

Design Considerations for the RTTT Assessment Competition Notice Inviting Application (NIA)

Scott Marion, Brian Gong, Marianne Perie, and Chris Domaleski

Center for Assessmentⁱ

April 19, 2010

This document is the Center for Assessment’s initial analysis of the technical aspects of the recently released NIA. For each of the technical sections of the NIA—theory of action, assessment design, assessment development, and research and evaluation—the relevant direct excerpts are presented along with the Center’s initial thoughts about challenges, opportunities, and possible approaches.

NIA Requirements	Challenges, Opportunities, and Approaches
<p>(A)(2) <u>Theory of Action</u> (up to 5 points)</p>	<p>See the recently produced papers: Marion, S. F. (2010, April 16). Developing a theory of action: A foundation of the NIA response. Available at www.nciea.org Gong, B. (2010, April 16). Some considerations for designing summative assessment components with item types beyond machine-scorable item types. Available at www.nciea.org</p>
<p>(A)(3) <u>Assessment System Design</u> (up to 55 points)</p> <p><i>The extent to which the design of the eligible applicant’s proposed assessment system is innovative, feasible, and consistent with the theory of action. In determining the extent to which the design has these attributes, we will consider—</i></p> <p>(a) <i>The number and types of components (e.g., through-course summative assessments (as defined in the NIA), end-of-year</i></p>	<p>Challenges, Opportunities, and Approaches:</p> <p>. While subsection (a) is important, it does not make sense to decide on the number and types of components until the analyses completed for (b). While it appears that points will be awarded for the number of components, it is more important to create a defensible plan for assessing the full breadth and depth of the standards. Focus on the relationship between the various components and the theory of action.</p> <p>Design decisions:</p> <ul style="list-style-type: none"> ✓ What will be the general structure of the summative assessment? ✓ Will the consortium include through-course components in the summative assessment system? If so, how? ✓ Will the consortium develop an interim assessment component, and if so, what will be its design (e.g., common forms, common specifications, common item bank with

<p><i>summative assessments, formative assessments, interim assessments in mathematics and in English language arts in the assessment system);</i></p>	<p>specifications and forms left up to local choices, etc.)?)</p> <ul style="list-style-type: none"> ✓ How will the consortium ensure the coherence among the various components of the assessment system? ✓ Formative assessment is not something that can be replicated or sold, but in fact is a process by which teachers engage with students to collect information in order to monitor and adjust instruction. Will the consortium design professional development structures to ensure that formative assessment is included as a critical part of the assessment system?
<p><i>(b) For the assessment system as a whole—</i> <i>(i) How the assessment system will measure student knowledge and skills against the full range of the college- and career-ready standards, including the standards against which student achievement has traditionally been difficult to measure; and provide an accurate measure of student achievement, including for high- and low-performing students, and an accurate measure of student growth over a full academic year or course;</i></p>	<p>. This is a critical section with many key components that must be addressed clearly and persuasively.</p> <p>i. This section will require an analysis of the consortium’s common standards in order to generate a set of test specifications where the design team describes how each of the standards and objectives will be assessed. Ordinarily, this detailed and time-consuming process begins when the final set of content standards is in hand. Doing this work in a compressed time without having the final standards document probably implies the consortium’s responses to this section are subject to revision.</p> <p>The language in this section sends two fairly obvious signals. The first is that the consortium must decide how its assessment system as a whole is going to measure “traditionally difficult to measure” standards. This is probably not a direct call for performance assessments, but respondents must make clear how they intend to measure knowledge and skills not captured well by multiple-choice items. The second is that the consortium must decide how its assessment system as a whole is going to measure the range of performance and growth.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ Which standards (or parts of standards) will be measured by which component of the assessment system? In particular, which “traditionally difficult to measure” standards will be measured for the summative component (e.g., speaking, listening, estimation and other mental operations in mathematics, explanation or justification of problem solving in mathematics, complex writing assignments, reading of more than a passage of literature, etc.)?) ✓ Which aspects of the standards for summative assessment will be measured by particular item formats (e.g., multiple choice, other machine scorable, simple constructed response, extended constructed response, performance assessment, etc.)? How will you decide and why? Note: We suggest the proposal use illustrative examples of standards and associated potential item types instead of laying out a complete test blueprint.

