the success of the 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

Absolute Reviewer Comments:

The Absolute Priority seems to be reasonably well met in this proposal. The degree of actual demonstration in the proposal that accurate scores can be generated across the full performance continuum within the specified design seems potentially weak due to some of the technical characteristics as described in the Assessment Design section, such as standard errors over the continuum, change scores and the capacity

of the fixed form tests to generate sufficient information at four time points given the likely available test-taking time. However, the proposal has indicated a structure to consider and address questions of technical adequacy that provides reasonable assurance that such concerns if they arise will be identified and addressed.

Recommendation: An external evaluator to provide formative advice and to consider questions of technical adequacy described above would be helpful. Additionally, the Partnership should be asked to more fully describe the constructs to be measured, score(s) to be reported, test-taking time available for each student, and how the assessment design will be able to generate accurate scores over the full performance continuum for these constructs within these constraints.

Grand Total	220	144
-------------	-----	-----

Budgets

Level 1 Budget

Name: Level 1 Budget(s)

After combining costs on summary table for Governance, Fiscal Agent Budget, Support for Governing States and Project Management, the total cost of governance and management together seems high as a percentage of the total budget for the Level 1 Budget Module.

Professional Capacity and Outreach seems under-resourced for the needs of the schools and teachers. It is not clear where additional resources will be obtained or drawn from existing resources repurposed for this need.

Research and Evaluation funds seem appropriate as a percentage of the total Budget 1 module; however, the Partnership acknowledges they will be insufficient for all the studies described in the tables of research studies. It would be important to see the prioritizing of the studies, and how the money will actually be spent to achieve or not achieve the various objectives.

Higher Education Engagement funds grow over time. It seems essential that the involvement of higher education be prioritized to better include the initial stages, such that alignment studies are established from the outset.

Travel funds seem high and perhaps possible to reduce through more virtual meetings and dialogue, especially considering the strong technology focus and priority of the proposal.

Level 2 Budgets

Name: Formative Assessment Tools for K-2

The content development, labor, production, scoring, travel and coordination costs for the Budget Module 2/Priority 1 of Formative Assessment for K-2 seem reasonable. However, whether this priority should be implemented at this time should be carefully considered and some thought given to whether it is well positioned within this proposal at this time. First, there is a very limited agenda for formative assessment in the main proposal. This has been detailed as a limitation in other sections of this review. Rather than trying to expand outside of the goals of the competition, it seems more attention to teacher use and decision-making for instruction should be addressed within the main priorities before taking on additional efforts. Secondly, efforts at achieving success for career and college readiness at the grade levels specified in the main proposal should be established, researched and evaluated before attempting to expand the range to earlier ages.

However, it is true that many teachers and schools would like more assessment information for the K-2 grades. Funding this budget priority might be an opportunity to provide them with useful information, if

questions of how PARCC will use information formatively and how the data collected in such early grades connects with constructs of career/college readiness can be addressed.

Name: Technology Enhancements

The \$10 million in this portion of the budget request is divided into two parts. The first portion of \$5 million is for strategic investment in scoring and automated scoring approaches. This could be an important addition to this proposal, because the \$5 million in requested funding would be helpful to move the field forward and better support the ability to use innovative and richer assessments for higher order thinking. However, the budget detail for how this funding will be used is sparse, and therefore it is not possible to fully evaluate how it will be spent.

The other half of the requested funding for this proposal under Budget Module 2/Priority 2 is to support technology infrastructure. This is an operational cost engendered by the assessment design in the main proposal. As such, it seems as if it should not be considered a technology enhancement, but should be incorporated within the main funding as an operational cost, or the partnership should shown how existing state resources will be repurposed for this need.



Race to the Top

Comprehensive Assessment Systems Technical Review Form



PARCC Application #PARCC (b)(6)

(A)(1) Consortium Governance

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

(A)(1) Reviewer Comments:

Overall, the evidence provided by the proposal indicates that the consortium is capable and qualified with regard to governance.

(a)

PARCC's vision is clear and concise, i.e. to develop an assessment system that will provide the kind of information parents and educators need to improve student college and career readiness and thereby ensure that more students graduate from high school prepared for postsecondary success.

However, the vision emphasizes "postsecondary success" at the expense of "career success." Balance is needed regarding students who do not continue at an institution of higher education (IHE). The proposal discusses collaborating with higher education stakeholders to develop a clear definition of "college and career readiness." It does not sufficiently consider the needs of students who do not continue their education at an IHE.

