STATE OF OHIO
DEPARTMENT OF ADMINISTRATIVE SERVICES
GENERAL SERVICES DIVISION
OFFICE OF PROCUREMENT SERVICES
4200 SURFACE ROAD, COLUMBUS, OH 43228-1395

MANDATORY USE CONTRACT FOR**:** The Ohio Performance Assessment Pilot Program (OPAPP) – Race to the Top (RttT)

CONTRACT NUMBER: CSP904911 EFFECTIVE DATES: August 12, 2011TO August 31, 2014

The Department of Administrative Services has accepted Proposals submitted in response to Request for Proposal (RFP) No. CSP904911 that opened on June 17, 2011. The evaluation of the Proposal responses has been completed. The Offeror listed herein has been determined to be the highest ranking Offeror and has been awarded a Contract for the services listed. The respective Proposal response including, Contract Terms & Conditions, any Proposal amendment, special Contract Terms & Conditions, specifications, pricing schedules and any attachments incorporated by reference and accepted by DAS become a part of this Services Contract.

The agency listed herein is eligible to make purchases of the contracted services in any amount and at any time as determined by the agency. The State makes no representation or guarantee that department will purchase the volume of services as advertised in the Request for Proposal.

This Requirements Contract is effective beginning and ending on the dates noted above unless, prior to the expiration date, the Contract is renewed, terminated, or cancelled in accordance with the Contract Terms and Conditions.

This Requirements Contract is available to the Ohio Department of Education (ODE) as applicable.

Questions regarding this and/or the Services Contract may be directed to:

Dana King, CPPB
dana.king@das.state.oh.us

This Requirements Contract and any Amendments thereto are available from the DAS Web site at the following address:

http://www.ohio.gov/procure

1. CONTRACT SUMMARY. The objective of this Pilot Program is to develop and pilot test a learning and assessment task assessment system (as defined in the Ohio performance task dyad system) for elementary and high school students that align to the Common Core Standards for ELA and Mathematics as well as Ohio's New Revised State Standards in Science, Social Studies and Career Technical Pathways. The intended outcome of the Pilot is to end up with a suite of field-tested performance tasks that could be used in classrooms for both formative and summative assessment purposes and a system of development, scoring, and delivery of both the tasks and the associated professional development that is sustained by a source independent of the State.

   In summary, the Contractor shall:

   a. Develop and validate, with the State, a task specifications document that includes a format for the tasks as well as parameters for the required parts of the tasks.

   b. Develop and implement, with the State, external review committees to be used for task development (e.g., a Fairness Committee and a Content Committee).

   c. Develop, co-develop, and submit for review Learning and Assessment tasks aligned to the Common Core and New State Standards in ELA, Mathematics, Science, and Social Studies. The review process shall include an internal review where revisions can be made as well as an external review where revisions can be made. The Contractor shall make the revisions that ODE approves.

   d. Identify a platform that will allow for electronic delivery and retrieval of both the learning and assessment tasks.

   e. Develop and implement a central moderation panel and several regional moderation panels.

   f. Develop and implement a reporting system for reporting student scores on the assessment tasks.

   g. Develop, in collaboration with the State, a program for involving experts from institutions of higher education (both in specific content areas and in education/pedagogy).

   h. Develop and provide professional development for teachers on how to use the learning tasks and generalized rubrics to guide feedback that teachers will give to students.

   i. Develop and provide professional development for teachers on how to score the assessment tasks using task-specific rubrics.

   j. Develop and provide professional development for teachers on how to construct performance tasks.

   k. Develop, in collaboration with the State, a program for coaches that includes their roles and responsibilities for the Pilot.

   l. Develop a plan for, and a schedule of participant (teachers, coaches, and experts) activities, each semester.

   m. Conduct an evaluation of psychometric considerations.

   n. Collaborate with a business partner, identified by the State, who will provide sustained support for the system developed.

   o. Prepare, provide, present and/or submit annual USDOE, ODE, and DAS project reporting requirements.

   p. Ensure the Project is implemented, executed, and transitioned successfully.

2. The general Terms and Conditions for the Contract are contained in Attachment Three of the RFP for Project. The Contract consists of:

   a. The original RFP and any addendums.
   b. The documents and materials incorporated by reference in the RFP.
   c. The Contactors' Proposals, as amended, clarified, and accepted by the State.
   d. The documents and materials incorporated by reference in the respective Offerors' Proposal and subsequent acepted clarifications.
   e. Any related amendments issued subsequent to Contract award.

3. The ODE and the Contractor shall notify the DAS, Office of Procurement Services within ten (10) business days in the event of a change in personnel, financial, or contact information.

4. The term of the Contract will be from the award date through August 31, 2014. The State may solely renew this Contract at the discretion of DAS for a period of one (1) month. Any further renewals will be by mutual agreement between the Contractor and DAS for any number of times and for any period of time. The cumulative time of all mutual renewals may not exceed one (1) year and are subject to and contingent upon the discretionary decision of the Ohio General Assembly to appropriate funds for this Contract in each new biennium. DAS may renew all or part of this Contract subject to the satisfactory performance of the Contractor and the needs of the Agency.

MUTUALLY AGREED UPON CLARIFICATIONS AND MODIFICATIONS. (To include the finalization of the terms agreed upon during negotiations, which concluded on August 10, 2011).

1. Timeline flexibility due to limited ODE resources. The Contractor agrees that:

   The dates listed on the Gantt charts provided to the State on July 6, 2011 along with its responses to clarifications are merely placeholders. The intent of both parties is to finalize activity dates at the initial planning meeting with the State's input and sign-off. The Contractor understands that dates may need to change to accommodate key stakeholders' schedules and that it will need to be flexible. For example, line 56 of the Gantt chart, the task that was estimated to take .5 days of an ODE resource could take place anytime during the span of 5 days shown in the schedule.

2. Project Management Meetings. The Contractor agrees that the project management meetings (in each stage of the Project) include updates to the Work Plan, review of the Project Gantt chart, calendars, and schedules by the State and the Contractor. All scheduling of contract activities will take place with the State's input.

