

Table of Contents

PART 3: Project Narrative Attachment Form

Application Narrative.....	2
ABSTRACT.....	2
Response to Selection Criteria and Sub-criteria	3
a. Theory of Action.....	3
b. Assessment Design	6
c. Assessment development plan	17
d. Research and evaluation	22
e. Professional capacity and outreach	26
f. Technology approach	27
g. Project management.....	29
Description of How Proposed Project Addresses the Absolute Priorities	39
a. Design Principles	41
b. Technical quality:	43
c. Data.....	43
d. Compatibility	44
e. Students with the most significant cognitive disabilities	44
Description of How Proposed Project Addresses Competitive Preference Priority 1	45
Governance – Executive Board: Thought and Policy Leadership	48
Consortium Council.....	49
Task Management Teams	50
Project Management Partner.....	59
Technical Advisory Committee (TAC)	60
Third-Party Research and Validation	60

APPLICATION NARRATIVE

The purpose of this proposal is to enhance the quality of assessment instruments and systems used by states for measuring the development of students' English language proficiency (ELP). It is our belief that no ELP assessment currently exists which corresponds in deep and meaningful ways to the Common Core State Standards (CCSS) in mathematics and English language arts/literacy.

Our proposed project, the English Language Proficiency Assessment for the 21st Century (ELPA21), will use multiple measures of students' English language proficiency—among them a diagnostic/screener test, elective interim benchmark assessments, and an annual summative assessment. Not to be undervalued is information that is gleaned from individual teachers' formative assessment practices. The assessment design ensures that instruments for measuring students' ELP are comprehensive and include assessment items that are performance-based as well as technology-delivered.

The ELPA21 is valid, reliable, and fair for its intended purposes; will be used by multiple states who have agreed to a common definition of English Learner, including common criteria for entry, placement, and exit; and will correspond to a set of English proficiency assessment standards which in turn correspond to college- and career-ready standards in English language arts (ELA) and mathematics.

ABSTRACT

The English Language Proficiency Assessment for the 21st Century consortium (ELPA21), led by Oregon as the governing state in partnership with twelve other states, Stanford University, and CCSSO, has formed to develop an English Language Proficiency Assessment that is aligned to the Common Core State Standards (CCSS). ELPA21's proposed assessment design is intended to ensure the valid, reliable, and fair assessment of the critical elements associated with English language acquisition and mastery of the linguistic skills linked to success in mainstream classroom environments. In addition, ELPA21's proposed assessment will support ongoing improvements in instruction and learning that are useful for all members of the educational enterprise, including students, parents, teachers, school administrators, members of the public, and policymakers. This assessment will incorporate principles of Universal Design and will comply with Accessible Portable Item Profile (APIP) standards. ELPA21 development

will be based upon the prior successes of member states (for example, the Kansas writing tool, the Michigan diagnostic screener, test items from Iowa and Louisiana, and online test delivery specifications from Oregon).

The deliverables for the diagnostic screener and summative components of ELPA21 will include open-source: performance level descriptors, item banks for diagnostic/screener, for practice and for operational delivery, psychometric scale, performance levels (cut scores), test design and delivery specifications, test specifications and blueprints, professional development resources, and administration and security protocols.

With participating states who are currently part of PARCC, Smarter, National Center and State Collaborative, and Dynamic Learning Maps, ELPA21 will strive to work with these consortia to maximize compatibility and interoperability across user platforms. These resources as well as model Request for Proposal language will be available to states for use (individually or in multi-state partnerships) to contract with vendors for operational assessment in the 2016-2017 school year.

RESPONSE TO SELECTION CRITERIA AND SUB-CRITERIA

a. Theory of Action

The proposed assessment system is based on the tenet that in order for resulting data to be used in meaningful ways, strong alignment must exist between college- and career-ready standards, English language proficiency (ELP) standards, language instruction, and multiple measures of English language proficiency. As a way of anchoring this alignment, a framework developed by the Council of Chief State School Officers (CCSSO), and the ELP/D Framework Committee, titled *Framework for English Language Proficiency/Development Standards Corresponding to the Common Core State Standards and the Next Generation Science Standards*, hereinafter referred to as the “Framework,” will be used to guide the development of common ELP assessment standards that will be adopted for use by all of the consortium states. The Framework authors, experts in the field of ELA, mathematics, science, English language acquisition, and English language learner (ELL) education, will ensure that the ELP assessment standards correspond to the Common Core State Standards (CCSS) in English language arts and

mathematics (and additionally science in anticipation of the next generation science standards). These ELP assessment standards will in turn align with the ELP assessment system described in this proposal.

Furthermore, data produced by the ELP21 will inform language instruction in the classroom on an ongoing basis as students progress toward college- and career-readiness. Specifically, the assessment system will (for multiple grade bands: k, 1, 2-3, 4-5, 6-8, 9-12) include a screener/diagnostic test, an item bank from which interim benchmark assessments may be locally constructed, a summative assessment, recommendations to teachers with respect to formative assessment practices that can be integrated into classroom instruction, and a secure consortium student outcome data cooperative for reporting and research. The work supported by this grant will first develop the screener diagnostics and summative assessments to be used by this multi-state consortium. The development of interim benchmark assessments, supporting professional development, recommendations on formative assessment practices, and the cooperative data reporting will begin, but will not be completed during this grant period. Additional grant money will be sought from other sources to complete the development of these components of the assessment system. The assessment system to be developed has been designed to ensure that it is logical, coherent, credible, and capable of resulting in improved student outcomes. The following section provides a description of how the proposal meets these qualifications.

1. How will the assessment results be used?

The assessment results will be used to make determinations of school, local educational agency (LEA), and state effectiveness for the purposes of accountability under Title I and Title III of the Elementary and Secondary Education Act (ESEA). More specifically, state departments of education will use summative assessment data to determine which districts are meeting annual measurable achievement objectives (AMAO) targets, and identify school districts in need of assistance. At the LEA level, summative assessment data will help determine which schools are effectively moving students at various levels of English proficiency forward. Schools that are not moving students to proficiency should be carefully reviewed and corrective action taken. LEAs can also use summative assessments over time to

determine program effectiveness for various subgroups of students and adjust educational programming as needed.

The practices of successful instruction can then be observed, analyzed, and shared with others. At the classroom level, teachers will use formative assessment data as a part of their instructional practice to give students immediate feedback and determine whether instruction results in positive outcomes and whether grouping strategies result in success (Heritage, 2001). At the student level, data from interim benchmark assessments will be used to set goals and monitor progress during the school year. Students will become familiar with assessment rubrics written in student-friendly language in advance of testing so that when they receive their scores they will understand what the scores mean.

2. How will the assessment and assessment results be incorporated?

All ELPA21 assessments will be aligned to common ELP assessment standards which in turn correspond to college- and career-ready standards. All tests across the 13 states in the consortium will therefore measure progress on the same constructs. ELPA21 consortium states will adopt common cut scores and performance levels on screener and summative assessments for purposes of determining program entry, ELP level, and program exit. This facilitates a cohesive assessment system across states which will be particularly valuable in responding to high student mobility. An ELL who is a level 2 in one state who moves to another district or even to another state within the consortium will have his/her ELP data available within the requirements of the Family Educational Rights and Privacy Act of 1974 (FERPA). The student will be more properly placed in his/her new school, thereby enabling the student to continue benefiting from appropriate standards-based instruction.

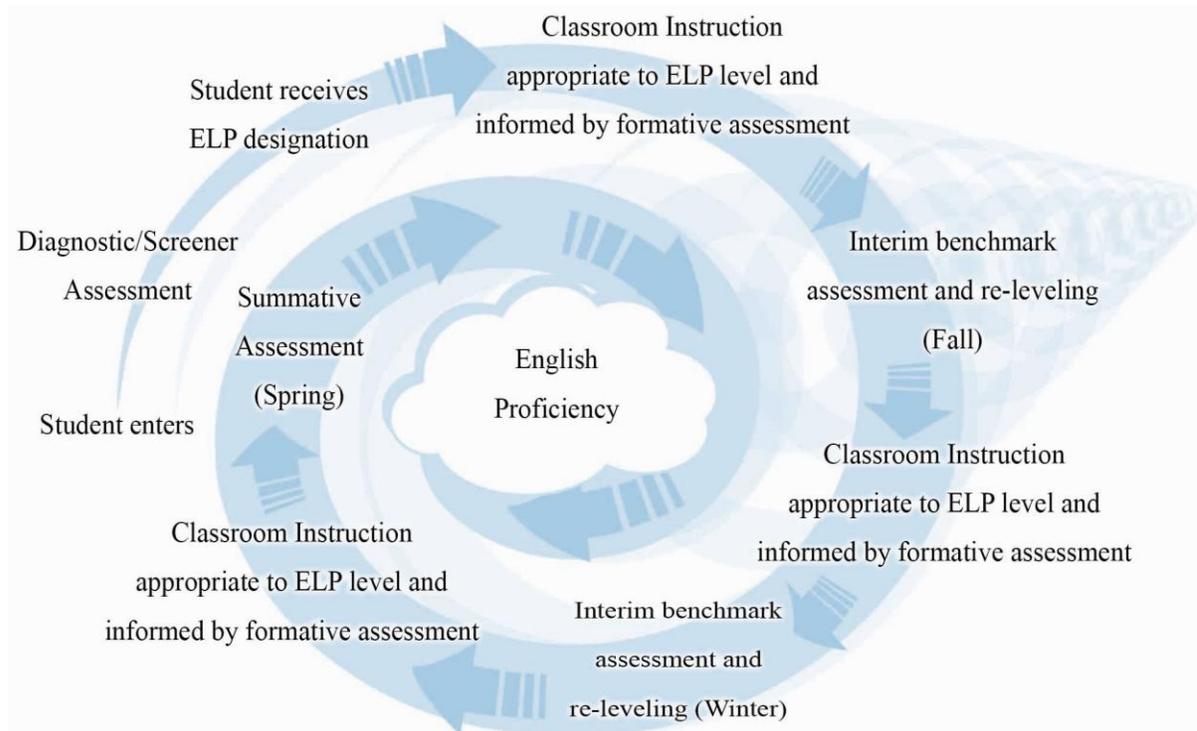
3. How will these educational systems as a whole improve student achievement?

Each assessment component plays a critical role with one building on another and each contributing to the overall efficacy of the assessment system and then ultimately to student achievement. For example:

- Accurate screening facilitates the proper identification and placement of students, which results in appropriate instruction;

- Interim benchmark assessments, state or locally drawn from the ELPA21 secure assessment item bank, provide teachers with timely data as to the effectiveness of instruction throughout the academic year and allow them to re-group students to receive better targeted instruction; and
- Formative assessment practices allow teachers, on a day-to-day basis, to determine which students need additional assistance in meeting specific standards.

The assessment system, therefore, establishes a continuous feedback loop to districts, schools, and teachers who are better able to adapt instruction, improve teacher training, restructure programs, and/or adopt better materials to respond to students’ developing needs. The following is the sequence of ELPA21 assessments administered throughout the academic year.



b. Assessment Design

The purpose of this project is to develop an English language proficiency assessment system that provides reliable and valid measures of students’ level of English proficiency in the four domains of reading, writing, speaking, and listening. The ELPA21 resources will be available for all ELL students K-

12 in the consortium states for use individually or in multi-state partnerships to contract with vendors for operational assessments in the fall of the 2016-2017 school year for diagnostic and summative purposes.

1. Number and Types of Assessments

The proposed assessment system includes one diagnostic screener and 2 forms of a summative assessment of English language proficiency for each of six grade bands. The assessments will include selected-response and performance items delivered via advanced technology as much as feasible.

2. Measuring Student Knowledge and Skills

At every point feasible and valid in structuring the assessment design, the ELPA21 consortium will make use of consortium states' already existing materials, artifacts, assessment items, procedures, and protocols after engaging in a valid and rigorous vetting process. The consortium will leverage the good ELP assessment work already implemented in consortium states in order to conserve precious development costs, build on successful and proven ELP assessment tools and practices for the required screener and summative assessments, and provide additional, no-cost supporting components of a complete ELP assessment, such as formative instruction guidelines and administration training for ELL teachers.