	<ul style="list-style-type: none"> ✓ How long will the summative assessment “session” be (e.g., one class period, two, three)? ✓ Will the consortium propose that certain standards be measured by particular item formats for the non-summative assessment components and not by the summative assessment? Why? ✓ What “feel” do you want for the assessment in terms of how much emphasis will be given to performance assessment, human-scored constructed response, and machine-scored items (or if adaptive, what is the general intent)? <p>The second signal might be seen to favor adaptive testing or multi-stage testing: “provide an accurate measure of student achievement, including for high- and low-performing students.” One might argue that current designs yield “accurate” measurements of student achievement, but not very “precise” (was the technical term chosen purposely in the NIA?) at all score points. The most important decision for a consortium is to decide what uses it will make from the comprehensive assessment system, and what are the desired levels of accuracy and precision for each use (e.g., high stakes vs. lower stakes)? The ways to achieve acceptable levels of accuracy and precision may differ by level of aggregation¹. So, using adaptive summative assessment is one way to meet this requirement. Another option might be to use a longer fixed form test or multi-stage test specifically designed to provide adequate information targeted to a specific use. It might also be fruitful to consider expanding the measurement continuum through the use of an interim component of the assessment system rather than trying to address all issues with the summative test.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ How accurate and precise does the measurement need to be for which report, decision, and use? Is this required precision the same all along the achievement continuum? ✓ What approach should be taken to get the desired accuracy and precision? ✓ What are the trade-offs among the possible options? Which are best aligned with the theory of action? ✓ How should non-summative components be considered to fulfill this purpose? <p>The last point in this section about the need to provide an accurate measure of student growth is easier to imagine in 3-8 than high school. Ideally, we could design tests built from learning progressions to produce content-referenced measures of student growth,</p>
--	---

¹ For example, it is typically possible to get a much more reliable school assessment score than individual student score by aggregating across multiple students and/or matrix sampling across items; it is possible to get a very unreliable and invalid school accountability score using highly reliable and valid student scores.

	<p>but a more conventional way of addressing this component is to document the plans for creating at least a vertically articulated set of achievement standards. Consider, too, what “good enough” growth means and how the information from the assessments can be maximized to provide useful information on student movement towards the goals. However, there is clearly the expectation, as revealed in the ESEA reauthorization blueprint, for measuring student growth at the high school level.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ What is the consortium’s definition of growth or learning? ✓ What will be the consortium’s plan for measuring and evaluating student growth over time? ✓ How will the consortium address the need for measuring student growth at the high school level? ✓ Are the high school decisions affected by the choice between end-of-domain or end-of-course testing systems?
<p>(ii) <i>How the assessment system will produce the required student performance data (i.e., student achievement data and student growth data (both as defined in the NIA) that can be used to determine whether individual students are college- and career-ready (as defined in the NIA) or on track to being college- and career-ready (as defined in the NIA));</i></p>	<p>ii. This section is a call for a coherent system design whereby the end-of-high school test is anchored in college/career ready expectations. Once this anchor point is established, test design, including in part the standard setting approach for the lower grades, is tied to ensuring that the achievement standards are articulated across grades such that “on track” performance in lower grades actually leads to college/career readiness. Of course, one of the major challenges in accomplishing this task is to do so without data (later in the document, the NIA calls for empirically-based approaches for validating these college-ready standards).</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ What is the consortium’s definition of college-readiness, and what will count as sufficient evidence? ✓ What is the consortium’s definition of career-readiness, and what will count as sufficient evidence? ✓ How does the consortium propose to ensure that the end-of-high school tests measure college/career ready knowledge and skills? ✓ How does the consortium propose to ensure that the assessments below high school are able to measure students’ performance in terms of being on track to meet college and career readiness standards. ✓ What is the consortium’s plan for developing common performance level descriptors