(b)

PARCC's structure and operations with regard to states are well described and reasonable. The explicit reference to by-laws for the Governing Board and to charters for committees is laudable. The explicit involvement of the chief state school officers should help to provide necessary leadership. Roles and responsibilities of the participants were well described.

A qualified project management partner (ACHIEVE) has already been selected through a competitive process. (See section (A)(8) for evidence of capacity and qualifications.) Although Kentucky conducted the procurement process, the agreement is to be renegotiated with Florida. Because different states have differing contracting procedures and regulations, it is to be hoped that the negotiation will go smoothly. The organizational chart is vague with regard to the project management partner's (PMP) lines of authority and reporting.

PARCC wisely includes a recommitment requirement for states that experience a change in political leadership.

(c)

The memorandums of understanding (MOU) are consistent across states. Appendix 2 documents states' plan for identifying barriers. However a number of states provided a formulaic plan that merely refers to a review of law and regulation. An immediate study is needed to describe existing barriers and to begin to assist states in implementing the assessment by 2014-15.

(d)

The MOU documents states' commitment to the procurement process, which will conform to the lead procurement State's (Florida) laws and regulations.

Recommendations

PARCC should consider augmenting its vision statement to address the interests of non-college bound students and community/junior college job training programs.

(A)(2) Theory of Action

	Available	Score
(A)(2) Theory of Action	5	5
(A)(2) Reviewer Comments:		
<p>Overall, the evidence provided by the proposal indicates an exceptional theory of action.</p> <p>The proposal clearly lays out a logical extension of the theory of action that underlies current ESEA policies – standards aligned with curriculum, teaching and assessment, plus an accountability system that provides incentives. The extension is to include CCR as a goal using common standards and assessments across states. Assessment components, uses, systemic application and processes are clearly and credibly described.</p> <p>The theory of action clearly relies on school and staff accountability to leverage improvements in student achievement. Also present in the theory is the concept that the assessment should help to guide instruction. For example, summative tests in the form of "through-course-assessments" can be used formatively to guide teaching.</p>		

(A)(3) Assessment System Design

	Available	Score
(A)(3) Assessment System Design	55	51
(A)(3) Reviewer Comments:		
<p>Overall, the proposal describes a high quality assessment design.</p> <p>(a)</p> <p>The proposal describes in clear detail the number and types of components of the assessment system. Included are several components administered throughout the year for English Language Arts (ELA) and Math, followed by an end of year component. The summary tables for ELA and Math are comprehensive.</p> <p>Laudably, the tests are to be heavily weighted with performance tasks and constructed response (CR) items.</p> <p>Useful formative tools include an on-line resource center, a tool to assess student reading levels, and optional performance tasks. Resources already under development by member states are to be disseminated more widely.</p> <p>(b)</p> <p>The proposal explicitly describes how test content is to align with the CCSS and assess the core standards in each grade/cluster.</p> <p>A strength of the proposal's approach to Math testing is the flexibility with regard to both traditional and integrated curricula.</p> <p>Decisions regarding the approach to high school assessment are to be guided by consultation with IHEs and policymakers.</p>		

(c)

The proposal's commitment to timely reporting of all assessment components, so that teachers may act upon the information, is commendable.

A strength of the proposal is its recognition that some paper/pencil administration and scoring may be needed for younger students with limited computer skills, versus computer administration for older students. However, computer administration will be used where feasible and is needed to keep costs within reason.

The proposal suggests that final College and Career Readiness (CCR) standards cannot be set until after the 2014-15 operational administration (after the grant period is ended). While it is reasonable to wait until after the operational administration to finalize standards, it is possible to set good preliminary standards on the basis of a well-designed field test.

Testing in grades 9 – 11, not just once in high school, should yield better measures of growth towards CCR.

The proposal confirms that data will be available to identify professional development needs. However, a study is to be conducted to assist states in developing or refining their teacher/principal accountability systems. This suggests that PARCC does not intend to implement a centralized approach to measuring student growth for teacher and principal evaluation.

The proposal describes an comprehensive approach to testing SWDs and ELLs, embracing item development, the use of technology, training, administration, scoring and reporting.