The ELPA21 consortium states have had individual, well-established, and operationally defined ELP standards for many years. Our preliminary review of the ELP standards from the participating states as well as the national Teachers of English to Speakers of Other Languages (TESOL) ELP standards shows a substantial commonality among state ELP standards. As no national or widely held ELP standards exist, the participating states and organizations working on this project are currently establishing a common set of ELP assessment standards as the basis of the work of this grant.

Appropriate attention will be paid to the need to develop proficiency in the discipline-specific language students will encounter in their academic content classrooms. In the context of assessments, language proficiency tests could vary in the degree to which they measure such discipline-specific language. Therefore, it is essential that ELP tests correspond to the college- and career- ready standards and measure the type of language proficiency needed to be successful in the academic content classrooms.

The term “correspond” indicates that ELP assessment items should include the language that facilitates content and learning, not to be aligned with the technical academic content. Thus, the ELP assessments should not be a measure of students’ content knowledge (e.g., math and science), but rather a measure of English proficiency that facilitates content learning.

The ELPA21 test blueprint will correspond to college- and career-ready standards as defined by the CCSS in mathematics and ELA, literacy in history/social studies, science, and technical subjects. Decisions about item types will be directly related to the ELPA assessment standards to be assessed and the best options that technology allows to provide the most authentic form of assessment. We will incorporate the use of technology-enhanced assessment items and performance tasks where necessary for valid assessment of particular standards. The blueprints will detail the standards appropriate to assess, as well as the number and types of assessment items that will be used to measure them.

ELL students arrive with wide ranges of English and academic proficiency, so it will be important that each grade band assessment assesses a wide range of ability. A vertical psychometric scale of ELP will allow ELPA21 to capture the progress students make between annual administrations of the summative assessment. With the subsequent addition of interim assessments, the vertical scale will allow progress to be monitored within school years.

3. How the assessments will produce the required student performance data

Student performance data produced by ELPA21 will include raw scores, scale scores, and performance levels. Data will be reported for each of the four domains, as well as a composite ELP score that will place students on a vertical ELP scale that goes from K-12 and tracks student progress over time. Performance levels will be defined through the bookmark standard setting process to establish cut scores separately for each grade level that create ELP performance levels such as Pre- Emergent, Basic, Intermediate, and Proficient (National Research Council, 2011). The standard setting team will include a diverse and representative group of teachers, administrators, and other key stakeholders from ELPA21 states who have expertise in areas such as language acquisition and serving LEP students with disabilities. The cut scores obtained through the standard setting process will be validated using external criteria such

as students' scores on states' current ELP and reading language arts assessments and teacher evaluations of students' ELP levels.

Decisions about including any additional subscores is unlikely, but will be considered and will be based on the relevance of the subscore for making educational decisions and the number of assessment items that would be available to support a reliable and meaningful score.

4. How and when during the academic year different types of data will be available

The diagnostic screener will be scored primarily by computer though some parts of the writing and speaking assessments will be scored by local educators who have been trained to do so. Results of the diagnostic screener will be available quickly and used for initial identification and placement into programs. The outcomes of this assessment will also be made available to teachers to inform instruction and curriculum planning.

The ELPA21 summative assessment will be administered near the end of the academic year depending on each member state's controlling state assessment schedule. Results of this assessment will be used as a measure of student learning which can provide data for accountability systems. The summative assessment results will be used to measure ELL students' progress toward acquiring ELP (AMAO1) as well as their attainment of ELP (AMAO2). At the student level, the annual summative assessment will inform decisions about student reclassification for the subsequent school year, and will provide critical information about students' ELP levels to the following year's teachers. Because all summative ELP assessments will occur in the same window, and because the summative tests will be moderately longer than the screener, some more complex item formats may be used, and external scoring will be utilized to manage the turn-around time which will be marginally longer than for the diagnostic screener.

5. The types of data that will be produced by the assessments

The assessment system will produce data that meet all of the criteria specified in Absolute Priority #5, and will take into consideration, to the extent possible, the varying contexts, conditions,

practices, and policies of the individual states. The consortium will develop materials and protocols to guide the standardized administration of the assessments.

The data will include student attainment of English proficiency as well as student progress, and can be disaggregated. Data will be collected in each of the four domains, as well as a comprehensive score based on all domains. Each domain will contribute significantly to scores at each proficiency level. Decisions about how the domains will contribute to a comprehensive score (e.g., weighting) will be finalized once field test data are available. One consideration is the dimensionality of the data. If the data support a unidimensional trait of English proficiency, a compensatory model based on a composite of all four domains will be meaningful. However, if the domain scores are not unidimensional, then a conjunctive model based on both domain scores and total scores will be considered.

Research has shown that in a compensatory model, ELL students may reach proficiency by being proficient in some, but not all, of the ELP domains of reading, writing, speaking, and listening. For example, ELL students who are above proficient in listening and speaking but below proficient in reading and writing may be deemed proficient in English because their high scores in listening and speaking compensate for their low scores in reading and writing (Abedi, 2008a). ELPA21 will study methodologies for creating composite scores and adopt a policy based on the results of that study to ensure the validity of reporting ELP assessment results.

Students can be classified by English proficiency status, and data will be collected on the effectiveness of schools, LEAs, and states. Data may be aggregated at the school, LEA, or state level for use in accountability systems. Furthermore, data tagging will allow information to be aggregated for all students of a particular teacher or principal.

The data-tagging capacity noted will apply the agreed-upon English proficiency standards to report at what level students are considered English learners and whether they qualify for particular interventions; how much their English proficiency has progressed; and what their current proficiency level is relative to the performance standards for their grade level. The tagging system will allow users to drill down to various ELL subgroups including the number of years in a language instruction educational

program, interruption in formal education, etc. Options will also exist to allow individual states to specify some disaggregation categories that are meaningful to their specific populations.

6. The uses of the data

i. Determining student achievement and progress: Data from the ELPA21 will serve multiple purposes. At the individual student level, data will include the student's scale score on the k-12 vertical scale of overall English language proficiency, the amount of growth the student made from the previous annual administration, the student's performance level in relation to grade-level ELP performance standards, and more detailed information for each domain. The aggregation of data is a useful measure for identifying professional development needs, and for informing a teacher about the instructional needs of his/her students.

ii. Informing teaching, learning, and program improvement: The ELPA21 information reporting templates on the performance of individual students will allow teachers and parents to monitor the student's progress, make placement/reclassification decisions, and adjust instructional strategies as needed. The project intends to specify results formats to inform families in ways that will help them understand their children's progress. Where possible, that means that student results would be translated into the language that makes them most accessible to families. Focus groups will be conducted with parents to explore the best format for presenting the assessment results.

Aggregate data at the teacher and principal levels can help educators to know what is working and what needs to be improved. The data can help teachers develop more effective instruction for ELL students, and it can help school officials know the types of professional development and/or support that will help teachers to better address the needs of their students.

The ELPA21 will draw on existing materials, artifacts, and protocols from consortium states to provide professional development modules for all teachers (including academic content teachers) on how to best use the ELPA21 assessment results, including how to discuss them with families and students. As appropriate, data regarding student progress on acquiring ELP may also be used as one of multiple measures to contribute to a state's own developed educator effectiveness system. At the consortium level,

a wealth of data will be available allowing for research to be conducted across states with substantially larger pools of students than is typically available when states operate their assessment systems independently.

7. Frequency and timing of assessment administration and rationale

The ELPA21 includes two assessments: (1) a diagnostic screening assessment will provide information for identification and placement, and (2) the summative assessment will be used to monitor student progress, accountability, reclassification, and instructional improvement.

The ELPA21 diagnostic screener will be administered at the time a student enters the school system, k-12, and may be re-administered as needed. The primary purpose is to determine whether, and at what level, the student requires services to develop his/her English proficiency. The diagnostic screener is a more brief assessment than the summative assessment, but it does assess the four domains. Both the diagnostic screener and the summative assessment will be administered by computer, using primarily selected-response assessment items for reading and listening. The writing and speaking portions of the diagnostic screener and the summative assessment will require students to produce output that is more validly captured using constructed responses.

The summative assessment will be administered annually near the end of the school year consistent with each state's test schedule. We believe that a comprehensive assessment system for ELLs should include formative assessment at the time of instruction and interim assessments to monitor progress throughout the school year. While these components are beyond the fiscal scope of this proposal, we hope to subsequently develop them through leveraging existing, contributed resources from consortium states and refine them with funding from other sources.

Once a student is identified as an ELL and therefore eligible for placement in an English learner program, he or she will be assessed annually using the annual summative assessment. Two forms of the annual summative assessment will be available for each of the six grade bands. The computer-based summative assessment will more fully assess the same four domains as the diagnostic screener. The item

types, however, will be more varied on the summative assessment, taking full advantage of the opportunities presented by technology to offer more extensive and richer performance tasks.

The rationale for the proposed ELPA21 has many advantages over the existing ELP assessments developed based on ESEA Title III guidelines. The proposed ELPA21 will produce assessments that serve multiple purposes and whose outcomes can be used to improve curriculum planning and instruction. The use of vertically articulated grade band assessments will provide additional information to teachers and administrators. Results from vertically articulated summative ELP assessments can inform teachers about students' areas of strength and weakness and can also be used by teachers to plan instruction based on individual students' knowledge and understanding of English. The outcomes of ELP assessments that are used diagnostically can help identify areas in which ELL students need assistance in order to succeed in core academic subjects and the corresponding assessments.

8. The number and types of assessment items

The question of *how* to measure that content (e.g., how many assessment items of what types) will follow through the collaborative development of a test blueprint which will serve as the basis for item development. For each of the four domains, content advisory panels of seven subject-matter experts will be convened from ELPA21 consortium states. They will be charged with describing the mix of assessment items, by item type and cognitive demand, which will best assess the ELP assessment standards of the domain.

Since the exact test blueprint will be developed under this project, specifics will evolve from those discussions. We do expect, however, that different ELP domains will require different item types to fully assess the constructs they represent. To the extent practicable, constructed-response or performance-based assessment items will be included in the assessment of all four domains.

Abedi (2010) suggests that performance-based assessment items may provide more opportunity for ELL students to present what they know and are able to do. These types of assessment items encourage more participation and reduce guessing. Generally speaking, given the make-up of validated assessment items contributed to the item bank by ELPA21 consortium states, we anticipate a listening test which is 25%

constructed-response following technologically delivered prompts; a speaking test which is 80% rubric-controlled rating from student constructed response; a reading test which is 20% rubric-controlled reader rating from student constructed response; and a writing test which is 80% rubric-controlled scoring of an extended writing sample and 20% rubric-controlled scoring of short-answer written responses.

Following are two sample assessment items, one selected-response and one performance-based. Both are based on the CCSSO-commissioned Framework.

Sample Item: Selected-Response. The CRSS response item that follows is a speaking task intended for middle grades.

Vitamins and Minerals

Vitamins		Minerals
<p>Source: plants and animals</p> <p>Water or fat soluble</p> <p>Organic</p> <p>Help immune system</p>	<p>different</p>	<p>Inorganic</p> <p>Help bone health</p> <p>Water soluble</p> <p>Source: soil and water</p>
<p>same</p>		
<p>Help immune system</p> <p>Support growth and development of body</p> <p>Help body work correctly</p> <p>Help to create strong bones</p>		
<p>different</p>		

Using the information provided in the Venn diagram, please answer the following questions. Use scratch to prepare your answer before recording, if necessary.

1. Compare vitamins and minerals, and briefly tell what they do for your body.
2. Name 2 differences between vitamins and minerals.