	<p>focused on college and career readiness standards?</p> <ul style="list-style-type: none"> ✓ How does the consortium propose to establish cutscores that are grounded in college and career readiness standards²?
<p>(iii) <i>How the assessment system will be accessible to all students, including English learners and students with disabilities, and include appropriate accommodations (as defined in the NIA) for students with disabilities and English learners; and</i></p>	<p>iii. Section (iii) should be familiar to most folks involved with test design and development during the past several years. It calls for thoughtful and intentional attention to issues of access and inclusion. A successful proposal should design in this access from the beginning and not simply do so as an add-on later.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ What framework (e.g. UDL) will be used to ensure that SWD and ELL students will have appropriate access to the assessments? ✓ How will the movement to computer based testing alleviate issues of access? What challenges will it raise? ✓ How will the consortium deal with tradeoffs between preserving the construct and trying to avoid any construct-irrelevant factors? This will be an issue as the consortium deals with the important contextual factors often included as part of complex performance assessments? ✓ For CAT designs, how will the item selection algorithms deal with students who learn and demonstrate that learning in non-traditional (and non-linear) ways? ✓ Will the consortium require all states to adopt identical accommodations and administration policies? If not, how will the consortium make decisions about the tolerance for differences in these policies and practices?
<p>(iv) <i>How and when during the academic year different types of student data will be available to inform and guide instruction, interventions, and professional development; and</i></p>	<p>iv. Section (iv) appears to require that the assessment system produce data well before the end of the school that can be used to inform instruction in some way. There is no requirement that these data be derived from the summative assessments, and it might make more sense for these data to come from non-accountability assessments such as interim or formative rather than summative tests, even if these data are derived from through-course components.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ What does the consortium mean by “inform and guide instruction, interventions, and professional development”? What is the state’s role in this? The district’s? ✓ What “student data” does the consortium propose will be made available as a result of this

² See Perie, M. (2010, April 15). Thoughts on Developing College and Career Ready Achievement Standards that are Internationally Benchmarked

	<p>project that will “inform and guide instruction, interventions, and professional development”?</p> <ul style="list-style-type: none"> ✓ How will the consortium ensure that instructionally-relevant data are provided throughout the school year? How does providing these data link back to the theory of action on improving teaching and learning? ✓ Will the assessment system include through course summative assessments that will also be used to provide instructionally-useful information? ✓ Will the consortium create a common interim assessment system to provide these data throughout the school year? If not a common interim assessment system, how does the consortium propose to provide this instructionally useful information? ✓ How will the consortium evaluate the instructional usefulness of the information generated from the assessment system?
<p>(c) For each component in mathematics and in English language arts in the assessment system--</p> <p>(i) The types of data produced by the component, including student achievement data (as defined in the NIA), student growth data (as defined in the NIA), and other data;</p>	<p>ii. This subsection requires the consortium to discuss how the member states intend to use the various forms of data for both accountability and instructional purposes.</p> <p>i. In this section, the consortium must indicate which of the assessment system components will produce status (i.e. point-in-time) information and which of the components will be used to provide student growth information. It is likely that all components will be used to produce status information of some kind, but clearly not all data will be useful for growth calculations.</p> <p>Design decisions:</p> <ul style="list-style-type: none"> ✓ How will the consortium propose to measure growth? ✓ Which components of the assessment system will be used to measure student growth? ✓ How will the consortium ensure that these growth measures yield valid growth data? <ul style="list-style-type: none"> ○ For grades 3-8? ○ For high schools? ✓ Is the consortium interested in creating a vertical score scale to link the various summative assessments (and perhaps interim assessments)? If so, why and how will this scale be created so the score and growth inferences are defensible? If not, how does the consortium propose to measure student longitudinal growth?
<p>(ii) The uses of the data produced by the component, including determining whether individual students are college- and career-ready (as defined in the NIA) or on track to being college- and career-ready (as defined in the NIA); informing determinations of school effectiveness for</p>	<p>ii. There are many purposes and uses of assessment data packed into this one requirement. Consortia will need to unpack this requirement carefully to make sure it addresses each of the sub-requirements. The requirement asks the consortium to specify how each of the data components is used. It is important to remember that the NIA does not require that each of these uses are met for each data sources, but it seems like a clear signal that the system as a whole must address each of these uses. Most critically, the specific uses, as outlined in this requirement, for each of the assessment system components, should be</p>