PARCC recognizes that an essential component of calculating student growth for teachers, principals, and schools evaluation is a system for identifying and tracking students, teachers, and principals.

Recommendations

Decisions regarding high school tests should take into account the interests of non-college-bound students.

PARCC should consider setting preliminary CCR standards on the basis of field-testing. Depending on the quality of the field-test, the final standards would change little, if at all. As time is allowed to pass, political difficulties can arise in setting the standards.

PARCC should implement a centralized approach to determining growth for teacher and principal evaluation. Encouraging States to implement customized approaches will likely result in inconsistencies. The same concern applies to school accountability and CCR measures.

(A)(4) Assessment System Development

	Available	Score
(A)(4) Assessment System Development	35	31

(A)(4) Reviewer Comments:

Overall, the proposal provides evidence of a good plan for assessment system development.

(a)

The evidence centered design (ECD) and universal design (UD) approaches to developing items reflect industry standard practices. The proposed assessment development committees will possess the expertise needed to analyze the Common Core State Standards (CCSS), develop the test blueprint, and begin development of performance level descriptors (PLDs), prior to contracting for item and test development. The proposal clearly describes the committees and processes employed to assure content validity, fairness, and fidelity to UD principles.

The plans for pilot and field-testing, commendably based on National Assessment of Educational Progress (NAEP) sampling designs, allow for good studies of item bias and accommodations for students with

disabilities (SWDs) and English language learners (ELLs). The plan for early pilot-testing should give teachers a valuable "heads-up" regarding the future tests.

The proposal rightly foresees the difficulty in obtaining adequate participation for the field-tests. While it sets reasonable goals for participation, it is not fully persuasive as to how those goals can be met.

(b)

PARCC intends to develop an accommodations manual that member states must adopt. A preliminary policy document, based on existing practice and current research, is to be developed early so that it can be used to guide test development. PARCC wisely foresees the need to assist member states in the difficult tasks of making changes to existing law/regulation in this area and training Individualized Education Program (IEP) teams.

(c)

The proposal envisions a mix of human and AI scoring, each serving to check and balance the other. PARCC clearly lays out the issues related to consistency and comparability of human scoring and explicitly describes reasonable strategies (e.g., vendor training, standardized rubrics) to address them. An opportunity for all teachers to use online modules for unofficial scoring will provide useful professional development.

(d)

PARCC foresees the need for managing IDs and unique identifiers. However the proposal does not consider the extent to which changes in state law/regulation may be needed for states to participate in the data system.

The proposed data system would permit tracking of students/teachers/principals, cleaning data, and comprehensive reporting. However, this section of the proposal does not discuss plans for timely reporting. This topic was briefly addressed in (A)(3)(c), but more discussion would be appropriate here.

(e)

The proposal documented quality/control (Q/C) processes in the earlier sections of (A)(4). However, it does not address the issue of detecting and managing irregularities in test administration.

Recommendations

Inadequate participation in stand-alone field tests (versus embedded field-tests) is often a serious problem. PARCC should develop a plan (possibly including teacher/school incentives, SEA encouragement, etc.) to assure the needed participation.

Requirements and management of student/teacher identifiers are often set in state law/regulation. PARCC should consider earlier implementation of plans to assist states in making the needed changes.

PARCC should commit to a specific schedule for timely reporting to teachers and parents.

PARCC should develop policies and procedures for detecting and managing irregularities in test administration.

(A)(5) Research and Evaluation

	Available	Score
(A)(5) Research and Evaluation	30	28
(A)(5) Reviewer Comments:		
Overall, the proposed research and evaluation plan is excellent.		

PARCC's approach to research and evaluation frames a coherent research agenda in terms of a validity argument for the assessment system. An exhaustive set of studies is proposed which would inform the design of the system and which would evaluate the system.

Wisely, the implementation of the research plan is to be guided by recommendations of the Research Study Group and the Technical Advisory Group.

Recognizing the challenges posed by the innovative nature of the assessment, PARCC will commission a series of research papers at the beginning of the project on various topics, including item development, artificial intelligence scoring, teacher effectiveness and evaluation, text complexity, accessibility, standard setting, and vertical scaling. The research papers should provide valuable guidance regarding test development and further research needs.

Recommendations

The timing of some but not all of the proposed studies is dictated by the schedule for assessment design and development. PARCC should develop a comprehensive agenda for conducting and reporting the proposed studies.