Record your response when ready:




Once the test blueprint is finalized, the project will harvest appropriate secure items from ELPA21 consortium states for possible screener or summative assessment use and non-secure items for inclusion in the interim benchmark test bank. After assessment items have been harvested and a gap analysis conducted, the project will enter into the item development process to complete the requisite assessment item pool.

9. The assessments' administration mode

The ELPA21 will employ computer-based delivery. While the assessments could be presented in a paper-pencil format if required, the computer-based mode is preferable because it will enhance standardized administration of the assessments across multiple states while providing unique opportunities to present accommodations recommended in the Individualized Education Program (IEP) and Section 504 plans of LEP students with disabilities. These accommodations, consistent with the planned common core test administration of the future state assessment systems (e.g., PARCC and Smarter-Balanced), would be difficult to apply through paper-pencil delivery. Rater training can be facilitated quickly and economically through computer-based practice and feedback. Innovative items and ELP assessment standards difficult to assess using paper-pencil would be available through computer delivery.

We are mindful of issues raised by researchers about the reality of the “digital divide.” Students’ and families’ socio-economic background and students’ access to computers at home are worthy of consideration (Goode, 2007; U.S. Department of Education, 2007). As access to technology continues to widen across all schools through the impacts of the other assessment development consortia, especially Smarter-Balanced and PARCC, we believe these issues will have less impact by the time the ELPA21 test is ready for administration in the 2016-17 school year. The ELPA21 project team will ensure that it is possible to take the ELPA21 using a paper-pencil version of the test and will verify the comparability between the computer version and the paper-pencil version.

10. Methods for scoring student performance, turnaround times, and rationale

The computer-based mode will allow for selected-response assessment items and some constructed-response assessment items to be scored in real time. Other assessment item types will be scored according to rubrics developed in conjunction with the development of those assessment items. For the diagnostic-screener, quick turn-around time is critical. Students entering school need to be appropriately placed and receiving the services they need as soon as possible. Assessment items that necessarily require human scorers, predominantly in the areas of speaking and writing, will be scored by

trained local educators. As soon as educator ratings are entered into the computer, reports will be available summarizing scores for each domain and an overall English language proficiency level.

For the annual summative assessment the premium is on information with the highest technical quality possible. Where constructed response, even extended response, assessment items can do a better job of assessing the desired construct, we will make every effort to incorporate those assessment items into the testing blueprint. We understand that such assessment items require extended time and resources for scoring. States may choose to use an external vendor for scoring the annual summative assessment, or they may have groups of educators trained to work online, over weekends, or possibly in the early summer. Because ELPA21 is providing the materials and protocols for the assessments, their administration, scoring, and reporting, the conduct of those activities continues to rest with the individual state to be best managed as their assessment programs have identified through their years of experience.

To ensure scalable, accurate, and consistent scoring of assessment items, any person scoring the ELPA21 (including teachers) must have completed the ELPA21 scoring certification course. The course will be developed after rubrics are finalized and exemplar responses for each item type are determined. The course will include a calibration component requiring participants to obtain a minimal percent perfect agreement and not exceed a pre-set maximum score bias in either direction (scores on a reference set of responses will be set by a group of expert raters).

11. Reports on the assessments will include key data, its use, target audience, and presentation

Materials and protocols for a web-based reporting system will be architected for access from an interactive platform for different audiences to access a variety of indicators at multiple levels of aggregation from existing practices and tools in the ELPA21 states. Other consortium members will be able to access/adapt as their own existing state systems by leveraging the data structures and process flow described in the ELPA21 tool.

Individual Student Reports will provide the overall composite ELP score on the k-12 vertical scale, scores for each of the four domains, a composite score, the student's performance level in relation to grade level expectations, a representation of the student's growth over time, and suggested

interventions for a typical student scoring at the student's level. As is required by FERPA, information about students will be carefully protected through a tiered system of secure passwords and permissions. Roster reports will provide lists of students and their associated scores (as described for the student report). Groups of students for roster reports can be defined based on many different criteria (e.g., students of a particular teacher, students that have exited the English language instruction program, students with a particular first language). Roster reports containing student names are again closely protected through security protocols.

Summary Reports will provide aggregate data only for groups of students. Performance can be aggregated as average proficiency scores, average growth from the previous annual summative assessment, and percentages by proficiency levels. Summary reports for groups such school, district, or state will be useful for accountability purposes.

In order to ensure that reports are providing useful and understandable information, and that the system provides user-friendly access, a series of focus groups will be held separately with families, teachers, and administrators to gather feedback to guide revisions to the system. To the extent practicable, student reports will be printable and translated into the language that is most accessible to the student and his/her parent(s).

c. Assessment development plan

Validity is the main concern in assessment development, and the current unified theory of validity will be used. Item development and validation is the most important activity in this section because items are the building blocks of any test. Technical reports, validity studies, and external, third-party evaluation provide a basis for evaluating the quality of this assessment system and ensure that valid interpretations and uses of these scores will be attained.

As often stated in this proposal, the assessment system produces test scores for summative evaluation that may qualify an ELL for exit from the ELL program, if other data (multiple-measures) also provide satisfactory evidence of English proficiency. Consortium states will decide how and what combination of multiple measures will be used, and recommendations will be made as to how this is best

done. Additionally, the assessment system will produce data for other purposes, and the design of the system will address the validity of these different uses.

Important influences on the development of this assessment system are: *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999); the unified approach to validity (Kane, 2006a, 2006b; Messick, 1989) which addresses the accuracy of test score interpretations and uses; and, the *Handbook of Test Development* (Downing and Haladyna, 2006).

1.i. The approaches for developing assessment items: The current technology for item and test specifications appropriately considers complexities of the ELP assessment standards. The item and test specifications document will direct the item development phase of this project and enable test design.

For each of the four language domains, content advisory panels of seven subject-matter experts will convene from member states. Each member will be highly qualified by experience and education. One of the first tasks is to develop test specifications that identify the content categories by which tests are designed and the appropriate cognitive demands. A simple taxonomy for cognitive demand will be used that has three categories: recall, comprehension, and application. The test specifications will provide a basis for filling the item bank with appropriate numbers for both the summative and screening tests. The ELP assessment standards will emphasize application, which entails the use of knowledge and skills in complex ways to perform tasks that have a decidedly higher cognitive demand. For instance, reading will involve testlets that are content oriented; writing will have a specific task objective or purpose.

Item formats include selected-response (multiple-choice) (SR), constructed-response with objective scoring (CROS), and constructed-response with subjective scoring (CRSS). SR item formats are numerous and very effective for measuring knowledge and many cognitive skills; this format will be used where appropriate. The CROS format is similar to the SR format, but generally requires higher cognitive engagement. The CRSS format is appropriate for content with a high cognitive demand, such as that involved with performance testing, but requires human scoring. Speaking and most writing items will use the CRSS formats.

The final step is the development of a plan for satisfying the item and test specifications leading up to a calculation of the number of items needed for the summative and screener item banks. The most valuable property in any assessment system is the bank of test items. According to Haladyna and Rodriguez (2012), the cost of a professionally developed item ranges between \$800 and \$2,200 each. The value of an item bank for this project will exceed six million dollars. Fortunately, economies and strategies will enable the production of these many validated items at a much lower cost by harvesting from states' existing assessment item banks those items which have sufficiently high psychometric properties, and then engaging new and targeted item writing activities following thorough gap analyses. This project intends to develop a validated item bank of considerable size that will be useful for screener and summative evaluation. In fact, this aspect of the assessment system is a primary activity. For instance, if a grade band summative test contains 50 items and two forms are needed for each of the six grade bands, the bank needs to minimally have 600 validated items arrayed to match the controlling test blueprint. For writing, the item bank concept will be different as most items will be in a CRSS format supplemented with some CROS formatted items. For the listening and reading domains, SR and CROS items will be used. CCSSO has successfully used a three-point CROS format for its speaking test, but new technologies using speech recognition promise superior results.

ii. Types of personnel involved in each development phase and process

Consortium states' ELP assessment specialists and psychometricians will contribute existing items and statistics. The Item Development Task Management Team in conjunction with the Lead State's dedicated staff and contracted vendors will align the existing items to the ELP assessment standards, identify gaps, and develop any necessary additional items.

2. Approach and strategy for accommodations

Accommodations will be provided in instances where student performance is affected by construct-irrelevant factors that diminish performance under regular administration practices. Such guidance on accommodations will be provided in consultation with experts via NCEO. Key questions about Accessible Test Item Characteristics for English Learners include:

Does the item:

- a) Meet general criteria for measuring what it is intended to measure?
- b) Have an overall appearance that is clean and organized?
- c) Have clear format for the text?
- d) Have clear pictures and graphics (when essential to item)?
- e) Allow changes to its format (and manipulatives, if used) without changing its meaning or difficulty (including visual or memory load)?
- f) Have flexible, easy-to-navigate presentation and response formats (for computer-based testing)?

3. Approach and strategy for ensuring scalable, accurate, and consistent scoring of assessment items

Vertical scales will be constructed for screener tests for each of the four language domains. Thus, any ELL can have progress accurately measured as they move upward through benchmarks. The screener tests will also measure each domain but will be grade-level appropriate and have diagnostic capabilities. Vertical scales will be constructed for the composite total score for many purposes including evaluating the progress of groups of students, evaluating programs, for research, and for evaluating staff performance. This composite score will be weighted as determined by an expert panel defined by a task management team (please see proposal section on Competitive Preference Priority 1 for further information) and informed by performance data. The vertical scaling will have test forms that are grade-band-appropriate, but also have some overlapping items so that continuous scales can be formed. Our member states have extensive experience in building vertical scales that will provide valid interpretations of measured growth from grade to grade.

Benchmark progress points will be set for each screener test using contrasting groups (grade appropriate) and expert judgment using the Bookmark method. This combination provides an empirical and judgmental basis for cut scores that identify levels of progress in ELP. The use of two methods provides a convergent validation of any cut score recommended by the expert committee for the

respective domain being measured. Cizek (in Downing & Haladyna, 2006) provides useful information about the desirability of both methods for standard setting. It is our intention to use both methods as a means of cross-checking and validating any decision made about student assignment to the regular classroom instruction.

This team will determine the weighting for the combination of language domain scores that comprise the total score. The team will also recommend the cut score for determining English proficiency. Determining cut scores follows standards in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999). Finally, the team will recommend how a combination of measures will be used to arrive at the final determination of English proficiency.

For SR items, scoring is automated. The degree of random error is very small. But it is important to note instances in which a student omits or fails to respond to items. ELL students have a higher frequency of non-response, as reported in National Assessment of Educational Progress (NAEP) studies. When non-response occurs, a test score can be invalidated or scored on the basis of items tried. CROS items may be used for the speaking and writing domains. CCSSO has found that its listening test for ELLs is highly reliable, but newer technologies may be superior. Thus, the CROS item may be used for the efficient measuring of speaking ability using speech-recognition software.

The CRSS format will be used for speaking and writing. Range finding methodology will help mitigate any risk to validity, and we intend to respond with e-rating technology wherever feasible and practical.

4. The approach and strategy for developing the reporting system

Score reports will be developed that are responsive to the specific purposes of this assessment system and in consultation with the users of these score reports. ELPA21 will develop a straightforward reporting system by which student data may be securely accessed and reports, aggregated by user role (e.g., teachers may aggregate for their own classroom) may be downloaded. However, we anticipate that nearly all consortium states will prefer to download the student data directly into their own state

assessment reporting platforms. Therefore, significant attention will be paid to defining data structures to enable this.