<p><i>the purposes of accountability under Title I of the ESEA; informing determinations of individual principal and teacher effectiveness for the purposes of evaluation; informing determinations of principal and teacher professional development and support needs; informing teaching, learning, and program improvement; and other uses;</i></p>	<p>linked directly and rationally to the theory of action.</p> <p>The NIA refers to formative assessment, but the language is not consistent with the way that it is defined by most formative assessment researchers. The consortia will have to decide carefully how to approach the topic of formative assessment, particularly in terms of providing the professional development structures and supports necessary to provide teachers with the knowledge and skills necessary for implementing formative assessment strategies.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ What is the consortium’s definition of “college and career ready”? What is sufficient evidence to make that determination? ✓ How does the consortium intend to determine that students are college and career ready? Will these be an end-of-high school measurement, a collection of evidence throughout high school, and/or an earlier assessment(s) such as the end of 10th grade? ✓ How does the consortium intend to document that students are on-track (or not) towards these college and career ready expectations? Will grade level determinations of proficiency be aligned with this conception of being on track to CCR? ✓ Which data sources will be used for these two purposes and uses? ✓ Which of the various assessment components at each level (3-8, HS) will be used to address each of the following purposes/uses: <ul style="list-style-type: none"> ○ School “effectiveness” for accountability? ○ <u>Informing</u> evaluations of teacher and leader effectiveness? ○ Providing data to guide program evaluation purposes such as professional development needs? ✓ What is the consortium’s strategy for addressing the issue of formative assessment? <ul style="list-style-type: none"> ○ Will the consortium create a general structure and set of supports for formative assessment, while leaving the implementation up to individual states? ○ If the consortium intends to design a consortium-wide approach, how will the individual states, districts, and schools ensure that the formative assessment strategies are adapted appropriately to local contexts?
<p><i>(iii) The frequency and timing of administration of the component, and the rationale for these;</i></p>	<p>iii. The response to the NIA will not only need to describe each of the components of the assessment system, but should be expected to describe when and how often these assessments will be given. Again, these decisions should be grounded in the theory of action.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ Who (e.g., policymakers, parents, teachers) needs what information and how often do they

	<p>need it?</p> <ul style="list-style-type: none"> ✓ Will decisions about when certain components are administered be common across the consortium or will states have the flexibility to make some of these decisions individually? If the latter, how will the consortium make these decisions such that the security of the assessments, if applicable, be maintained? ✓ Do these data needs differ by subject area and grade level? If so, how and how does the consortium propose to address these differences? ✓ What else affects frequency and timing of administration of the assessment components, such as budget, when information is needed (e.g., some assessments may be moved earlier in the year to allow time for human scoring; some assessment information may be collected at one point in time but stored until a later time for various reasons, including combining with other data elements needed to make a decision)?
<p>(iv) The number and types of items (e.g., performance tasks, selected responses, brief or extended constructed responses) and the distribution of item types within the component, including the extent to which the items will be varied and elicit complex student demonstrations or applications of knowledge and skills (descriptions should include a concrete example of each item type proposed); and the rationale for using these item types and their distributions;</p>	<p>iv. USED appears to be asking for an incredible level of specificity at this stage in the process. It is not clear how, before the CCSS are even finalized, a consortium can be so specific in the design. As noted above, this type of effort requires a careful and detailed analysis of the CCSS as well as any intermediate materials (e.g., model curricula, learning progressions). We suggest providing a description and rationale of the types of items/tasks proposed in the response, but most importantly, it will be critical to outline the processes by which the consortium will make these decisions. Additionally, we suggest using the most recent draft of the CCSS to illustrate the types of standards/objectives that will be assessed by each item type. Finally, we suggest that the consortium describe, at least in general terms (but not overly general) the types and ratios of items and tasks for each of the major component parts of the system.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ Is the consortium willing to provide the level of specificity apparently required in the NIA? <ul style="list-style-type: none"> ○ If so, how does the consortium propose to make these decisions? ✓ Does the consortium envision including any items/tasks that are not used operationally at this point (i.e., still in the R & D stage)? If so, how will the consortium write about these developing formats in the proposal? How will they incorporate them into scores later? ✓ Does the consortium envision different types of items/tasks for the different components of the system? If so, what is the rationale for the differences in distributions?
<p>(v) The component's administration mode (e.g., paper-and-pencil, computer-based, or other electronic device), and the rationale for the mode;</p>	<p>v. This straightforward requirement asks the consortium to define for each component how it will administer the various assessments in the system. To the extent the computer-based and/or computer-adaptive tests are proposed, the response should indicate the proportion of test takers participating via each mode.</p>