3. ODE's limited facilities for meetings. The Contractor agrees to accommodate up to as many as 20 participants in each of the "kickoff" and/or other periodic leadership and management meetings. The Contractor views participation by as many ODE staff members as possible as a benefit to the Pilot. Therefore, the Contractor suggests that these meetings be held either at the State Library of Ohio or at the State Capitol. These facilities offer meeting rooms at free or reduced cost to state agencies. The Contractor's program management team shall arrange for meeting space as needed, at no additional charge to the State. Alternative meeting locations, such as at The Ohio State University or at a Columbus School District auditorium, are other possible options. Additionally, the Contractor is able to offer the State the option of providing video conferencing for remote participants or if none of these options are acceptable.

4. Site Selection. The Contractor agrees that the State shall have the right of review of sites selected and obtained by the Contractor or Battelle.

5. Technical Advisory Committee. The Contractor acknowledges that the State does not feel the need for a performance assessment specialist to serve on the technical advisory committee as the expertise already exists. The State does not consider it appropriate for the Contractor to identify members of the technical advisory committee. The State requires that the interpretation of Performance Assessments be within the conceptual learning/instructional framework as well as the assessment embodiment of the ODE and may differ considerably from past embodiments practiced or known to any third party.

6. Item Development beyond Content Standards. The Contractor acknowledges that at this time there is no intent to devise items and operational methods for Alternate Assessment of students with severe cognitive disabilities. College- and Career-Readiness shall be embodied in this work product through alignment to established academic content standards. There may be a separate analysis in the context of College- and Career-Readiness, but it is not anticipated that there will be items developed beyond the content standards (or Career Tech Competencies) to assess College- and Career-Readiness. The State does plan to use the OPAPP in the context of formative instrumentation, summative instrumentation, and for purposes of developing improved instructional practices.

7. The state of Ohio approach to Performance Assessments. While The Contractor has considerable experience with performance assessment and the State may gratefully accept some input from The Contractor based on the Contractor's past experience and expertise. The development of items as either learning tasks or as assessment tasks, as well as development of scoring rubrics and scorer training shall conform to the approach prescribed by ODE.

   For example, while teachers may learn how to develop their own performance assessment tasks and scoring rubrics for their own idiosyncratic tasks, the primary thrust of this Project is not one of independent teacher initiated developments but one of uniform tasks and rubrics that ensure opportunities to learn (through proper alignment) and provide results that are shown to be on an invariant scale. In other words, the State strives to provide for sufficiency of instruction and fair testing, as a minimum.

   Additionally, disclosure of developments of this Project is restricted in use per state laws and this Contract. Press releases are also subject to state law and this Contract.

8. Teamwork. The performance of the Contract may or may not involve democratic processes depending on the situation. While the State realizes the importance of a collaborative approach in the development process, there may be times or specific state goals that may not allow for a wholly democratic process.

9. Nimble Innovation Lab. The da Vinci Awards® annually recognizes the most innovative adaptive and assistive technologies that enable equal access and opportunity for all people, regardless of ability. In 2008, Nimble Assessment Systems, Inc. received a da Vinci award for NimbleTools, a universally designed test delivery interface that increases access to test content for students with disabilities and special needs. NimbleTools employs a single test delivery interface that flexibly tailors the accessibility tools that are available for each student based on his or her individual need. NimbleTools ensures access for students with physical disabilities who use any one of a variety of assistive/adaptive communication devices (e.g., switch mechanisms, sip-and-puff devices, Intellikeys, alternate mice, touch screens, etc.). NimbleTools also ensures that the reading and signing of content is provided in a high-quality, equitable manner to students across a testing program.

10. Federal Requirements for Accessibility. NimbleTools exceeds federal requirements for accessibility, per Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. 794d) as amended by the Workforce Investment Act of 1998 (P.L. 105 - 220), August 7, 1998. Section 508 Standards are available at this government Web site: http://www.section508.gov/index.cfm?fuseAction=stds

11. Scoring Expertise.

    a. Levels of inter-rater reliability. The State's expectations are levels of inter-rater reliability between 60%-80% exact agreement, depending on the task, teacher/scorer population, double-blind rate, etc. The Contractor recommends at project outset a standard inter-rater reliability agreement rate of eighty percent (80%) exact agreement in terms of accuracy on the double-blind, embedded check set, and/or third-party read behind procedures. If, however, at some point during work on the Contract, the State prefers applying an alternate rate of inter-rater reliability agreement (e.g., 70% exact agreement), the adjusted agreement rate will be instituted and adjusted without difficulty. The scoring of performance tasks in a similar fashion to the Contractor's experience with scoring of some extended response prompts (i.e., writing) are scored. In this case, the Contractor's expectation of scorer accuracy for the scoring of extended response prompts begins around 70% exact agreement.

    b. Remedies for inadequate inter-rater reliability. As described in the Contractor's original proposal, (page 124, paragraph 3), the TaskStream system shall flag for the scoring moderator any reader whose accuracy rate falls below the specified agreement rate (e.g., 80% exact agreement). If a reader's work or overall scoring in general falls below the minimum ODE-established inter-rater reliability agreement rate, the scoring moderator will have the opportunity to answer a reader's questions, provide the reader with additional coaching, feedback, or retraining, or choose to stop that reader's scoring activities altogether. This interaction between reader and moderator will create not only a means of corrective measure toward maintaining inter-rater agreement standards, but also create added value in the form of peer-to-peer professional development training.

       A final available remedy for any reader who falls below either the minimum inter-rater reliability agreement rate or another ODE-established alternative rate is the removal altogether of that reader's work from the scoring system. If a scoring moderator decides to remove a reader's data based on less-than-minimum established inter-rater reliability agreement performance, the TaskStream system will allow the moderator to remove all of that individual's collective scores and redistribute those student responses to other readers for re-scoring.