5. Overall approach to quality control

The most effective means for quality control is multi-fold. First, a well-developed project management plan should identify tasks, timelines, and personnel responsible for outcomes. Second, validity evidence will be collected during the project. Third, a Technical Advisory Council (see proposal section on Competitive Preference Priority 1) will be convened three times during each project year by the third party evaluation contractor to evaluate the assessment program. Finally, the *Standards* (AERA, APA, & NCME, 1999) address the need for documentation. We will produce summaries, minutes, and reports that are dated and titled for the activities of this project. Not only does this body of information contribute to the annual technical report, it also constitutes validity evidence supporting the development of the test. Our third-party evaluator needs this information to complete their work.

d. Research and evaluation

1. Plan for identifying and employing psychometric techniques

The national Center for Research on Evaluation, Standards, and Student Testing (CRESST) will serve as the independent evaluator for the proposed Consortium EAG effort. Under the direction of CRESST co-director, Dr. Li Cai, CRESST's effort will focus on validation of the comprehensive assessment system and particularly the two components targeted for development, a diagnostic screener, for initial identification and placement of new English languages learners, and summative assessment to measure the ELP performance of all students, including students with disabilities.

Validity is the overarching concept that defines quality in educational measurement. In essence, validity concerns the extent to which a variety of evidence demonstrates that an assessment measures what it is intended to measure, provides sound evidence for its intended decision-making purpose(s), and well serves that purpose(s). Tests themselves are neither valid nor invalid. Rather, it is specific interpretations and uses of test scores as evidence that are subject to validation. From the CRESST

perspective, validity is built into a test from its initial design and development and validation is an ongoing process throughout the test development and use cycle.

Validation involves first defining the interpretive argument that justifies the use of a measure for a specific purpose and then gathering evidence to evaluate the argument (Kane, 2004, 2006). The interpretative argument is comprised of a series of propositions that link scores from an assessment to specific interpretations of the meaning of the scores and to specific conclusions or decisions and uses to be made on the basis of test performance. For example, backward chaining from the ultimate use of ELP summative assessment results to measure school effectiveness in promoting individual student growth, relevant propositions at the most general level might be:

- Assessment *instruments are designed* to accurately and fairly measure ELP that students are expected to develop;
- *Scores* from individual instruments accurately and fairly represent students' ELP performance;
- *Scores* accurately and fairly reflect the ELP that students have developed over the course of the school year, or from one school year to the next; and
- Student growth based on the assessments can be accurately and fairly attributed to the contributions of schools.

Claims underlying each proposition provide fundamental criteria for substantiating each proposition. For example, claims for proposition #1 may include that the assessment items are aligned with ELP assessment standards and college readiness standards, the instruments reflect the depth and breadth of intended standards, they are designed to be developmentally appropriate, they are designed to be accessible for students with disabilities, they are free from design flaws that could introduce construct irrelevant variance, etc.; while claims for proposition #2 involve those related to the psychometric characteristics of the scores and their instructional sensitive. Appropriate designs, samples, and measures and/or evidence sources that can be used to evaluate each claim then are identified and implemented throughout the test development and validation process.

The interpretive argument thus is the measurement version of the consortium's theory of action and a tool for guiding and providing feedback for the Consortium's development efforts and for informing the design of on-going validity studies that can be used to strengthen the effort. CRESST's first task will be to work with ELPA21's Consortium Council and Executive Board (please see proposal section on Competitive Preference Priority 1 for further information) to specify the interpretative argument justifying each of the ELP assessments (diagnostic screener and summative).

2. Plan for determining whether assessments are implemented as designed and theory of action

The CRESST team will be a critical friend to the ELPA21's Council and Executive Board during the first two years of the project to help the Consortium support and substantiate claims that the assessment instruments are appropriately designed to meet their intended purposes. Substantiating these claims, in general, involves systematic expert review with established protocols and quantitative analysis of review results. We would expect that there would be systematic reviews of alignment, accessibility, fairness, and the technical soundness of item design, and that such feedback would be used to improve test items and instruments. As part of initial reviews, we also would recommend that intended score reports be reviewed for usability and utility by intended users. Attending early to score reporting requirements helps to ensure that utility is built into the instrument design and that the assessment can deliver technically sound scores at the desired grain size – or that realistic expectations are established early on.

During years 2 and 3, we will collaborate with the Council on planning for appropriate pilot and/or field test designs to evaluate both the psychometric and other technical qualities of ELP scores and scales and construct evidence supporting intended score interpretations and use. Feedback about item and instrument design and reporting also will be solicited from participating teachers. We anticipate representative field test samples of at least 20 schools with high proportions of English learners and, where possible, diverse language and SES groups, from each state, for each level of each assessment.

CRESST would act as advisors in design, sampling, and data collection and would provide oversight and independent analysis of technical findings. Among these would be psychometric analyses

including the range and distribution of scores, means, standard deviations, and standard errors of measurement. We will report item difficulties (P -values) and interrater reliability for constructed-response items scored by teachers and/or project staff, and we will compute IRT and generalizability statistics to evaluate item characteristics and score reliability. Appropriate dichotomous or polytomous IRT models will be applied to the item response data, and model fit will be examined to ensure that the items are contributing to the measurement of the construct. Misfitting items will be amended or dropped. Location parameters within the polytomous items will be examined to ensure that the items and the scoring rubrics are successfully distinguishing among different levels of ELP performance. As a measure of convergent validity, we will test the difference between the scores of students who are rated by their teachers as exhibiting high levels of English proficiency and with those of students who are judged to be low science achievers. Analysis of variance or independent t tests will be used to test these differences. Cognitive labs in which a sample of students think aloud as they respond to items will also be considered for new item formats.

During years 3-4, CRESST will advise on standard setting studies and implement additional field studies to evaluate the psychometric quality, validity, utility, and use of the assessments by intended stakeholders (e.g., district and school administrators, teachers) and will supervise review and feedback for professional development materials for teachers, specialists, and other stakeholders. Drawing on performance level descriptors, student test performance data, and predictive relationships with school success, CRESST will support appropriate standard setting and level setting in each of the four specified language domains, as well as identifying mastery of linguistics.

In addition to student responses, teachers also would be asked for data on students' opportunity to learn relative to ELP performance standards, and to provide collateral indicators of individual student language proficiency and progress so that issues of instructional sensitivity and convergent validity could be further studied. Studies of utility and use would involve interviews and surveys of intended users. Protocols will be specially developed to evaluate and provide feedback on professional development protocols and surveys designed to elicit feedback from users.

e. Professional capacity and outreach

Ensuring stakeholders are informed and invested is key to the success of any education transition – particularly when that transition involves assessment policy and practice. The ELPA21 will affect not only students in their classrooms, but teachers, administrators, parents, communities, districts, legislators, and state education agency staff.

Two key resources will be utilized in outreach and communications with teachers and administrators for implementation of the assessments and improving and informing instructional practice, and with the public and key stakeholders: the consortium states' education communication and assessment directors, and an expert-led task management team specific to communications and outreach made up of 2-4 SEA staff from the consortium states.

1. Plan for supporting teachers and administrators

The phases of support for teachers and administrators will follow the development and implementation of the assessments. Engagement with district-level staff will begin early using the experience and expertise of the ELPA21 communications and outreach task management team (please see proposal section on Competitive Preference Priority 1 for further information). Information disseminated through various media such as email, school meetings, online networks, web-based information sessions, and newsletter articles will focus on explaining the project timeline and the adjustments to the assessments and its resulting data. Identifying and training teacher and administrator advocates for ELPA21 is another key element in the plan. These advocates can support the assessments for their peers and can share their instructional practices and successes with the broader LEA, SEA, and consortium states.

2. Strategy and plan for informing the public and key stakeholders in the consortium states

Similar to the outreach plan for teachers and administrators, the outreach to the public and key stakeholders must begin early in the assessment system development process and must be phased. The strategy will highlight the benefits of the system to students, teachers, and ultimately the community and state. Leveraged resources, higher standards, college- and career-readiness, and multiple measures of

student achievement are all topics the public and key stakeholders will understand to be the resulting benefits of ELPA21.

Existing ELP programmatic and test-related communications from consortium states will be collected and gaps will be identified. New messaging materials will be developed or existing materials will be refined to address identified gaps in message or means of delivery. A comprehensive media package will be developed to communicate to the public and key stakeholders the essential features (including changes), benefits, and values of ELPA21. This package will include materials across multiple media to guide SEAs and LEAs in informing their various constituencies about the progress, implementation, and outcomes of the new assessment system. Various messages, built around these central themes, will be developed for dissemination within states specific to their state context and needs.

f. Technology approach

Technology will be used to provide quality, accessible, cost-effective, and efficient means of assessing ELLs. To the maximum extent feasible, the ELPA21 consortium will make extensive use of technology in test development and test administration, as well as for scoring and reporting, always with an eye to compatibility with other large-scale assessment consortia efforts.

1. Description and rationale for the ways technology will be used

The two primary technology-based issues that will be addressed as part of the development effort are interoperability and test item bank management. ELPA21 products include test items and blueprints for validated, calibrated test forms, not a software platform from which the tests may be delivered. There are already several significant delivery platforms on which the ELPA21 assessments may be delivered. The ELPA21 consortium will communicate regularly with PARCC and Smarter-Balanced consortia to understand the emerging plans for delivery; this will help ensure that the ELPA21 assessments can be administered using these platforms with the minimum of adaptation.

To maximize the likelihood of interoperability, it is important to develop and encode test items using a common standard. The APIP standard has become the industry standard for formatting and

encoding test items, and it focuses on item presentation accessibility, providing specifications for how to encode accommodations for special populations (e.g., Braille and tools). Item development will be facilitated through a digital banking and management system. The assessments will be designed to be computer-delivered, computer adaptive, and conform to UDL standards and APIP specifications with paper-pencil alternatives available. However, this grant's limited resources will provide only two fixed from assessments per grade band.

Recognizing that some schools may not have the technology to deliver the new ELP assessments, we will develop a paper-pencil alternative for any items that require the use of technology. Items developed for this effort that do not specifically require technology will already be XML-encoded to the APIP standard, allowing production of paper-pencil tests. This approach is consistent with the approach currently being considered by both the PARCC and Smarter-Balanced consortia.

Technology will play a significant role in the scoring and reporting process. Both individual student results and summary reports at the student, class, school and district level will be provided with a basic ELPA21 reporting software module. Students, with appropriate permission and following the security protocol will be able to obtain their results online via a secure website. Similarly, teachers and school or district administrators will obtain summary reports (rosters and summary data) online. Teachers or administrators, with appropriate permission and a secure log-in will follow security protocol to obtain these summary reports online via a secure website. The data structures in the ELPA21 reporting package will be freely shared with consortium member states so that those who wish may import the ELPA21 student performance information into their own report packages for display and dissemination.

2. How technology-related barriers will be addressed

To provide accessibility to all students, the ELPA21 consortium will employ the principles of Universal Design (UDL). In short, UDL in this context means the design of a computer interface that is usable to all students. More importantly, this usability will be provided without the need for special adaptations for individuals or subgroups; the accessibility accommodations will be built into the design of the item. The focus on UDL is to help ensure that each item is accessible to all students and that items

include tools and adaptations within to avoid the need for separate interfaces or platforms for students with special needs.

g. Project management

Project management expertise and oversight will be provided by CCSSO, which has developed resource management plans, budgets, and timelines with both phased and incremental development philosophies, oversight constructs, and numerous project checkpoints to bring the proposed ELPA21 system to fruition. The project will be managed by a PMI-Certified Project Manager using a variety of widely accepted and successful methodologies (PMI, 2008; Office of General Commerce, 2009), as appropriate at each stage of the project lifecycle.

1. Project workplan and timeline, including each key deliverable

The ELPA21 project will be managed as a series of sub-projects, each one phased and managed toward a single critical deliverable – the output(s) of each of the expert-led Task Management Teams (TMTs). These deliverables include:

- Multi-state English Language Proficiency Assessment Standards – Spring 2013
- Item Bank platform – Fall 2013
- Field Test Forms – Spring 2014
- Final Summative and Diagnostic Test Forms – Fall 2015
- Performance (Benchmark) Standards, Weighting, Cut Scores – Summer 2015
- Report Templates – Winter 2014
- Data Protocols – Spring 2013
- Professional Dev. Field Tests, including ELPA21 Scoring Certification Course – Spring 2014
- Final Professional Development Materials – Fall 2015
- Media Package – Summer 2015

Each of these critical deliverables will be managed on its own timeline with its own sub-deliverables and discrete tasks. This timeline allows SEAS and LEAs within the consortium to do either a pilot or a wholesale adoption in the second half of year four of the assessments, scoring components, assessment item bank, reports, and professional development program developed by the consortium as a result of this project. This also allows project leadership to serve in an advisory role through the year four, to facilitate ease of implementation. A more detailed timeline follows on the next two pages.