(A)(6) Professional Capacity and Outreach

	Available	Score
(A)(6) Professional Capacity and Outreach	15	10

(A)(6) Reviewer Comments:

Overall, the proposal's approach to professional capacity and outreach is adequate, but could be improved.

(a)

PARCC proposes to build a leadership cadre in each state by involving a small group of teachers in the design and development process, which is good. However, it is not clear how this cadre is to carry out its activities.

PARCC laudably proposes to develop an online training and certification program to certify staff for administration of the test. However, it does not describe how it will obtain SEA and LEA cooperation in allocating teacher time (a scarce resource) to training.

The proposal wisely considers providing materials to encourage alignment of local curricula with the CCSS. However, successful alignment will likely depend more on the existence of suitable textbooks and states' adoption policies, which is beyond the scope of the grant.

(b)

The proposal describes an appropriately scaled, effective set of public relations campaigns that are focused on important constituencies: the public, parents, K-12, and IHEs. Specific issues have been identified (e.g., too much testing, the benefits of the CCR approach to testing), and talking points outlined. Opinion leaders and state and national leadership organizations would be enlisted. Outreach activities would be conducted in-phase with the development of the assessment system.

Recommendations

PARCC should develop a training of trainers plan that each state could use to expand professional development using its leadership cadre.

PARCC should consider that non-certified staff may be drafted by LEAs and schools to assist in test administration, particularly if large amounts of time are needed. Minimum qualifications for test administration (e.g. English speaking, education level) should be developed.

In a project of this scope, complexity, and duration it is impossible to foresee every possible public relations issue. PARCC should consider conducting periodic focus groups with teachers who will be working directly with the assessment, students, and their parents. Such focus groups will be more valuable once the assessment is operational in early years when large-scale administration uncovers inevitable unforeseen problems, but may provide important insights during pilot- and field-testing.

(A)(7) Technology Approach

	Available	Score
(A)(7) Technology Approach	10	9
<p>(A)(7) Reviewer Comments:</p> <p>Overall, the proposal's use of technology is excellent.</p> <p>(a)</p> <p>The proposal thoroughly described the types and uses of technology in the sections (A)(3) – (A)(6). The descriptions are briefly recapped in (A)(7).</p> <p>PARCC's commitment to open source and interoperable technology is commendable.</p> <p>The proposal's reliance on artificial intelligence scoring is essential to manage scoring costs.</p> <p>PARCC intends to integrate various existing separate data systems to improve overall management of test administration.</p> <p>(b)</p> <p>PARCC wisely foresees a potential barrier in the technological readiness of state and local education agencies (SEA/LEA) to operationalize the assessment system. Part of the proposed solution is to provide incentives for the technology marketplace to emphasize interoperability and open source standards. PARCC would also coordinate with various existing national projects that have a similar focus.</p> <p>Recommendations</p> <p>Despite good planning and best efforts, it may be overly optimistic to assume that all SEA/LEA partners will be technologically ready to implement the assessments by 2014-15. PARCC should conduct a study to identify which agencies are best equipped to implement the assessments for pilot- and field-testing. Lessons learned from these agencies could then be used to inform capacity building elsewhere.</p>		

(A)(8) Project Management

	Available	Score
(A)(8) Project Management	30	22
<p>(A)(8) Reviewer Comments:</p> <p>Overall, the proposal's management plan describes a feasible approach that will implement the system on-time and within budget. Long-term financial sustainability will depend on replacing most human scoring with computer-based scoring.</p> <p>(a)</p> <p>The project manager partner (Achieve) possesses the qualifications and experience required to implement the assessment system (e.g., the American Diploma Project (ADP) Network, the ADP Assessment Consortium, the College & Career-Ready Policy Institute, and the Common Core State Standards</p>		

Initiative). Achieve personnel currently on staff are suitably qualified and experienced. If the grant is awarded, Achieve intends to hire a new director and staff to manage the project. Because these individuals have not yet been hired their experience and qualifications cannot be evaluated.

(b)

Summary Table (A)(8)(b) displays a thoroughly detailed workplan and timeline that summarizes the work described in previous sections of the proposal. However, some of the more general tasks should be milestones in their own right. For example, under the technology platform milestone, a task is for schools to develop and implement specific plans to address technology barriers. It is unrealistic to expect that this task can be completed by 6/30/2012 as proposed. There are too many tasks listed under the milestones for higher education engagement and under public outreach.