12. Psychometric Data. The Contractor may use proprietary software for tasks such as cleaning data and concatenating files. Software for computation of psychometric properties shall not be proprietary but publically accessible, shall have full documentation and support, and shall be approved by the State. In general, this is commercially available software that is commonly used within the industry for these purposes. The Contractor shall not utilize proprietary software to conduct any item response theory (IRT) analyses. All item response theory analyses shall be conducted using commercially available software.

13. Staffing Plan. The Contractor understands the scope of development work required, especially in the first 18 months of the Contract, and has planned for this workload accordingly. The agreed upon plan proposes contract leads for each content area that will have the support of additional staff during the item writing and development process. The Contractor's contract leads shall manage the work of the additional staff and ensure the quality of our deliverables to the state of Ohio.

    Support shall be organized into two (2) roles:

    a. Item writer; and

    b. Curriculum & Assessment (C&A) Specialist.

The Contractor's item development process, and the tasks to be performed by these two (2) roles, is described as follows:

a. Item Writing—Freelance Item Writers. The Contractor maintains relationships with a number of individual freelance item writers as well as organizations that provide item writing services (item writing houses). Freelancers shall have undergone a rigorous selection process and have receive training on the Contractor's expectations for quality and timeliness in completing assignments. The training shall include a thorough understanding of the criteria with which the Contractor C&A Specialists evaluate items.

   1) Content—Does the item writer show in-depth understanding of the content area?

   2) Alignment—Are the items properly aligned with the standards?

   3) Curriculum—Are the items appropriate to the grade level?

   4) Creativity—Does the writer show the ability to write items that use a variety of approaches and that measure higher cognitive levels?

   5) Mechanics—How well do the items meet the many psychometric conventions such as parallelism of options?

   6) Fluency—How well written are the items?

   At the start of a round of item development, contract lead C&A Specialists meet with the Item Writer Coordinator to determine the scope of freelance work and to check on the availability of freelance item writers. When freelance item writers are contacted, the Item Writer Coordinator provides them with details of the contract such as the distribution of items, the delivery date, the client's curriculum standards, and the item specifications. Item specifications are particularly important when dealing with freelance item writers because they provide examples of grade level appropriate items, sample scoring guides for constructed response items and performance tasks, and interpretation of the curriculum standards that the freelance item writer needs to understand in order to provide items that meet the expectations of our clients.

   Freelance item writers shall deliver their work to the Contractor via TaskStream. The item writers shall e-mail the Item Writer Coordinator to indicate that items have been posted. The Item Writer Coordinator shall be responsible for ensuring that new items are delivered to C&A Specialists in a timely and secure fashion.

   In addition to the text of the item, freelance item writers are expected to provide additional information where appropriate.

   1) Scoring Rubrics – Where appropriate, freelance item writers shall be expected to develop scoring rubrics for individual items.

   2) Graphics and Media – While freelance item writers will not expected to generate graphics or other media, they will be expected to describe what they expect the graphic or media object to look like.

   3) Source Material – In all instances where there are verifiable facts involved with the questions, the item writer shall submit copies of reference materials or hyperlinks to on-line sources.

b. Internal Review – C&A Specialists. The Contractor C&A Specialists shall produce quality items and tests that meet the needs of their clients. The Contractor's freelance item writers shall continue to be thoroughly trained and given ongoing feedback. It is the C&A specialists who shall have extensive knowledge of the intricacies of a state contract that can be developed only through personal contact with members of the State and the content advisory committees. When the Contractor receives freelance items, the items shall undergo an extensive content review by the Ohio contract lead C&A Specialist and other Specialists assigned to the Contract.

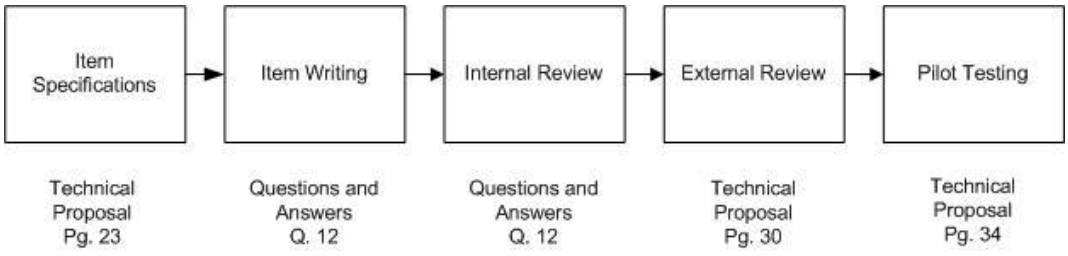
Items shall then be reviewed, edited, and, if necessary, replaced to ensure a consistent level of high quality. C&A specialists review each set of items to ensure that it meets the requirements described in Table 1 below.

TABLE 1 – ITEM REVIEW REQUIREMENTS

REQUIREMENT	EXPLANATION
Alignment to standards	Every item shall be checked for alignment to standards as called for in the specifications. In addition, the item set shall be evaluated to ensure that the item set as a whole adequately and properly samples the domain of the standard.
Range of difficulty	Each set of items shall be reviewed for appropriate range of difficulty to ensure adequate measurement of students along the ability scale.
Range of cognitive levels	Cognitive levels shall be checked to ensure the set meets the requirements for Depth of Knowledge or other cognitive measure as specified by the State.
Item correctness	All items shall be reviewed by the C&A specialists to assure that every item has one (1) and only one (1) correct answer. Attention shall be given to assure that a reasonable misread of the question will not lead to an incorrect answer. Correctness of items shall be considered an important part of every internal review.
Accuracy of information	When real-world contexts are used in an item, the given information shall be checked for reasonableness. The Contractor shall require that reliable sources be provided for all real-world data.
Language	C&A specialists shall continue to be aware of the great diversity in the current school population. A major part of the initial content edit and all levels of content review shall involve the structuring of items for maximum clarity and brevity. Various vocabulary lists shall be employed to ensure that non-technical words are at or below grade level.
Bias and sensitivity concerns	Items shall be screened to avoid racial and cultural bias. Every effort shall be made to avoid using contexts for items that might be sensitive to a student and thus comprise the measurement of the student's understanding.

c. Item Development Process. The diagram in Figure 1 describes, at a high level, the item development process for the OPAPP. This diagram's purpose is to put the above information into context with the other aspects of the item development process discussed in the Contractor's Technical proposal and agreed upon negotiated technical components.