English Language Proficiency Assessment System																
Timetable of Major Deliverables																
MAJOR DELIVERABLE	SY 2012-13				SY 2013-14				SY 2014-15				SY 2015-16			
	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
Proficiency Assessment Standards (TMT)																
Amass EL Proficiency standards																
Compile proposed multi-state list																
Submit for public comment																
Finalize list of multi-state standards																
Consortium states adopt standards																
Item Acquisition and Development (TMT)																
Establish item-sharing agreements																
Convene content advisory panels																
Specify item requirements																
Inventory submitted items; gap analysis																
Develop necessary new items																
Finalize and hand off item bank																
Technology Utilization (TMT)																

Develop secure hosting platform																			
Collect incoming item data																			
Ensure correct APIP-tagging																			
Specify data management protocols																			
Data and Reporting (TMT)																			
Compile reporting best practices																			
Survey audiences and define needs																			
Specify reporting requirements																			
Develop report templates																			
Refine reporting with field test feedback																			
Create RFPs for reporting platforms																			
Assessment Design (TMT)																			
Develop assessment blueprints																			
Construct field test forms																			
Analyze field test data																			
MAJOR DELIVERABLE	SY 2012-13				SY 2013-14				SY 2014-15				SY 2015-16						
	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3			
<i>Assessment Design (TMT), continued</i>																			
Conduct item analysis and cull																			
Refine and vertically scale items																			
Produce final forms																			
Field Testing (TMT)																			
Establish field test parameters																			
Secure field test sites																			
Conduct field test																			
Compile field test data																			
Accommodation and Accessibility (TMT)																			
QC review items under development																			
Review field test data																			
Advise on final forms																			
Advise on PD materials																			
Standard Setting (TMT)																			
Set Benchmark Standards																			

development cycles and foresee any derailments. This foresight allows the project manager to shift resources and apply contingency plans when necessary to maintain the timeline.

The ELPA21 consortium has consciously refrained from naming vendors for key functions until such time as a thoughtful, competitive RFP process can be put in place, following notice of grant award. The use of vendors for key tasks, such as Item Acquisition and Development and Field Testing, will be critical to project success. The most demanding day-to-day responsibilities in the project rest on these contracts. Vendors bring the necessary technical and professional competence to bear and are accustomed to managing projects, reporting to constituents, documenting specifications, and communicating to all stakeholders. They are effective custodians of the timeline – and their financial compensations is tied to on-time, budget-or-better delivery. This places responsibility for the timeline (as well as for meeting specification and budget) in the hands of those entities most experienced in managing it.

Risks also exist around **resources**. Resources can be human—the often overtaxed and in high-demand SEA leadership that will make up the Consortium Council and Executive Board. Budget and space are also resources. The best way to manage resources is constant monitoring. Budgetary resources can be monitored by monthly invoice reconciling, forecasting, and cash flow management. Human resources are best managed through communication, accountability, and creativity. Planning the review of major deliverables to coincide with in-person meetings is one effective technique—travel costs are conserved and the members’ availability is maximized. Triggers will be defined in advance, so that, when a resource meets a certain level of constraint, a contingency plan goes into effect.

A third area of risk concern is **quality**. Any project on an accelerated timeline or restricted budget runs the risk of quality impairment. Constraints in timeline can lead to poor specification gathering or problematic execution. Constraints of budget can create a lack of viable vendor options, leaving a project to choose among second-rate or inexperienced vendors. A well-managed project, even one on a restricted budget, can still be managed in such a way as to maintain multiple, viable vendor choices. Work can be re-bundled, budgets can be shifted to higher priority targets while reducing scope in less critical components, and timelines can be extended to allow for higher quality to prevail.

3. Extent to which the budget is adequate to support the development of assessments

ELPA21 has carefully designed its work tasks to provide maximum output to consortium states at a minimum cost in project dollars. Grant funds will be used to develop screener and summative ELP assessment forms sufficient for states to operate their Title III assessment programs for two years following the conclusion of the grant. To do this successfully and within funds available, consortium states will need to harvest and contribute to the ELPA21 project validated ELP assessment items complete with UDL and AIP meta-tags as well as requisite psychometric data. These harvested items will be mapped against the ELP assessment standards and test blueprints, and where gaps exist, new items will be commissioned. Consortium states will likewise leverage materials and protocols already existing and previously vetted to construct the necessary professional development modules around test administration and classroom formative assessment strategies and techniques.

ELPA21 states will continue to advertise and competitively secure test delivery contracts, paying for those as they always have from state/local funding. Thus, the project's costs are lessened (by comparison with the large assessment consortia (Smarter-Balanced and PARCC) with states continuing to pay test delivery costs. Consortium states costs are also lessened (by comparison with normal state test program operations) when ELPA21 absorbs the assessment development costs. We will ensure funding is sufficient to reach our consortium targets for providing screeners, an interim test item bank, a summative test, and materials and protocols to support professional development sessions locally delivered by SEA and/or LEA personnel. Additionally, ELPA21 will work with the Assessment Solutions Group (ASG) for their services to develop cost models for current state ELP assessment programs, ELPA21 state ELP assessment programs, and to help the consortium complete final negotiations with vendors for project deliverables at the lowest cost possible.

4. Estimated costs for each consortium state and how it will be sustained over time

The ELPA21 delivery model consolidates and leverages development costs while leaving delivery costs to the member states. During the ELPA21 development years, member states will continue to deliver their current ELP assessments. When the ELPA 21 products are completed and validated,

member states—individually or in groups or even as an intact consortium—will be able to engage the vendor community in competitive RFP processes to select the best option for their needs. Thus, states maintain responsibility for and control of their own assessment programs.

ELPA21 is taking focused, proactive steps to arm states with empirical data about ELP assessment cost projections. ASG will gather current fiscal data on states' ELP assessment costs and then use their highly accurate proprietary software to analyze current practice and generate future cost models. The consortium members all want robust discussion and wise decisions about the future of the ELPA21 work beyond USED funding. There are key issues to be resolved around the benefits of maintaining the consortium structure, and if so, governance, maintenance, and funding in the longer term. Future governance policy options will be formulated by the Executive Board and then vetted and passed on to the member states by the Consortium Council in democratic debate. With this key information, states will be well positioned to make the best individual and group decisions about future governance structures, working relationships, and assessment maintenance.

5. Quality and commitment of personnel

Project Director - Doug Kosty - Mr. Kosty is Assistant Superintendent in the Oregon Department of Education (ODE) responsible for assessment, accountability, and data functions. Prior to joining ODE he was an operations systems consultant with Deloitte. He brings significant large scale project development and implementation experience, including guiding the launch of the country's first online adaptive state assessment program.

Principal Investigator - Kenji Hakuta - Dr. Hakuta is Director of the Understanding Language Initiative housed at Stanford University. He is an internationally recognized expert in language acquisition and one of the lead contributors to the CCSSO commissioned document, *Framework for the Creation and Evaluation of ELP Standards Corresponding to the CCSS and Next Generation Science Standards*.

Project Management Partner - Council of Chief State School Officers - ODE has identified the Council of Chief State School Officers (CCSSO) as their intended General Project management partner.

Founded in 1927, CCSSO is a nonpartisan, nationwide, nonprofit organization of the public officials who head departments of elementary and secondary education throughout the U.S. and beyond. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. CCSSO seeks its members' consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public. Through its structure of standing and special committees, CCSSO responds to a broad range of concerns about education and provides leadership and technical assistance on major educational issues.

CCSSO is qualified to assume the role of Project Management Partner for ELPA21. CCSSO has a proven track record of managing Enhanced Assessment Grant (EAG) projects, serving as the primary management partner for six EAGs. CCSSO along with member states has successfully developed and managed an operational English Language Acquisition assessment through its English Language Development Assessment (ELDA) EAG. ELDA is a battery of tests designed to allow schools to measure annual progress of non-native English speaking students in grades 3-12 in acquiring ELP skills.

CCSSO has a long-standing history of supporting and facilitating collaborative work that empowers state education leaders to come together to address common challenges, share resources, and produce work that ultimately supports all states as they strive to improve the educational outcomes of all children. A majority of CCSSO consortia/collaboratives bring states, thought leaders, and funders together in long-term, multi-year work. These state consortia encompass content-specific, subgroup, and technically relevant interests in areas such as accountability systems, assessment for LEP students, teacher and leadership standards development, comprehensive assessment system for ESEA Title I, and extended learning opportunities.

The ELPA21 scope of work will connect with the Council's established on-going state collaborative efforts to include the college- and career- ready standards, Implementing the Common Core Standards (ICCS), the State Collaborative on Assessment and Student Standards (SCASS) system, and access to all member chief state school officers. These established connections create a ready mechanism for

conducting and disseminating the work; providing administrative, meeting, and logistical support; and ultimately helping to inform a larger transformational agenda in education policy and practice.

- **CCSSO SCASS Director - Robert M. (Bob) Olsen** - Mr. Olsen directs the State Collaboratives on Assessment and Student Standards (SCASS) system at CCSSO, of which the ELL SCASS is one group. Prior to joining CCSSO he served in three different LEAs as director of testing and evaluation and was an associate research professor in the Oregon University System.
- **CCSSO SCASS Budget Manager - Adam Petermann** - Mr. Petermann oversees the fiscal aspects of CCSSO's involvement in cooperative projects operating as the liaison between ELPA21, CCSSO, and ODE's fiscal interactions. He will assist and coordinate contractual negotiations with project subcontractors.
- **ELP-EAG Project Manager: Cathryn (Cat) Still** -- Ms. Still is a PMI-trained project manager with over 15 years of experience shepherding education and assessment initiatives to fruition. Prior to joining CCSSO, Ms. Still managed multi-year contracts with SEAs and LEAs for an assessment development and testing platform vendor.

TMT Advisor/Leads: TMTs are lead by contracted experts in each area. TMTs are not work teams. Their role in the project success is oversight of the competitively selected vendor who will undertake and complete the work assignments necessary to deliver the products and/or protocols. For example, the Field Testing TMT, working with the Assessment Design TMT will develop the criteria for the field test, arrange the release of the RFP by the Executive Board, vet proposals and recommend the most effective and cost efficient vendor, and then oversee and monitor that vendor's work to successful completion, resulting in sufficient data collected from a valid sample, in secure, consistently administered settings, finally producing full spectrum item statistics sufficient to support item selection for screener and summative test construction.

The project has identified ten (10) work components for TMT action. Those TMTs are shown in the table following. A short description of the key deliverable products/protocols for each is noted along

with the consultant arranged to advise/lead the group. Advisor/leader resumes are included in this proposal.