	<p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ What platform is the consortium proposing for test delivery? To the extent that the consortium proposes computer administrations, what will the consortium expect from member states in terms of degree of implementation? What is the transition plan for states not yet fully computer ready? ✓ Will the types of administration vary by assessment type (e.g., formative, interim)? If so, how and why?
<p>(vi) <i>The methods for scoring student performance on the component, the estimated turnaround times for scoring, and the rationale for these; and</i></p>	<p>vi. The consortium needs to specify how the items/tasks will be scored. This is obviously important for open-response items/tasks. It seems clear that USED, which has expressed interest in teacher scoring, wants to see how the consortium proposes to score the open-response tasks in ways that are economical, efficient, and if possible provide opportunities for professional learning. Additionally, USED would like the consortium to discuss turnaround times for the open-response tasks as well as for the various components as a whole.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ How does the consortium propose having the open-response tasks scored? ✓ Do these scoring approaches vary by assessment system component (summative, interim) and/or by timing of administration (e.g., through-course or end of year)? ✓ Will teachers be used for scoring? Does the answer vary by assessments system component? ✓ For consortia that propose teacher-scoring, particularly for summative components, the consortium should foreshadow how it proposes to deal with security, scorer quality, and training issues (these issues are addressed in-depth in the assessment development section). ✓ How will the consortium decide the speed with which the results from the various components of the assessment system need to be returned to stakeholders? ✓ For systems that propose incorporating scores from performance tasks with multiple-choice, machine-scored (including computer administered tests), will the consortium wait until all components are completed before returning scores or will it produce scores when the component parts are ready?
<p>(vii) <i>The reports produced based on the component, and for each report, its intended use, target audience (e.g., students, parents, teachers, administrators, policymakers), and the key data it presents.</i></p>	<p>vii. Ron Hambleton likes to say that assessment reports are the major way in which we communicate with the public about the assessment system, but it is the last thing we tend to address in design. Ironically, the NIA—at least in terms of ordering—does the same thing. Therefore, we strongly recommend that the reporting system design occur as early in the overall design process as possible. Further, the reporting system should be an explicit part of the theory of action. It is important to think about many of the things</p>

	<p>discussed already, such as the function and use of each of the assessment system components. Articulating a report design as part of the overall design process requires clarity in thinking about who needs what assessment data, in what form should the data be reported, and how often these reports needs to be made available. A comprehensive assessment system requires a comprehensive reporting system so that the various intended uses of the assessment results can best serve their purposes.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none">✓ What data will make the greatest impact on teaching and learning? How will these data be clearly communicated and made accessible so they can have the desired impact?✓ How does the consortium propose to develop a reporting system to serve the multi-faceted purposes of the assessment system?✓ How will the reporting system incorporate the various components of the assessment system? Will there be one reporting system or multiple systems? Will some reports be developed at the state level or will they all be consortium-developed?✓ How will technology be incorporate into the consortium’s reporting structure and how will this use of technology support the purposes and uses of the assessment system?
--	--