FIGURE 1: OPAPP ITEM DEVELOPMENT PROCESS



14. Work Plan - Performance Task Medium and on-line delivery.

    With regard to the Table 4 exemplar in the Contractor's Technical Proposal: It is not the intent to create portfolios of student work to demonstrate acquired skills as part of the assessment task. Instead, there may be (probably will be, for all content areas) a constructed response that is narrowly responsive to some explicit skill and knowledge the student should have been taught, the student should have practiced, and the student should have received feedback on the practice during the learning task.

    The Contractor agrees that the assessment task shall align narrowly to the learning task – so narrowly that the student should have practiced the assessed skill during the learning task. It is also imperative that the assessment task be clearly aligned to the content standards. It is understood that the State will be vigorous at insisting that the expectation for learning be narrow.

    The Contractor openly welcomes dialogue to increase its understanding of the the State's needs for ELA dyad tasks and to fine tune its approach to meeting those needs.

    Sample dyad tasks shall be created in such a way that as the client, the State will have the ability to choose the portions it wishes to keep, and the Contractor will be responsive to any changes the State requests for the remaining elements. The Contractor shall ensure collaborative efforts with the State to rework the dyad in any way that that will assist the State in meeting the objectives of the OPAPP. Both parties understand and agree that the working relationship between the State and the Contractor staff is essential to creating the learning and assessment system that is desired for this endeavor.

    The Contractor's understanding of the Common Core Standards for Mathematical Practice is that they should infuse the entire curriculum rather than be taught directly. The Contractor also agrees with the State that it is difficult to assess these practices. Thus, the Contractor understands the State may wish to refrain from citing an alignment of performance tasks to the CCSS mathematical practices.

15. Learning task and goals. The Learning tasks shall be explicitly aligned to the content and should have an explicit learning goal that can be specified in the context of what the student should be able to do after completing the learning task.

16. Task dyad system. While the Project may have or perhaps still conform to the "Common Assessment Consortia definitions", the relevant definition is that of the ODE and the state of Ohio.

    The Contractor understands that the operational definition of a performance task dyad system for this Project will be developed by the ODE. The Contractor shall work with the State, through an iterative and collaborative process, to arrive at a common understanding of an operational definition for the performance task dyad system. The Contractor shall then document that definition in its item specifications and prepare training materials regarding performance task dyads for freelance item writers and cohort members.

17. Task Development Timeline.

    a. The Contractor understands that the teachers and maybe the students may not be a static cohort across years. E.g., students in the pilot with a given teacher will not be the same for year 1 as in year 2).

    b. The duration of the assessment task should be maintained and, preferably, be compartmentalized to guard against security threats when activity extends across multiple testing sessions.

    c. The second bullet on page 34 ("Some content….") should be interpreted in the context that the assessment task is a sample or subset of the content domain and of the learning task domain.

    d. There is, at this time, nothing called an "extended task".

18. Technology Platform. TaskStream operates on a Software as a Service (SaaS) model; the software cannot be installed on local machines to limit access to designated computers. If the State does not want students to have the ability to access TaskStream unsupervised, the State could consider providing student usernames and passwords to teachers or other appropriate school officials, instead of providing this information directly to the students. Alternatively, teachers could manually turn on/off the visibility of tasks for their students to restrict student access to review information and provide their responses during set times in the day. Should these options not meet the State's needs, shall work with the State to understand the need more clearly in order to develop a more appropriate and/or desirable solution.

19. End-user products. Users (e.g., students) shall be able to access the TaskStream site using mobile browsers; however, the site is not currently optimized for this use and TaskStream does not yet offer a special app for such devices that might build in additional security measures. However, TaskStream has a group dedicated to Integration Services and Mobile Platforms that is responsible for developing new features and enhancements for mobile platforms. This group is currently focusing on optimizing the use of the site by evaluators (e.g., teachers) using Safari on the iPad, and is continually prioritizing the list of desired features and functionality to add in the future.

20. Task Delivery, Retrieval and Storage System. All assessment tasks shall be individual student work.

21. Task Scoring. The Contractor understands the State's expectation that student work may be electronically redistributed for multiple scoring using a variety of decision rules. TaskStream shall continue to have the capability to maintain separate scoring records for each rater. The system does not currently support a variety of automated decision rules for redistributing student work for additional scoring; however, it is designed to support up to 25 evaluations of the same work submission.

22. Reporting and Data Analysis. The State expects data provided from the TaskStream system be provided in tables of records (one [1] for each rater and student work combination) that can be read as either a fixed field flat file or as a .csv file.

    TaskStream provides three (3) options for exporting data records from the system:

    a. All reports generated in TaskStream can be exported to Excel (or .csv), including performance reports that display scores along with student and rater information.

    b. TaskStream offers an optional bulk export capability for exporting assessment data from TaskStream assessment programs in XML format with associated data schema. In addition to evaluation results (i.e., scores), the export includes student data (e.g., name, student IDs), history logs of all submission and evaluation events, rubric scores for all rubric criteria, and descriptions of student work (uploaded artifacts are not included).

    c. Via Web services, TaskStream can provide real-time data on evaluation events occurring in the system (e.g., scores when an evaluation has been completed).

23. Scoring.

    a. Learning Task Scoring. Details for learning task scoring shall be resolved through the mutual agreement of the State and the Contractor subsequent to Contract inception. Topics for discussion include criterion levels for decision rules and the computation of indices of rater agreement.

    b. Assessment Task Scoring. TaskStream includes the ability to manually group students with specific scorers, which can, in effect, lead to the same result as assigning individual work submissions to specific scorers, depending on how the system is configured. The system does not currently offer the ability to automatically limit or view statistics regarding the number of student work samples assigned to specific scorers.

    c. Mechanism for determining when scorers are expected to reliably produce consistent scores. There is no mechanism in the system for determining when scorers are expected to reliably produce consistent scores. TaskStream does offer functionality to facilitate scorer training, including the ability for multiple scorers to evaluate the same benchmark sample and the ability to generate reports to examine inter-rater reliability. The determination of when scorers are expected to reliably produce consistent scores is left to human judgment, not a calculation in the system.