Task Management Team	Key Deliverable Products/Protocols	Advisor/Lead
ELP Assessment Standards Development/ Implementation	Implementation of ELP Assessment Standards corresponding with college and career-ready common core standards	Dr. Kenji Hakuta Stanford University
Item Acquisition and Development	Tagged bank of items [a] acquired from member states' existing assessment collections and [b] newly written.	Steve Marban Latin American Educational Services
Assessment Design	The specifications for assessment items including meta-data tags, test blueprints, administration delivery and timing, validation procedures and thresholds, reporting criteria, security and release.	Dr. Tom Haladyna Professor Emeritus, Arizona State University
Accommodations and Accessibility	ELPA21 assessments accessible by all students; UDL and APIP tags ensure accommodated delivery to special needs students.	Dr. Martha Thurlow National Center on Educational Outcomes
ELP Performance Standard Setting	Multi-state benchmark proficiency standards definitions and values.	Dr. Bill Auty, Ed. Measurement Consulting
Field Testing	Validly and reliably gathered and analyzed, full spectrum item statistics sufficient to support item selection for screener and summative test construction.	Dr. Scott Elliot SEG Measurement
Technology	ELPA21 test delivery and reporting which is fully	Dr. Scott Elliot

Integration/Utilization	integrable with and leverages the technology specifications of Smarter and PARCC.	SEG Measurement
Data System and Reporting	A secure, state-accessible, bank of ELPA21 screener and summative test forms with meta-tags for online delivery, central repository of student performance data with data architecture specifications.	Dr. Jay Pfeiffer Consultant Former Deputy Chief Florida Dept. of Education
Professional Development Support	Materials and protocols to support member states' professional development activities to ensure valid and reliable assessment administration.	Dr. Charlene Rivera George Washington University
Communications and Outreach	A communications plan to inform various stakeholder audiences about the impacts of ELPA21.	Kara Schlosser Consultant Former Communications Director, CCSSO

DESCRIPTION OF HOW PROPOSED PROJECT ADDRESSES THE ABSOLUTE PRIORITIES

Absolute Priority 1: Collaborations. Collaborating with institutions of higher education, other research institutions, or other organizations to improve the quality, validity, and reliability of state academic assessments beyond the requirements for such assessments described in section 1111(b)(3) of the ESEA.

The consortium’s leadership assets include national experts from high ranking institutions of higher education and nationally known research organizations in English language acquisition, in assessment and accommodations of LEP students along with experts from the ELPA21 states. These organizations include Stanford University's Understanding Language Initiative; CCSSO; the National Center for Research on Evaluation, Standards and Student Testing (CRESST) at UCLA; the National Center on

Educational Outcomes (NCEO); and thirteen states. This expert leadership directly addresses issues concerning the content and psychometric quality of the assessment system.

Absolute Priority 2: Use of Multiple Measures of Student Academic Achievement. Measuring student academic achievement using multiple measure of student academic achievement from multiple sources.

The assessment system that we are proposing entails use of multiple measures, in the four language domains, of students' English language proficiency—collected at various times and with different instruments, among them a diagnostic/screener test, individual teachers' formative assessments, interim benchmark assessments, and an annual summative assessment. These multiple measures provide information and data that can be used to identify students as ELLs, impact decisions about whether a student should exit from ELL instructional programs, and inform determinations of school, LEA, and SEA effectiveness for the purposes of accountability. And at the consortium level, the results of multiple-measures will be collected allowing for research to be conducted across states with substantially larger pools of students than are typically available when states operate their assessment systems independently.

Absolute Priority 3: Charting Student Progress Over Time.

The psychometric scale and performance standards will be a strong basis for charting students' longitudinal progress. The data from this grant will be available for states' and researchers' use.

Absolute Priority 4: Comprehensive Academic Assessment Instruments. Evaluating student academic achievement through the development of comprehensive academic assessment instruments such as performance and technology-based academic assessments.

ELP standards and assessments have always had alignment to the academic content standards. This has been a requirement of ESEA for ELP assessments. However, with the adoption of the common core state standards, the ground has shifted considerably with respect to the content-specific language demands featured in the new standards. As analysis by the Understanding Language Initiative of the language underlying the common core state standards in English language arts and mathematics, as well as the Next Generation Science Standards makes amply clear, the new standards give pronounced attention to the language demands contained therein. A similar conclusion was reached by the team of experts who

developed the *Framework for the Creation and Evaluation of English Language Proficiency Standards*. This analysis explicitly delineates the language demands separately for each of the academic content areas.

The result of these efforts to establish correspondence between the content standards and ELP standards reveals the multiple ways in which language is part of academic content, and therefore an ELP assessment that is based on standards that correspond to academic assessments will necessarily evaluate student academic assessment. Naturally, there are aspects of academic content that are not covered by language; however, the new standards place an unprecedented emphasis on language as it is used in the context of the content. We will not be generating an academic achievement score per se. Rather we will be generating a language proficiency score which is likely to be predictive of (or at least correlate with) students' linguistic readiness to achieve the standards.

Absolute Priority 5: Developing an English language proficiency assessment system.

a. Design Principles

1. ELPA21 which will be developed under this project will be based on states' current products and protocols, and designed in collaboration with and for implementation in 13 states.
2. A common definition of English learner will be defined, clarified, and adopted by all consortium states. Among the initial tasks will be the adoption of common performance level descriptors for classifying students as English learners for purposes of eligibility for targeted services. These descriptors will subsequently be utilized in the process of setting performance standards for levels of proficiency. The ELP Performance Standards TMT will ensure this is completed successfully.
3. ELPA21 will include a diagnostic screener and a summative assessment. A strong, comprehensive assessment system, will also include formative and interim components. However, due to resource limitations, we are proposing to develop only the diagnostic screener and summative assessment through the funding for this grant. Consortium member states will

contribute, without cost, existing materials, artifacts, and protocols and so add to the components and functioning of a complete assessment system. Additionally, the ELPA21 consortium will pursue grant money from other sources to support and supplement specific refinements and enhancements to the “donated” resources to complete development of a comprehensive assessment system.

4. The system will measure students’ English performance against common ELP assessment standards. CCSSO has developed a Framework to support states as they strive for correspondence between ELP assessment standards and standards for college- and career-readiness, as defined in the CCSS. This Framework will serve as a guide as the participating states work to establish an agreed-upon set of ELP standards and common ELP assessment standards that will guide test development.
5. The assessment will measure students against ELP standards that correspond to a common set of college- and career-ready standards.
6. The assessment will cover the full range of ELP standards across the four domains of reading, writing, speaking, and listening. The CCSSO Framework will guide development of the full range of ELP standards.
7. The assessment will consider the students’ control over the linguistic components of language.
8. The assessment will produce results that indicate whether individual students are ready for instruction in English and whether they are meeting college-career readiness standards. These benchmark decisions will be pivotal in the process employed to set standards for English proficiency levels.
9. The system will provide an annual measure of English proficiency as well as progress in English proficiency for each English learner in grades k-12 in each of the four domains based on vertical scaling techniques.
10. The system will accommodate participation of all English learners, including those with disabilities and those with a limited formal education. We understand that the students with the

most significant cognitive disabilities may be placed in an alternate assessment based on alternate achievement standards.

11. The system will be accessible to all English learners by taking full advantage of the accommodations that are available through a computer-based assessment delivery model.

b. Technical quality:

The assessment system will measure English proficiency in ways that:

1. Are consistent with nationally recognized professional and technical standards. Three of the sources we will rely on are *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999); the unified approach to validity (Cronbach & Meehl, 1955; Kane, 2006a, 2006b; Messick, 1989); and *The Handbook of Test Development* (Downing and Haladyna, 2006).
2. Elicit complex student demonstrations of comprehension and production of academic English through the use of a variety of item types, including constructed response designed to elicit complex student responses.

c. Data

1. The assessment system will produce data that includes student attainment of English proficiency and student progress in learning English and that can be disaggregated by subgroup.
2. The system will provide a valid and reliable measure of students' abilities in each of the four language domains and a comprehensive English proficiency score based on all four.
3. The data can be used for (i) identification of students as ELLs, (ii) decisions about whether a student is ready to exit the English language instruction program, and (iii) determinations of schools, LEA, and state effectiveness for accountability purposes.
4. Data can be provided to consortium states for use as one of multiple measures included in the state's own system of educator effectiveness to inform (i) evaluation of individual principals and teachers to determine effectiveness, (ii) to determine professional development and

support needs, and (iii) to improve teaching, learning, and language instruction education programs.

d. Compatibility

The assessment system will use compatible approaches to technology, assessment administration, scoring, reporting, and other factors that facilitate the coherent inclusion of the assessments within states' assessment systems.

Rather than developing a completely different set of applications and resources to support ELPA21, this grant will leverage the work of the assessment consortia to the extent practicable. The Consortium has established within its system architecture a multi-Tenancy approach such that additional assessments beyond mathematics and ELA may be supported. For example, ELPA21 will leverage the Smarter item bank/authoring tool, test administration platform, and reporting systems necessary to deliver an ELP assessment. In addition to saving scarce development dollars, collaboration like this will provide states with the option of using a single set of tools for delivering the federally required summative assessments. This grant will be used to conduct a gap analysis and to make enhancements such as to the applications user guides as necessary to support the additional uses required by the ELP Assessment. In addition, monies will be allocated for hosting and help-desk support that are required to complete the field tests necessary for the design of ELPA21.

e. Students with the most significant cognitive disabilities

ELPA21 consortium states will be critical for ensuring that accessibility features and accommodations policies are an aspect of every phase of the development and implementation process.

Accessibility for all students who take the ELPA21 is a cornerstone of ensuring that the assessment is valid, reliable, and supports appropriate inferences. Accessibility has been defined in a variety of ways. In the initial notion of universally designed assessments, the kernel of the concept emerged in the definition that the assessment was designed and developed from the beginning to allow participation of the widest range of students, and that would result in valid inferences about student performance (Thompson & Thurlow, 2002; Thompson, Thurlow, & Malouf, 2004). The definitions and early work of

researchers nearly always mentioned all students, including English learners, yet invariably focused on students with disabilities.

Accommodations are changes in the materials and procedures of an assessment that provide the student with a way to show knowledge and skills rather than the effects of their limited English. In 2007, Albus and Thurlow examined state accommodation policies for English learners with disabilities taking ELP assessments. They found that states varied considerably in their policies, and that the accommodation allowed by most states was the “repeat, reread, and clarify directions” accommodation.

With the development of common assessments, there is a need for states to weigh in on the development of the policies for the new ELP assessment (see NCEO, 2011a). It will be helpful to develop a set of principles to guide the development of accommodations policies and practice, similar to what has been created by the National Center on Educational Outcomes (Thurlow et al., 2008) and by the National Accessible Reading Assessment Projects (Thurlow et al., 2009).

Implications for ELPA21. States working through Task Management Teams will be critical for ensuring that ELPA21 accessibility features and accommodations policies are an aspect of every phase of the development and implementation process. The Accessibility and Accommodations TMT, with NCEO observing, monitoring, and supporting its work, will be at the table as a critical friend during all aspects of the project. NCEO will synthesize and report to the TAC on observations and recommendation for accessibility and accommodations.

DESCRIPTION OF HOW PROPOSED PROJECT ADDRESSES

COMPETITIVE PREFERENCE PRIORITY 1

Sixteen state education agencies have expressed active commitment to joining with their colleagues to form the ELPA21 consortium.

Of these states, MOUs have been enacted and received from AR, CA, FL, IA, KS, NE, SC, WA, and WV. Lacking sufficient time to yet fully enact MOUs through their state’s required processes, six

additional states have submitted indicated interest and support. These states include: CT, LA, MI, NY, OH, and TN.

These states have participated in the development of the governance model. They have identified a project management partner and collaborated with that partner in developing this proposal.

Because it is both evidence of collaborative commitment and of a strong and efficient management structure, we devote significant text to the consortium's structure and operation including methods and processes for decision-making, protocols, and the process and timeline by which the consortium will operate including the consortium's plan for managing grant funds received.

Each participating state (please see the memorandum of understanding included in this submission) agrees to adopt the recommendations of the consortium at the conclusion of the project. ELPA21 states will adhere to the governance structure of the consortium, participate in the development and decision-making processes surrounding the development of the item bank and assessments, and will agree to follow established timelines. These states will adopt the common set of English Language Proficiency assessment standards, which will be CCSS aligned, that are identified as the foundation of the new English Language Proficiency assessments. Consortium states will effect a statewide implementation of the consortium summative assessment in grades k-12 at the conclusion of the assessment development grant with the first full year of Consortium-wide implementation planned for the 2016-17 school year.