<p>(A)(4) <u>Assessment System Development</u> (up to 35 points)</p> <p><i>The extent to which the eligible applicant’s plan for developing the proposed assessment system will ensure that the assessment system is ready for wide-scale administration in a manner that is timely, cost-effective, and consistent with the proposed design and incorporates a process for ongoing feedback and improvement. In determining the extent to which the development plan has these attributes, we will consider—</i></p>	<p>This next major section requires respondents to add considerably more detail to what was described in the design section above. However, the means for making the decisions about the assessment design in the previous section should be consistent with the approach for developing the assessment system described in this section (e.g., evidence-centered design). As a strategy for trying to discuss all of these important details in only about 20-25 pages between the design and development sections, we suggest that each consortium carefully integrate the narratives of these two sections.</p>
<p><i>(a) The approaches for developing assessment items (e.g., evidence centered design, universal design³) and the rationale for using those approaches; the development phases and processes to be implemented consistent with the approaches; and the types of personnel involved in each development phase and process (e.g., practitioners, content experts, assessment experts, experts in assessing English learners, experts in assessing students with disabilities, psychometricians, cognitive scientists, IHE representatives, career and technical education experts);</i></p>	<p>The USED heard from many assessment experts who suggested requiring all proposals incorporate a theory-based or principle-based design approach to test design and development. This section requires that the approach each consortia uses for item and test development are guided by a theoretical approach (e.g., ECD) designed to yield the most valid score inferences possible. The consortia should clearly describe how it intends to translate the CCSS into test and item specifications, including, if applicable, creating some intermediate products (e.g., model curricula, learning progressions).</p> <p>The consortia should clearly indicate the expected development processes used for the various assessments in the system in terms of subject, grade level, and assessment type (interim, summative). The consortia should also describe, to the extent possible, the types of stakeholders and experts that need to be involved in each of the assessment development processes. Depending on each consortium’s intended connections with the curriculum and instruction—which we hope is considerable—we suggest that content and curriculum experts should be one of the most important groups to include in both helping to translate the standards into test specifications and assessment items/tasks?</p> <p><u>Development decisions:</u></p> <ul style="list-style-type: none"> ✓ What theoretical approach will the consortium require its development personnel (e.g., teachers) and/or vendor(s) to use and how will it monitor and validate the implementation of this approach?

³ Universal design for learning” is used as that term is defined in section 103(24) of the HEA.

	<ul style="list-style-type: none"> ✓ How will the consortium translate the common standards into item and test specifications? Will there be intermediate products to help this translation? ✓ Will the consortium employ different design and development process for the different assessments in the system (types, subjects, grade levels)? If so, what is the rationale for these different choices?
<p><i>(b) The approach and strategy for designing and developing accommodations (as defined in the NIA), accommodation policies, and methods for standardizing the use of those accommodations for—</i></p> <ul style="list-style-type: none"> <i>(i) English learners; and</i> <i>(ii) Students with disabilities;</i> 	<p>Again, this was an issue addressed in the design section, but it is clear that USED expects details and rationale for how the consortia will design and implement an inclusive assessment system. The consortium should clearly distinguish among ELL and students with disabilities and how it intends to address the unique needs of each group.</p> <p>Most importantly perhaps, the consortium should specify how it will develop accommodations policies for the entire consortium. Ideally, the consortium could make a commitment towards a common accommodations policy, but it should test this idea with a few key examples (e.g., reading the reading test) to see if it will be able to achieve agreement across member states.</p> <p><u>Development decisions:</u></p> <ul style="list-style-type: none"> ✓ What model or approach will the consortium use to ensure that the assessment will be as inclusive as possible? This response needs to be considerably more specific than simply saying, “this consortium will employ Universal Design for Learning (UDL) principles in the design and development of its assessment system.” Therefore, what are the specific design and development actions the consortium intends to pursue in order to maximize inclusivity? ✓ Will the consortium adopt a common accommodations policy for each of the two special populations? If so, what is the proposed approach for facilitating such an agreement?
<p><i>(c) The approach and strategy for ensuring scalable, accurate, and consistent scoring of items, including the approach and moderation system (as defined in the NIA) for any human-scored items that are part of the summative assessment components and the extent to which teachers are trained and involved in the scoring of assessments;</i></p>	<p>The requirement is wisely focused on the summative assessment and not on the more instructional components of the system. This section requires the consortia to describe how it will monitor the technical quality of human-scored items and tasks, but it is surprising to see this emphasis on something we already know how to do quite well. There are many well-established processes and procedures for ensuring the accuracy and consistency of human-scored, whether teachers are part of the pool of scorers or not.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ Will the consortium include active teachers as part of the pool of scorers on the summative assessments? Why? Will the requirements for teacher involvement in scoring be the same across participating states? ✓ If the consortium will include active teachers as part of the pool of scorers: <ul style="list-style-type: none"> ○ How does the consortium propose using the teacher scorers? ○ How will the consortium maintain the security of the summative assessment with