24. The TaskStream Platform, Data Transmission, etc. The expected types and amount of data that will be transmitted to and from MP/TS on a regular basis. Internet bandwidth pales in comparison to the gigabit Ethernet links commonly found in enterprise LANs, and data transmissions that take a few minutes to transfer between servers in a school's server room might take hours to transmit to and from MP/TS, which is located across the country.

    TaskStream was designed to be bandwidth efficient. The bandwidth for normal use is proportional to the number of subscribers. Security Network filtering devices may also have an impact on performance; however, if they are efficiently configured, the user experience should be excellent with only nominal bandwidth requirements.

    Because the TaskStream software is Web-based, there is no need to transmit large data files between the ODE, the Contractor, and TaskStream on a regular basis. The State can generate reports in the system in real-time. If the ODE wishes to establish a data link between TaskStream and the ODE for transmission of evaluation data—such as submission and evaluation information—the amount of bandwidth required is not substantial. However, if TaskStream is required to retransmit every student artifact to the ODE, then the bandwidth requirements necessarily go up substantially.

25. Technical Support. TaskStream's technical team can support the ODE in assessing bandwidth needs and can provide strategic consulting regarding ongoing infrastructure plans.

26. Benchmarking – Learning Tasks. The Contractor understands the State's expectation that teachers be trained to score one or more learning tasks by the Contractor after the implementation process.

    For assessment tasks, the State expects that the benchmarking for these tasks will be done by the Contractor, but with the assistance of input from the Range Finding Committee. Following a meeting of the Range Finding committee, the State will approve training sets, anchor papers and validity sets proposed by the Contractor. Training sets and anchor papers will be used to train teachers to score the assessment tasks. (This would be done in a face-to-face training session.) Then, teachers would score papers that would be assigned to them along with a set of validity papers (that would not be identified as validity papers to the teacher).

27. The State does not expect that the Contractor will have a person on the ground for classroom video. The State is willing to work closely with the videographer hired (as accepted by the State) to review video for use in PD modules and for other purposes.

    The Contractor shall work closely with the State to create and fine-tune module deliverables' dates and milestones for completing the associated products. The Contractor suggests the first round of modules represent the face-to-face training sessions. This footage would be gathered in the early phases of the Project, reviewed by the State and the Contractor, edited accordingly, and formatted for the appropriate delivery mode(s). Capturing the live training sessions provides modules for scaling and sustaining the project after the pilot year.

    The Contractor suggests that classroom footage be gathered after teachers have had ample opportunity to implement and fine-tune strategies utilizing the actual learning tasks. This could be in spring 2012 for Cohort 2 and fall 2012 for Cohort 3. This timeframe will provide the State the opportunity to identify, with the assistance of on-site coaches, which strategies to capture and which teachers can provide the highest quality examples of practice to film. For scaling and sustaining the program after the pilot year, providing high-quality exemplars of practice to utilize within a PLC structure will give teachers an adult learning program to reflect on current practices and discuss what changes may be needed to improve to fit the exemplars. Footage that may reveal less than ideal practice could be set aside and possibly used to train leaders for evaluation purposes.

    During the interim time between days 1 & 2 and 3 & 4, teachers could practice the provided strategies, log their experiences, and bring examples to the following training days. During the subsequent training day, teachers could be guided through collegial discussions regarding their experiences and findings modeling a PLC protocol format.

    The Contractor shall work with the State on the format for the exemplar modules, associated materials to support viewing, and editing for delivery.

28. Professional Development. ODE has secured locations for the professional development for the fall semester. The cost of securing these locations can be taken out of the remainder of the budgeted amount contained in this Contract, rather than the budget submitted for this proposal. All meetings have been scheduled to occur at the Quest Business center in Polaris. The participating LEAs and coaches have already been notified of these meeting dates. The dates and times for the meetings scheduled are as follows:

Meeting 1 – Quest Center			
Cohort 2: Sep 12-13 Math	Sep 15-16 Science	Sep 19-20 Social Studies	Sep 22-23 ELA
Cohort 1: Sep 14 ELA	Oct 11 Science	Oct 19 Math	
Meeting 2 – Quest Center			
Cohort 2: Oct 3-4 Math	Oct 6-7 Science	Oct 17-18 Social Studies	Oct 20-21 ELA
Cohort 1: Dec 14 ELA			
Meeting 3 – Quest Center			
Cohort 2: Oct 31- Nov 1 Math	Nov 3-4 Science	Nov 9-10 Social Studies	Nov 14-15 ELA
Meeting 4 – Quest Center			
Cohort 2: Nov 28-29 Math	Dec 1-2 Science	Dec 8-9 Social Studies	Dec 12-13 ELA

29. Benchmarking. The establishment of a range finding committee means that this committee will review student work and score the field tested task samples. The actual development of training sets is to be up to the vendor to do, in all cases.

30. PD Overview. The entire cohort of teachers will be receiving the professional development. While the entire cohort does not always include all teachers at a school or LEA, it does in some cases.

    The Contractor proposes using "technology as a means to enhance and support all participants". An example is the use of on-line tutorials covering specific topics. The State agrees to the use of on-line tutorials and fully supports this idea. The State has specific ideas for how these on-line tutorials should be created. The State proposes the tutorials to be based on the professional development provided at the face-to-face sessions and based on real teacher work. Therefore, the State's expectation is that these modules be created by the Contractor employing an Ohio-based videographer to record the professional development sessions and to visit classrooms to record teachers and students implementing both formative teaching techniques and performance tasks (learning and assessment types). The State believes this would enhance the program beyond the typical PowerPoint WebEx webinars which only promote the "stand and deliver" type of professional development. It is the goal of the State to provide professional development that directly supports the Pilot, and as such, the uses of materials created for this Project are the only materials that will suffice for any on-line tutorials.