The consortium will create and deliver a complete and inclusive system of diagnostic/screener and summative assessments, consisting of a variety of item types strategically aligned with the mission of this grant around English Language Proficiency. These assessments will have a scope sufficient to assess the full range of foundational CCSS around the four language domains of English Language Proficiency (reading, writing, speaking and listening) and to identify mastery in linguistics. The finished outputs of the consortium will be as follows:

- A diagnostic screener at each grade band for initial identification and placement of new ELLs.
- Two linear forms of a summative assessment at each grade band that provides accurate performance information as described above for all ELLs, including students with disabilities.

- ELP performance standards and performance level descriptors benchmarked across the consortium.
- Field test scoring for individual students and groups that is fair, reliable, and statistically valid. This scoring will be the basis for future evaluation of growth at the student level and aggregate growth at the classroom, school, and district level.
- Secure item banks underpinned by psychometrically sound scaling and equating procedures to provide comparable scoring across consortium states.
- A data reporting architecture that will articulate benchmarks and goals to audiences consisting of students, families, teachers, principals, LEAs, and SEAs and develop and increase understanding of student progress toward ELP.
- Professional development protocols and materials focused on the examination and scoring of students' work in alignment with the assessment system, to guide and positively impact curriculum and lesson development.
- A phased communication plan and supporting materials across multiple media, to guide SEAs and LEAs in informing their various constituencies about the progress, implementation, and outcomes of the new assessment system.
- A timeline and resources for adoption and implementation at the state and district level, including boilerplate contracts and Request for Proposals (RFP)s to assist states in the selecting of and contracting with necessary vendors and consultants to launch and execute the operational ELPA21 in 2016-2017.

The consortium will be structured as to allow for the identification, collection, codification, and dissemination of best practices from member states; it will also provide a platform to incorporate input from state administrators, policymakers, school practitioners, and technical advisors to ensure an optimal balance of assessment quality, efficiency of development and implementation, and

maximum possible conservation of costs and time. The organizational structure of the consortium includes the following and is discussed in depth below:

- Governance – Executive Board
- Consortium Council
- Task Management Teams (TMTs)
- Project Management Partner
- Technical Advisory Council (TAC)
- Third-party Evaluation Partner

Governance – Executive Board: Thought and Policy Leadership

Governance of the ELPA21 consortia will be vested in a seven-member Executive Board comprised of a Chairperson, the Project Director from the Lead State (Oregon), the Deputy Superintendent of Public Instruction lead of the 2011 EAG Lead State (California), and four at-large representatives from state membership. The five non-lead state members, in addition to the state leads (2011 (CA) and 2012 (OR)) will be selected by vote of the Consortium Council members; the Chairperson will be one of these seven delegates, elected by vote of the Executive Board members. The 2012 lead state (OR) will facilitate the Executive Board nominations so that the Board includes states with small and large EL populations, and active membership in all consortia (PARCC, Smarter, DLM and NCSC). Additional, ex-officio Executive Board members will include the Principal Investigator, up to three principles designated from the Project Management Partner, the Third-Party Evaluator, and the Technical Advisory Committee, as deemed necessary by the Executive Board. The Executive Board will coordinate policy formulation with the four current CCSS assessment consortia: PARCC, Smarter-Balanced, and the two 1% Consortia, National Center and State Collaborative (NCSC) and Dynamic Learning Maps (DLM).

The Executive Board will meet twice monthly for the duration of the project. Responsibilities specific to the Executive Board include oversight of the English Language Proficiency assessment program's development, the project's expenditure of funds, and the Project Management Partner, as well

as other ex-officio members. Using a collaborative approach and a consensus voting process, the Executive Board will develop project plans and agendas and act on reports and recommendations from project support members, such as the USED, CRESST, TAC and consortium states. The Executive Board will also serve as the final voice and decision-making entity on all issues and decisions resulting from the Consortium Council.

The Executive Board will benefit from extensive interaction with and input from the Understanding Language Initiative, at Stanford University's School of Education, the mission of which is to "heighten awareness of the language and literacy issues embedded within the new Standards." Dr. Kenji Hakuta, co-chair, is one of the nation's leading authorities on bilingual education and English language acquisition by immigrant students; he has constructed the Theory of Action that will underpin all of the consortium's efforts and outputs.

Consortium Council

One member from each consortium state will make up the Consortium Council. These may be the SEA's chief state school officer or a designee. Members must have prior experience in one of the following areas: curriculum or instructional supports to ELLs, assessment system policies, or assessment implementation. Consortium Council members will meet twice monthly and will serve as the liaison between the TMTs and the overall consortium activities.

The Consortium Council will bear the responsibility for determining the general scope of the proposed assessment system and the review of the recommendations from the TMTs prior to submission for approval to the Executive Board and Lead State (Oregon). The Council will receive regular reports from the TMTs, Project Management Partner and various advisors throughout the project. In collaboration with the Lead State, the Consortium Council will also be the initial point of contact for the expenditure of funds within the project.

Decision making within the Executive Board and Consortium Council will be conducted with a goal of consensus on all decisions. Each Consortium Council member will have one vote. Any vote of the Consortium Council members that is split by three votes or fewer (voting outcomes of six-to-nine or

seven-to-eight) will be referred back to the Consortium Council for further discussion and elaboration. The Executive Board may, in these cases, prepare and provide additional information to the Consortium Council to aid in these decisions.

Task Management Teams

In order to provide a mechanism for the creation and dissemination of the best ideas and practices of the consortium states around the development of the comprehensive ELPA21, a Task Management Team (TMT) structure will form the foundation of the project. Each TMT will consist of 2-4 State Education Agency (SEA) members from the ELPA21 consortium states. These TMTs will be led by an Advisor-Expert with industry-leading experience in the specific area of responsibility charged to each TMT. The Project Management Partner, CCSSO, has identified an Advisor-Expert to lead each TMT and direct their outputs; they are listed, in the table beginning on page 41.

State representatives to the TMTs may also serve as members of the Consortium Council or the Executive Board, or may be other SEA employees with expertise and skills in the target area. All consortium states are expected to commit support toward one or more of the TMTs, based on the skills, expertise, and interest within the state, in order to maximize contributions and distribute expertise and responsibilities effectively. TMT membership will be appointed by the Consortium Council, taking into account the recommendation of the Project Managing Partner and each TMT's Advisor-Expert. To the extent practicable the scope of RFPs will encompass the tasks of two or more TMTs. Each TMT (or joint TMT) will create specifications around its responsibility area and will competitively select, through an RFP process, contractor(s) to complete the TMT's identified tasks. Each TMT will manage and hold accountable its respective contractors, and will be responsible for all communication around its mission-specific progress and outputs. State participation in a TMT will require a minimal amount of engagement to review task progress and materials developed to support project outcomes. ELPA21 has identified the following list of task areas around which TMTs will be built:

- Proficiency Assessment Standards Development
- Item Acquisition and Development

- Technology Utilization
- Data Systems and Reporting
- Assessment Design
- Field Testing
- Accommodations and Accessibility
- Standards Setting
- Professional Development Support
- Communications and Outreach

The TMT for **ELP Assessment Standards Development and Implementation** is charged with recommending a comprehensive set of multi-state standards that will inform the development of the ELPA21 assessment across the four domains of reading, writing, listening and speaking. In cooperation with a pre-existing team of language acquisition experts who are consortium states and leading institutes of higher learning, this TMT will conduct a series of meetings both to identify those multi-state standards critical to building English Language Proficiency and to specify processes and sub-tasks for the amassing and refining of a final compiled list of multi-state assessment standards.

These standards will be submitted for public comment, which the working group will review and weigh, refining toward a complete list of proposed multi-state standards, which will be submitted to the Executive Board for approval. The TMT will then move forward with the approved set of standards for adoption by the consortium states and hand off to the TMTs for Item Acquisition and Development, Assessment Design, and Accommodations and Accessibility for development of an item bank and assessments. The team will ultimately identify and adopt a set of standards for English language proficiency development assessment that correspond to the CCSS in English language arts and mathematics. The standards will be adopted in sufficient time as to inform the development of the assessment.

The soon-to-be-released Framework for the Creation and Evaluation of English Language Proficiency Standards Corresponding to the Common Core Standards and the Next Generation Science Standards from CCSSO will be used to identify and determine the standards. The Framework contains a specific protocol to determine correspondence that will be followed. The standards will be drawn from the products of the following set of activities: (1) a multi-state analysis of how the English Language Proficiency standards in 16 states (including most of the states in ELPA21 consortium) correspond to the CCSS work conducted by the Assessment and Accountability Comprehensive Center and the MidAtlantic Comprehensive Center on behalf of the ELL SCASS at CCSSO; (2) the California State Department of Education revision of its English Language Development Standards, which has made significant progress on the development of standards that correspond to the Common Core and is scheduled for completion in October, 2012; (3) a revision of the WIDA Standards that utilizes the CCSSO Framework to determine the correspondence of the WIDA Standards to the College and Career-Ready Standards; and (4) any efforts that may be separately funded to develop a set of ELP assessment standards that build upon the Framework.

Kenji Hakuta and Martha Castellon of Stanford University will advise this TMT. Several individual ex-officio expert consultants may be invited, Edynn Sato (WestEd), Gary Cook and Marianna Castro (UW-Madison/WIDA), and Charlene Rivera (GWU-CEEE). The TMT will consider all of the possible components of ELP standards made available from the above sources, and will gain consensus among member states on standards.

The TMT for **Item Acquisition and Development** is charged with the identification and refinement of the assessment items that will be used to build the screeners and summative assessments, and populate the assessment item bank which users will access to construct such benchmark assessments as they chose to employ.

This TMT will kick-off with the identification of a working team of language acquisition experts and SEA representatives who, over the course of a series of meetings, will develop ELPA21 item requirements and characteristics and communicate those descriptors, as well as identified multi-state

proficiency assessment standards, to consortium states. Item acquisition and development will originate with both the execution of cooperative agreements with consortium states, which will agree to share their existing item banks, and the specification and competitive selection of an item development contractor.

The Item Development contractor will harvest assessment items and their performance statistics from state item banks. Under the guidance of the TMT, the contractor will secure reviewers and writers and establish item review teams who will evaluate assessment items based on criteria including, but not limited to, linguistic complexity, content demand, cognitive demand, and accessibility. These review teams will also conduct bias and fairness reviews and a gap analysis to identify any standards or skill areas that require additional assessment items. The contractor will hire and train and manage item writers, and will work with the Technology Utilization and the Accommodations and Accessibility TMTs to ensure that harvested and newly written items are UDL and AIP meta-data compliant, and are developed to be accessible to students requiring accommodations.

The Item Development contractor(s) shall also identify CCSS alignments of each item, and will provide mapping back to corresponding state standards. The expertise of the TMT will guide and inform the entirety of the item identification and development processes. The process of reviewing and refinement of the item bank will recur until the TMT has collected a sufficient number of items to comprise a multi-state item bank in support of the new assessments. The final deliverable of the Item Acquisition and Development TMT will be an item bank, complete with metatags, which will be handed off to the Assessment Development TMT for construction into field test forms for screeners and summative assessments, and ultimately, final test forms.

The TMT for Item Development and Acquisition will be helmed by Steve Marban, Steve Marban, the advisor for this TMT brings large-scale, multi-year item and assessment development expertise to ELPA21 through his project working with the Mexican ministry of education on the two essential skills of the Twenty-first Century: English and Computing Management. His project there is assessing the progress in English-language proficiency among 64,000 public school students and 630 teachers using custom instruments developed for the project.

Final forms and remaining items will be made accessible on a secure item and assessment hosting platform, which will be a major output of the **Technology Utilization TMT**. This team will kick off by soliciting and compiling specifications on Consortium states' data management systems and reporting platforms. These compiled specifications will serve two purposes: to inform ease of use and access in the design of the item and assessment platform, and to ensure compatibility of outbound data, meta tags and other test statistics into the reporting platforms of the consortium SEAs and LEAs. Recommendations of the TMTs for Accommodation and Accessibility, Field Testing, and Data and Reporting will also contribute to the specification and development of this platform.