	<p>teacher scorers?</p> <ul style="list-style-type: none"> ✓ How will the quality control of the scoring process be assured whether teachers are part of the pool of scorers or not?
<p><i>(d) The approach and strategy for developing the reporting system; and</i></p>	<p>As mentioned in the design section above, the reporting system is the main way that assessment results can become actionable. Designing reporting systems in the static formats that currently predominate will not take advantage of some of the newer technologies and designs of interactive reporting systems.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ Will the consortium’s “reporting system” include common hardware or software requirements that will apply to all participating states? ✓ How will the consortium, in its proposal, describe how its proposed reporting system, will take advantage of state-of-the-art designs and technology? ✓ Will the reporting system be developed separately for the summative and interim components? If so, how and why? ✓ How will the consortium secure the necessary expertise (e.g., Java, Flex developers) to ensure that the reporting systems are state of the art?
<p><i>(e) The overall approach to quality control; and the strategy for field testing assessment items, accommodations, scoring systems, and reporting systems, including, with respect to assessment items and accommodations, the use of representative sampling of all types of student populations, taking into particular account high- and low-performing students and different types of English learners and students with disabilities.</i></p>	<p>The section requires the consortia to specify a general approach to field testing everything from items to reporting systems. It appears to suggest a multi-stage and multi-faceted approach to field testing. For example, innovative items, innovative uses of items (e.g., curriculum-embedded tasks), accommodations, and/or new score reports might be tried out in a much smaller scale before full-scale field testing. This section also expects respondents to discuss the general timing and approach for ensuring that the field testing system will allow for the development of a high quality operational assessment.</p> <p>Further, the consortium respondents should lay out a comprehensive quality assurance/quality control plan. Most assessment vendors have such plans and other organizations (e.g., TILSA, ASR SCASS, ATP) have produced such QC/QA documents as well.</p> <p><u>Design decisions:</u></p> <ul style="list-style-type: none"> ✓ What is the consortium’s approach to “item tryout” and field testing? How will this approach assure that a high-quality operational assessment can be produced by 2014-2015? ✓ How will the consortium deal with challenges of producing innovative items/tasks and ensuring that there is adequate time and opportunity to produce and validate these tasks? ✓ Producing a high quality and specific QA/QC plan could require the entire 60 pages. What will be the consortium’s general framework for ensuring the design and development of a high quality operational assessment system?

(A)(5) Research and Evaluation (up to 30 points)

A forthcoming Center for Assessment paper will offer analysis and suggestions for addressing this aspect of the NIA.

ⁱ This paper is one of a series of brief documents produced by the Center for Assessment designed to assist state consortium leaders in thinking about and preparing their responses to the United States Department of Education Notice Inviting Applications for Comprehensive and High School Assessment System grants. Funding has been provided by the Bill & Melinda Gates Foundation. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Gates Foundation.