    Further, the inclusion of a videographer in the pilot allows for a more in-depth analysis of the program. Additionally, the work of the videographer could be more "action research" so that the material recorded could be use both to support the professional development and to supplement the information that the external evaluator gets.

    The Contractor acknowledges that the State is willing to work closely with a videographer for the professional development modules.

31. Overview of Initial Training Session for Cohort 1.

    a. Only the ELA participants in Cohort 1 will do any operational scoring. The other two (2) content areas will be engaged in deepening their understanding of the dyad system and refining and documenting their implementation techniques. They may also be engaged in developing "homegrown" learning tasks; as part of their professional development, this could be supported.

    b. The number of participants in Cohort 1 is now 17 for Math, 20 for ELA and 25 for Science. Each subject area shall be treated differently. Cohort 1 has already been trained on how to score assessment tasks, with the exception of the ELA teachers. Only they need to have training on this and that can occur in their second PD session for the first stage. The idea of including professional development on the use of the on-line system for Cohort 1 is consistent with the State's goals for the PD.

    c. Cohort 1 is already trained in scoring. The scoring training that will need to occur is for the ELA group in Cohort 1 and they will need to be trained on how to score the assessment task, not how to score the learning task, as they are already trained to do that.

32. Cohorts 2 and 3, Stages 1 and 2. Eight (8) days of professional development for Cohorts 2 and 3 will be conducted.

    a. The proposed plan for Days 1 and 2 shall be more focused on instructional strategies and less focused on learning tasks. The teachers will not be familiar with the learning tasks yet and the focus shall be on the more generic use of strategies that will be important to use during the implementation of learning tasks without emphasizing the learning tasks themselves, since they won't have any concrete examples to work with. The State proposes using a pre-survey to identify the interests and needs of the participants first, then to develop a two-day session that informs teachers, models the practices and gives the teachers an opportunity to try out some of the techniques taught.

       Following this first session (days 1 and 2) the videographer will record teachers trying the techniques taught in actual classrooms. The State and the Contractor PD team will work with the videographer to identify clips that could be used in the second session of PD. Another pre-survey could be used to identify new needs and interests of the participants. Days 3 and 4 would then focus on these new needs and re-cap what was done before, using participant video to highlight good/bad techniques and how they might affect student learning. To deepen the understanding of formative methods, ODE would like to emphasize how teachers and students can use the information gained from a formative assessment method to further student learning. The focus of this session would be on "completing the learning cycle" and re-engaging students in learning material they clearly did not learn the first time. Again, the videographer would be instructed to go and record actual classrooms and the footage would be reviewed by ODE and The Contractor for use in Session 3.

b. Session 3 (days 5 and 6) is where ODE would like the use of rubrics to be introduced. At this point in the development cycle, the State should have a generalized rubric developed for each content area. The focus of the third session should be on interpreting a rubric, using a rubric to guide feedback to students and teaching students to use rubrics as a guide to their learning. This is an extension of Session 2 in that "completing the cycle" can only be accomplished if the student and teacher are aware of the need and know how to move the student from not knowing or not being able to do to knowing or being able to do. As a part of the professional development at this stage, it might be worth having the videographer record one classroom visit per teacher by the end of the semester so that each teacher could analyze their own classroom video. This could be a very important part of the professional development that teachers receive at this time.

c. Session 4 (days 7 and 8) would be an introduction to learning tasks and a framing of the learning-assessment dyad system for participants in Cohort 2 or 3. By the time this PD session is occurring, there will be some idea by the Contractor of what makes a good task (we should be completing internal task review and holding committee reviews). The PD would center on the dyad system as a learning and assessment tool as well as the task specifications document. This PD serves a dual purpose: First, it introduces teachers to the system they will be implementing in the coming semester and second it starts them down the path of understanding the components of a learning task for the future when they will design tasks themselves. This PD should incorporate all of formative techniques used in Sessions 1-3 as required components for the learning task.

d. For Stage 3 for Cohort 2 (Stage 4 for Cohort 3), the PD shall consist of a one-day session early in the semester (pre-implementation), a one-day session in the middle of the semester (post-implementation, pre-assessment), and a two-day session at the end of the semester (for scoring the assessment tasks). ODE would accept one other arrangement of the PD sessions: a one-day pre-implementation session, a two-day post-implementation/pre-assessment session, and a one-day scoring session. Whatever pattern of professional development is chosen for stage 3 for Cohort 2 will be the pattern of PD for the remainder of the Project.

33. Coaches and Their Role. Both parties agree this Pilot is intended to be transformational and the imposition of a new structure in districts and schools is expected. The State will be asking schools and districts to do things that are outside of their normal modes of operation.

34. Participants. On p. 73, the Contractor mentions again the creation of on-line professional development modules. The State supports this model and requests that the modules be created by video-recording the live face-to-face professional development meetings as well as classrooms of teachers and students participating in this pilot. The State shall direct the creation of the modules with the advice of the Contractor and the Ohio-based videographer. The Ohio-based videographer shall obtain all necessary permissions for video-recording students and teachers for both the purpose of creating the modules as well as for the purpose of evaluating the Project.

35. The Contractor will seek an Ohio-based videographer, to be approved by the State, and will require that videographer to obtain all necessary permissions prior to recording.

36. Validity. p. 74: ODE is particularly amenable to Contractor's offer for performing the "…investigation of student cognitive methods using think-aloud protocols." The proposed addition of a videographer will allow for such a study to be founded in observational evidence.

    Scoring protocol/plan: for a given teacher, there will be "n" student samples. Teachers will score 2*n + 10 student samples. Of the 2*n student samples, none will be student work from the teacher's own students. The ten (10) student samples will be validity student samples that have been mutually agreed upon by the State and the Contractor. Teacher ratings of the ten (10) validity student samples will be used to monitor teacher scoring reliability. A teacher with low reliability on the ten (10) papers will be required to go through training again. The student work that the teacher with low reliability scored will be scored by the moderation panel and the teacher's scores will not be included in the calculation of the students' scores.