The proposal indicates that there will be a number of committees with state and national experts to ensure quality, including the Passage and Media Review Committee, Test Item Review Committee, Bias and Sensitivity Review Committee, Data Review Committee, and the Committee on Accessibility and Accommodations.

(c)

The budget summary table indicates that 10.6% of requested funds would go for project management. This percentage may seem high, but is actually quite reasonable given the scope and complexity of the proposed assessment system.

The proposal assumes that the costs will be less than or equal to the per student costs of a majority of states' current systems. This assumption is not realistic for those states that decide to use human (versus computer) scoring.

The proposed budget for the fiscal agent (Florida) wisely notes that the state will need four full time equivalent (FTE) staff to fulfill its role. The support for governing states provides one FTE per state. This is realistic given the scope and complexity of the proposed assessment system.

The proposal prioritizes two level 2 budget modules: formative assessment tools for K-2 (priority 1); and technology enhancements (priority 2).

The proposal to develop K-2 formative assessment tools that are "developmentally appropriate" does not address two important developmental issues: (1) individually variable rates of developmental growth in young children; and (2) individual variability in social skills (shyness, confidence, etc.). Both factors make K-2 test scores hard to interpret. The costs are adequate and reasonable.

The proposal for technology enhancement wisely recognizes the risks of system overload during administration and AI scoring of the assessment. The risk is serious since an overloaded and broken-down system threatens the credibility of the project. The costs are adequate and reasonable.

(d)

The proposal correctly notes that the cost of administration in each state depends on the type of scoring (e.g., human versus artificial intelligence (AI)) the state selects. However, an average cost of \$32.68 per student is forecast for 2014-15, which is reasonable given the content of the assessment.

Recommendations

PARCC should provide additional detail regarding the milestones for "higher education engagement" and "public outreach."

PARCC should require the fiscal agent and governing states to assure that the FTEs budgeted are in fact appropriately assigned to the project.

PARCC should conduct a more detailed study of the sustainability of the system after the project ends. Given the high costs of human scoring, it may be important to promote the acceptability of AI scoring.

If the level 2 budget proposal to develop K-2 formative assessment tools is funded, it should include test interpretive materials that alert teachers and parents to limitations in the meaning of the scores.

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:

The letters of intent signed by the heads of IHEs and IHE systems meet the requirements of this priority. Although the letters themselves do not provide numbers of direct matriculation students, the required numbers were provided in the proposal.

PARCC's strategy for engaging higher education commendably includes establishing advisory committees, including a College-Ready Advisory Committee and College-Ready Work Groups. PARCC will have a Director of Postsecondary Engagement to work with the committees and state liaisons to develop and strengthen relationships with IHEs.

The proposal indicates commitments from 188 public institutions representing 90% of direct matriculation students in public IHEs in all membership states.

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:

This is an extremely ambitious project, considering the heavy emphasis on performance tasks and constructed response test questions, together with the innovative approach to computer-based administration and scoring. However, the overall clarity of the proposal, the strength of the governance plan, the demonstrated capacity of the project management partner, the clarity of the assessment system's design, the strong plan for development, the exceptional use of technology, and the thoroughness of the plan for research and evaluation, all strongly suggest that it can succeed. The proposal specifically addresses all of the requirements for this program.

Grand Total	220	194
--------------------	-----	-----

Budgets

Level 1 Budget
Name: Level 1 Budget(s)
The proposed budget for content development assumes that 1.85 items must be developed for each operational selected response item and that four performance tasks must be developed for each operational

task. This errs on the conservative side. However, if more items survive than assumed they could be used later in operational tests.

Level 2 Budgets

Name: Formative Assessment Tools for K-2

The proposal to develop K-2 formative assessment tools that are "developmentally appropriate" does not address two important developmental issues: (1) the prevalence of individually variable rates of developmental growth in young children; and (2) individual variability in social skills (shyness, confidence, etc.). Both factors make K-2 test scores hard to interpret.

If the proposal to develop K-2 formative assessment tools is funded, it should include test interpretation materials that alert teachers and parents to limitations in the meaning of the scores.

Name: Technology Enhancements

The proposal for technology enhancement wisely recognizes the risks of system overload during administration and AI scoring of the assessment. The risk is serious since an overloaded and broken-down system will threaten the credibility of the project.