The Technology Utilization TMT will also oversee the collection of the incoming data itself – items and their associated statistics. As the final list of multi-state standards evolves, the Technology Utilization TMT will ensure that all items are UDL- and AIP-compliant and properly tagged. In this capacity, the Technology Utilization TMT will act as a quality control partner to the TMT for Item Acquisition and Development.

Lastly, this TMT will collaborate with ELPA21 consortium SEA data system teams to ensure seamless inbound and outbound integration, transparent access and ease of coordination with their existing Student Information Systems and assessment reporting platforms. Data structure components will be informed by the data systems evolving from the work of the four CCSS-focused Consortia: PARCC, Smarter Balanced, the National Center and State Collaborative Partnership (NCSC) and the Dynamic Learning Maps Alternate Assessment System Consortium (DLM).

Dr. Scott Elliot will serve as Technology advisor to this TMT. Dr. Elliot has considerable experience with educational assessment technology dating to the first installation of a computer-based state assessment program in the US. He is currently working with vendors developing the SBAC and PARCC assessments, and will provide a vital and robust linkage to their related work.

The reporting, data management and security specifications amassed by the Technology Utilization TMT will be shared with the **Data and Reporting TMT**, whose efforts will inform assessment data reports to serve the data needs of assorted audiences. The TMT for Data and Reporting will retain and

manage a contractor or consultant with expertise in assessment data management to survey policy experts, SEAs, LEAs, teachers, parents, and students from ELPA21 states to determine both universal reporting requirements and those specific to just one or a handful of key audiences.

This TMT will then oversee the specification and development of report templates that will compile and display the data elements required by each audience in user-friendly formats. A final output of this TMT will be a set of turn-key RFPS that SEAs and LEAs can use to solicit contractors to build or augment reporting platforms at the district and state level. Jay Pfeiffer, former Deputy Commissioner in the Florida Department of Education and lead architect of that state's seminal data system, will advise this TMT.

Assessment Design will be managed by a TMT consisting of consortium state members with appropriate expertise to act as an assessment design and construction review panel. In collaboration with the TMTs for Item Acquisition and Development, Accommodations and Accessibility, Field Testing, and Technology Utilization, the TMT for Assessment Design will oversee the development of blueprints for each assessment instrument that address domains and grade band. Using the secure item bank developed by the Item Acquisition and Development TMT, the Assessment Design TMT will construct assessment forms for field testing. Upon conclusion of field testing, the Assessment Design Contractor(s) will conduct item analyses to identify and cull poorly performing items while vertically scaling items that performed as expected. This final set of assessment items will be deployed into one (1) screener and two (2) summative test forms per domain and grade band, and will incorporate placeholders for future field test assessment items. All test forms and meta-data will be loaded into the secure item bank, as will any identified interim assessment items.

Tom Haladyna, Professor Emeritus, Arizona State University, will advise the Assessment Design TMT. Tom is the author of multiple books on item and test design, lead developer for state assessments, as well as a frequent consultant to complex performance test based licensing examinations such as the American Board for Facio-Plastic and Reconstructive Surgery.

Field test parameters will be defined by the **Field Testing** TMT, which will also be responsible for the collection and management of field test data. This TMT will identify and contract with vendor(s) with the expertise to conduct field testing, and will establish LEA field test sites in the consortium states. This TMT and its supporting workgroups will facilitate field testing with states' existing virtual testing platforms and make available paper and pencil test forms where necessary. Dr Scott Elliott will also lead this TMT, to ensure consistency in item data throughout the development, field test and final deployment stages of the project.

The development of each item and the composed assessments will also be informed and guided by the **Accommodations and Accessibility** TMT, operating in a “critical friend” role as a kind of internal audit to ensure assessment items are accessible for all, including students with special needs. This TMT will observe all phases of the project and provide guidance to ensure that established, standard practices around accommodation and accessibility figure prominently in all appropriate aspects of design, development, and deployment. This TMT will also participate in each meeting of the TAC to share observations and make recommendations.

Martha Thurlow, Executive Director for the National Center on Educational Objectives (NCEO) will advise this TMT, lending her expertise of 35 years research and advocacy for equal access to the education enterprise, including valid assessments. Dr. Thurlow has published extensively on all of these topics, authoring numerous books and book chapters, and publishing more than 200 articles and reports. In 2003, she completed her 8-year term as co-Editor of *Exceptional Children*, the research journal of the Council for Exceptional Children.

As item statistics and other field test data are collected, they will be transmitted to the TMT for **Standard Setting**. Working with a user-based standard-setting committee, this TMT will oversee the determination of cut scores to identifying progress and English language proficiency. The TMT for Standard Setting will follow processes for setting benchmark standards, selecting required experts, and overseeing their workflow and outputs. Process requirements will include combining contrasting groups

and applying expert judgment using the Bookmark method to identify benchmark standards, and the determination of weighting for domain (reading, writing, listening and speaking) scores to form a total composite score. This TMT will also review and apply all relevant processes cited in the AERA/APA Joint Standards for Educational Psychological Testing to the data review, benchmarking and score weighting processes.

Dr. Bill Auty, former Assessment and Research Director in the Oregon Department of Education will advise the Field Test TMT. Bill's 35 years experience in all levels of student assessment, program evaluation and data analysis, his leadership building Oregon's state assessment program during its transition to computer-based delivery and his participation with the NAEP standard-setting ideally position him to lead this group.

Professional Development Support will be managed by a TMT that will identify a contractor or contractors to develop and distribute Professional Development materials, curricula, scoring rubrics, protocols, and supporting tools, such as training manuals, communication plans and online videos. Working closely with the Assessment Development TMT and its experts, the Professional Development Support TMT will survey and document the testing process at the LEA and SEA level, including adherence to AERA/APA Joint Standards requirements. This TMT will be responsible for the production of a Professional Development curriculum, which will include the ELPA21 Scoring Certification Course, which will assist in developing continuity across states in the scoring of open-response items.

The proposed Professional Development curriculum will be field tested in six sites, in collaboration with the Field Test TMT. These field test sites will involve, at minimum, four (4) teachers and/or test administrators at each site who agree to pilot the training materials, videos and protocols developed by this TMT. For purposes of fidelity, 25% of the trained participants must administer at least three assessments each.

The Field Test TMT will collect post-training and post-testing participant data from the field test, as well as student performance data. The Professional Development Support TMT will compile participant and observer input to identify areas for improvement in materials, protocols or supporting tools. Based on

these dual analyses, PD materials will be refined, and the final outputs will be produced and packaged in preparation for distribution.

Dr. Scott Elliot will advise the TMT for field test work. Scott's expertise and experience with sampling structures to ensure valid and reliable field test results, his in-depth experience building web-accessible item banks, and his psychometric expertise are all assets to this TMT. Additionally Scott's experience as chief operations officer in an educational testing company will be invaluable to the team's monitoring responsibilities.

The identification of best practices and the creation of new recommendations around messaging will be overseen by the TMT for **Communications and Outreach**. This TMT will direct the solicitation and collection of existing ELP programmatic and test-related communications from ELPA21 states and conduct a gap analysis to identify missing components. In consultation with the Executive Board, this group will manage the identification of key messaging that will be disseminated proactively to essential program participants, partners, and consumers. New messaging materials will be developed or existing materials will be refined to address identified gaps in message or means of delivery. This group will also manage the compilation of a media package, consisting of compiled and newly created materials, which will be used to communicate to SEAs, LEAs, teachers, parents and policy makers about the essential features, uses and values of the English Language Proficiency Assessment Program. Messaging and the final media package will be submitted to the Executive Board for approval prior to dissemination. Third, the TMT for Communications and Outreach will, upon commencement, work with the Technology Utilization TMT to identify and retain a web developer to build a public-facing, informational website to serve as a communications scaffold for project vision, tactical information, findings, and key decisions as they develop. This will help ease the transition and adoption of the new standards set by member-states and will help build consensus and buy-in as the project unfolds. This website would also house a restricted-access component through which consortium members can store and access internal documentation, discussions, and decisions related to the execution of the project.

Kara Schlosser will lead this TMT. She has over 15 years of professional communications experience and previously served as the communications director for CCSSO. A former teacher, she also holds a master's degree in education and is currently consulting with national nonprofits in communications strategy and planning.

Project Management Partner

Project Management expertise will be provided by CCSSO, the Council for Chief State School Officers, whose mission is to “lead and facilitate collective state action to transform our public education system.” CCSSO has established a leadership position in the areas of education legislation and advocacy, workforce, information systems, research, next generation learners, standards, assessment and accountability, all of which dovetail into the mission of ELPA21. Furthermore, CCSSO's role in the CCSS initiative showcases the organizations' depth of familiarity with the details and climate around adoption and implementation of the standards.

The project will be managed using standard, widely accepted and historically successful project management methodology, as described by the Project Management Institute, a worldwide standard-setting and credential-granting organization. CCSSO will not only leverage experience in management of large scale projects, but will also draw upon a wealth of resources and tactics developed while leading similar, SEA-based projects under their State Collaboratives on Assessment and Student Standards (SCASS) initiatives.

The Project Management Partner will sit in oversight on the Planning Phase, which will rely on pre-scheduled and frequent check-ins to assess progress, constant communication to manage scope change and requirement creep, and daily plan management to monitor timelines and control slippage. Stopgaps and flexibility have been built into the timeline, as solid project management acknowledges inevitable realities such as communication delays and member schedule conflicts at start-up.

At the point of vendor selection, each sub-project will move to its own timeline, as described in the Timetable on pages 33-34. Sub-projects will primarily be developed and managed with a Waterfall methodology, which allows one complete phase of development to cascade into the next. So a

deliverables such as the development of field test forms, cannot commence until an earlier deliverable — the creation and approval of assessment blueprints — ends. Some of the sub-task timelines are aggressive, waterfall project development will keep the TMT focused on one specific task area at a time, which tends to facilitate consistent forward motion.

The presentation of critical deliverables for most of the TMTs is timed to occur in October, February or June, which are when CCSSO's ELPA21 consortium holds its meetings. Since the majority of the members of ELPA21 will be involved in the ELP Assessment System Consortium, this plan minimizes time constraints on members by leveraging their presence at these thrice-yearly meetings; this also reduces travel costs. There is sufficient flexibility built into the TMTs timelines and the overall project plan to allow for adjustment while still meeting major deadlines, should a critical deliverable fall behind. There are, however, numerous checkpoints and communication intervals built into the sub-project plans so as to avoid “surprises” and minimize slippage.

Technical Advisory Committee (TAC)

ELPA21 will also establish a TAC. This TAC will operate in the conventional sense, as advisor around assessment psychometrics. Convened by the third party evaluation contractor, CRESST, the TAC will also be charged with a broader mission: to act as a pervasive quality control partner. The Technical Advisory Council will act in an advisory capacity to all of the entities in the project, and will make itself available to inform and guide any of the decisions and processes developed and executed by the TMTs — not only around questions of assessment design, psychometrics and test validity, but also in the arenas of hardware and software requirements and applications as they affect the design and development of the ELP assessment system.

To maintain the organizational linkages in this project, the TAC membership will be recommended by CRESST to the Executive Board who will review, select, and appoint the members.

Third-Party Research and Validation

Third-party research and evaluation will be performed by CRESST, the Center for Research, Evaluation, Standards, and Student Testing at The University of California, Los Angeles. CRESST has led efforts around the improvement of learning and education in the US for over forty years, and is a widely respected authority in scientifically based evaluation and testing techniques.

CRESST will act as a trusted but critical friend to the Consortium, scrutinizing all aspects of the Consortium's decisions about students' assessment data and the conclusions drawn from these data. Specifically, CRESST will review the project's resulting formative and summative reports, and all related processes and outputs, through the lenses of technical quality, wisdom of process, utility, and impact. CRESST will also conduct independent external validation of field test assessment data, for purposes of replicating and corroborating the Consortium's findings and recommendations.