37. Inter-rater Reliability. The Contractor acknowledge the State's clarification about the commitment of random selection(s) of readers, and the potential research-based reasons for the selection and placement of particular student documents in the hands of particular readers. The intent to place the selection and control of responses to be scored by particular readers in the hands of the scoring moderators or other state staff responsible for monitoring and directing such workflows using the TaskStream system. As described previously, TaskStream includes the ability to manually group students with specific scorers, which can, in effect, lead to the same result as assigning individual work submissions to specific scorers, depending on how the system is configured.

The number of documents to be scored per reader may not be reasonable. For example, if teachers are doing the scoring, it may be difficult to get each teacher to score 300 papers. It may be necessary to use a higher ratio of validity papers or some other approach to ensure fidelity of scoring processes.

The Contractor acknowledges the potential benefit of using an increased ratio of validity papers in order to offset the scoring task expected of teachers (i.e., 300 papers). An additional, alternative approach for checking the validity of teacher-scored results includes embedded check sets containing five or more student responses. These would include moderator/committee-approved student responses which would be cross-checked against teacher scores.

The Contractor will not be compensating teachers to score Ohio program assessments.

38. Leadership, Staffing and Professional Development. p. 77-78: The State accepts the proposed staffing with the provision that the experimental nature of this Project will require some adaptability of the Contractor's staff to satisfy program contingencies. This Project will have some elements that are not "business as usual".

39. Building Capacity. On p. 86, the State is concerned that The Contractor does not mention the Range Finding committee in the first paragraph of this section. The Range Finding committee is an important part of the external review process and will be primarily responsible for verifying scores on student work samples as well as for identifying potential scorer training materials.

    The State is concerned that The Contractor is leaving this committee out of the loop here and in other areas of the proposal. The State's intended purpose for this committee is to review student work and score it and discuss how the rubric is implemented and identify areas where the rubric needs clarification or improvement.

    The Contractor acknowledges the importance of the Range Finding committee as an important part of the external review process as well as criticality for verifying scores on student work samples and selection of scorer training materials.

40. Sustainability Plan. On p. 87 at the top of the page, this paragraph describes purportedly, how the State can sustain this model. The State does not feel that the proposal of seeking funding from the Gates Foundation is a sustainability plan. The Gates Foundation is not intended to sustain assessment systems, but to ramp up new and innovative models. A sustainability plan should include a partner that has a vested interest in Ohio's assessment system, like a large employer within Ohio who is interested in building capacity within their own organization or developing a pipeline. Some candidates might be NASA, the Wright-Patterson Air Force Base, American Electric Power, Ohio United Health, Proctor and Gamble, or Honda. The ODE expects that The Contractor will continue to seek a partnership with one or more of these companies (not necessarily limited to those listed above) to achieve the sustainability requirement. The State will work with The Contractor to help identify potential partners.

41. Assumptions. The Contractor's intent is to take sufficient time to develop and revise quality learning and assessment tasks. This includes developing a small overage in the event that an idea for an assessment or learning task is not accepted during reviews.

    Both parties agree that exposure to some of the problems with tasks can be mitigated by writing and following the task specification meticulously and by working closely with state staff.

42. p. 102: As mentioned earlier, the benchmarking of tasks is to be done by the Contractor with input from the Range Finding Committee and the State. The State does not have the staff and the teachers do not have the expertise to do this. Perhaps, the Moderation Panel will be able to do this, but that cannot be piloted in the timeframe allotted.

43. p. 103: The pacing of the program must conform to cycles of school year and fiscal year. Therefore, in general, strategies and tactics must be developed and executed to ensure the project goals are accomplished. ODE has the expectation that the Contractor will be able to deliver the first batch of learning tasks for review in time for them to undergo external review by the end of the calendar year and be implemented in classrooms in the spring semester.

    The State is already engaged in recruiting participants for Cohort 3.

44. The State will pay only for services rendered.

45. There will be no separate or additional reimbursement for travel or other related expenses.

46. All Contractor invoices shall be submitted to:

    The Ohio Department of Education
    Office of Assessment
    25 South Front Street, Mail Stop 507
    Columbus, OH 43215

47. UNSPSC Category Code: 86141500 – Educational Guidance Services.

COST SUMMARY

The Ohio Performance Assessment Pilot Program (OPAPP) – CSP904911 EDU089

OPAPP OAKS Item ID No.: 000000000000019524
OPAPP Videography OAKS Item ID No.: 000000000000019525

DESCRIPTION	DELIVERABLES	COST
Stage 1 (anticipated completion 12/31/11):	Work with ODE Leadership to implement pilot: Develop schedule for Stages 1 and 2 with ODE. Work with ODE to prepare professional development for Cohort 1 and Cohort 2 for Stages 1 and 2.	$ 144,661
	Communicate with Pilot Participants: Communicate schedules for Stages 1 and 2 with all participants from Cohorts 1 and 2.	$ 4,204
	Task Development: Develop tasks for Stage 2 implementation and complete review process of tasks developed (number of tasks to be developed listed in Table 4). Make all necessary changes to developed tasks.	$ 476,117
	External Committees: Assemble high school external review committees. Secure locations and hold sessions for review of items developed for Stage 2 implementation.	$ 28,829
	Professional Development: Secure locations for professional development sessions. Deliver professional development for Stage 1 to Cohorts 1 and 2 (8 hrs and 64 hrs, respectively) to all participants (teachers, coaches, administrators, higher ed Education and Content experts).	$ 104,329
	Scoring: Benchmark student papers for ELA short task for Cohort 1. Hold scoring session with ELA teachers in Cohort 1.	$ 387,810
	Stage 1 Total	**$ 1,145,950**
Stage 2 (anticipated completion 6/30/12):	Work with ODE Leadership to implement pilot: Work with ODE to develop schedule for Stages 3 and 4. Work with ODE to prepare professional development for Stages 3 and 4.	$ 223,663
	Communicate with Pilot Participants: Communicate schedule for Stage 2 with all participants (including coaches and higher ed experts) from all cohorts.	$ 4,204
	Task Development: Develop tasks for Stage 3 implementation (both Cohort 2 and Cohort 3 -- number of tasks to be developed listed in Table 4). Complete review process of tasks.	$ 607,001
	External Committees: Assemble elementary school external review committees, secure locations and hold sessions for review of items developed for Stage 3 implementation. Assemble range finding committees for high school. Secure locations and hold range finding committee meetings.	$ 76,138
	Professional Development: Secure locations for Stage 2 to Cohorts 1, 2 and 3 (8 hrs, 32 hrs, and 64 hrs, respectively). Deliver professional development for Stage 2 to all cohorts.	$ 225,989
	Scoring: Secure locations for review of scoring practices and training for scoring. Score short assessment tasks implemented in Stage 2.	$ 161,110
	Stage 2 Total	**$ 1,298,105**

DESCRIPTION	DELIVERABLES	COST
Stage 3 (anticipated completion 12/31/12):	Work with ODE Leadership to implement pilot: Work with ODE to finalize PD for Stage 4.	$ 146,149
	Communicate with Pilot Participants: Communicate schedule for Stage 3 with all participants (including coaches and higher ed experts).	$ 4,331
	Task Development: Develop tasks for Stage 4 implementation (both Cohort 2 and Cohort 3 – number of tasks to be developed listed in Table 4). Complete review process of tasks developed for Stage 4 implementation. Make all necessary changes to developed tasks.	$ 500,376
	External Committees: Secure locations for review of items developed for Stage 4 implementation. Hold sessions for review of items developed for Stage 4 implementation. Secure locations for range finding committee meetings. Hold range finding committee meetings.	$ 66,528
	Professional Development: Secure locations and deliver professional development for Stage 3 to Cohorts 2 and 3 (32 hrs and 64 hrs, respectively).	$ 182,413
	Scoring: Secure locations for review of scoring practices and training for scoring. Begin training for regional moderation panels (16 hrs).	$ 184,831
	Stage 3 Total	**$ 1,084,628**
Stage 4 (anticipated completion 6/30/13):	Work with ODE Leadership to implement pilot: Work with ODE to prepare schedule and professional development for Stages 5 and 6.	$ 227,521
	Communicate with Pilot Participants: Communicate schedule for Stage 4 with all participants (including coaches and higher ed experts).	$ 4,331
	Task Development: Develop tasks for Stage 5 implementation (both Cohorts 2 and 3, number of tasks to be developed listed in Table 4). Complete review process of tasks developed for Stage 5.	$ 408,067
	External Committees: Secure locations and hold sessions for review of items developed for Stage 5 implementation. Secure locations and hold range finding committee meetings.	$ 53,222
	Professional Development: Secure locations for professional development for Stage 4 to Cohorts 2 and 3 (32 hrs each). Deliver professional development for Stage 4 to Cohorts 2 and 3 (32 hrs each).	$ 50,043
	Scoring: Secure locations for review of scoring practices and training for scoring. Hold panel meetings for review of scoring practices and training for scoring.	$ 253,522
	Stage 4 Total	**$ 996,706**

DESCRIPTION	DELIVERABLES	COST
Stage 5 (anticipated completion 12/31/13):	Work with ODE Leadership to implement pilot: Work with ODE to finalize PD for Stage 6.	$ 150,510
	Communicate with Pilot Participants: Communicate schedule for Stage 5 with all participants (including coaches and higher ed experts).	$ 4,459
	Task Development: Develop tasks for Stage 6 implementation (Cohort 3 only – number of tasks listed in Table 4). Complete review process of tasks developed. Make all necessary changes to developed tasks.	$ 251,389
	External Committees: Secure locations for review of items developed for Stage 6 implementation. Hold sessions for review of items developed for Stage 6 implementation. Secure locations for range finding committee meetings. Hold range finding committee meetings.	$ 66,528
	Professional Development: Secure locations for professional development for Stage 6 to Cohorts 2 and 3 (16 hrs and 32 hrs, respectively). Deliver professional development for Stage 6 to Cohorts 2 and 3 (16 hrs and 32 hrs, respectively).	$ 144,449
	Scoring: Secure locations for review of scoring practices and training for scoring. Hold panel meetings for review of scoring practices and training for scoring. Continue training for regional moderation panels (16 hrs).	$ 126,061
	Stage 5 Total	**$ 743,396**
Stage 6 (anticipated completion 6/30/14):	Communicate with Pilot Participants: Communicate schedule for Stage 6 with all participants (including coaches and higher ed experts).	$ 234,324
	External Committees: Secure locations and hold range finding committee meetings.	$ 26,635
	Professional Development: Secure locations and deliver professional development for Stage 6 to Cohort 3 (16 hrs).	$ 54,948
	Scoring: Secure locations and hold panel meetings for review of scoring practices and training for scoring. Complete training for regional moderation panels (16 hrs).	$ 239,810
	Stage 6 Total	**$ 555,717**
All Stages	Videography fees	**$ 400,000**
NOT-TO-EXCEED TOTAL PROJECT COST*		**$ 6,224,502**

All costs are in U.S. Dollars.

The State will not be responsible for any costs not identified.
There will be no additional reimbursement for travel or other related expenses.

* The Not-to-Exceed Total Project Cost shall be the sum of the respective stage's not-to-exceed cost, and includes fiscal years 2012 – 2015.

CONTRACTOR INDEX

CONTRACTOR AND TERMS:

OAKS Vendor ID: 194141

Measured Progress
100 Education Way
Dover, NH 03820

CONTRACT NO.: CSP904911-1 (08/31/14)

TERMS: Net 30 Days

CONTRACTOR'S CONTACT(s):
Ms. Dona Carling
Director of Client Services

Telephone: (800) 431-8901, ext. 2213
e-mail: carling.dona@measuredprgress.org