

MEMORANDUM

To: U.S. Education Secretary John King
From: KnowledgeWorks and the Center for Assessment
Date: March 21, 2016
Subject: Recommendations for Implementation of the Innovative Assessment and Accountability Demonstration Authority Authorized in Section 1204 of the Elementary and Secondary Education Act

Implementation of the Every Student Succeeds Act (ESSA) provides the U.S. Department of Education (ED) with a significant opportunity to support interested states in developing high-quality student centered systems of assessments that support competency-based learning. While there are a number of provisions in ESSA that states can leverage to begin building these systems, the Innovative Assessment and Accountability Demonstration Authority authorized by Congress provides states with an unprecedented opportunity to develop next generation approaches to assessment that transcend the standardized instruments that states have implemented for nearly two decades.

KnowledgeWorks and the Center for Assessment encourage ED to consider the recommendations presented in this memo for how to best support states as it develops its regulatory and application strategy for implementation of this demonstration authority. These recommendations are divided into two sections: 1) *A planning process for states interested in exploring this opportunity but need additional support and resources*, and 2) *An implementation process for states that have widespread stakeholder buy-in and a high-quality plan to begin implementation*. The success of this demonstration authority rests heavily on ED's leadership to design an application process that provides states with the appropriate flexibility to design, implement, and evaluate innovative assessment systems while incorporating guardrails to ensure high quality implementation.

Our Experience

KnowledgeWorks and the Center for Assessment bring important perspectives to this work, partnering with states, districts, and educators in the design and implementation of competency-based education systems.

- **KnowledgeWorks** is a social enterprise focused on ensuring that every student experiences meaningful personalized learning that allows him or her to thrive in college, career and civic life. By offering a portfolio of innovative education approaches and advancing aligned policies, KnowledgeWorks seeks to activate and develop the capacity of communities and educators to build and sustain vibrant learning ecosystems that allow each student to thrive. Our on-the-ground work includes partnerships with schools through competency education and EDWorks early college high schools, as well as with communities through StriveTogether. We also provide national thought leadership around the future of learning.
- **The Center for Assessment** is a nonprofit organization whose mission is to foster improved student achievement through enhanced policies and practices in educational assessment and accountability. The Center staff brings extensive expertise in all aspects of assessment and accountability, development, operations, and policy formation from multiple perspectives. To date, the Center has had contracts with over 40 states and currently serves as the lead technical partner in the design and implementation of New Hampshire's Performance Assessment for Competency Education (PACE).

The Benefits of Student-Centered Assessment

When Congress established the Innovative Assessment and Accountability Demonstration Authority, it did not intend for the program to serve as a rubber stamp for any assessment system. Instead, congressional leaders saw this as an opportunity to support states in the development of high-quality, innovative assessment systems. New assessment systems should provide a data-rich picture of each student's level of proficiency – including students who have not yet met proficiency—to ensure continuous improvement of learning and to inform annual determinations of student and school performance.

Stakeholders exploring this opportunity should consider the wide range of potential benefits inherent in student-centered assessment approaches that provide information about student skills relative to content standards or competencies. Educators collect this information through the use of complex performance tasks and/or other assessment forms that call for demonstrations of deeper learning. In addition, student-centered assessment systems must be designed to operate in concert with the curriculum, instruction, and learning. Such assessment systems cannot be designed and implemented by those external to the teaching and learning processes. Specifically, these assessment systems must be designed to:

- **Document and Support Student Achievement of Complex Knowledge, Skills, and Dispositions** – Students have the opportunity to engage in assessment tasks that elicit performance of meaningful knowledge and skills at the depth of understanding that they can transfer to new real-world situations.
- **Deepen Focus on Student Learning, Engagement, and Outcomes** – Assessments not only document what students have learned, they also inform educators and students of the impact of specific learning activities so educators make necessary adjustments in real-time to ensure continuous improvement of student learning. Additionally, complex performance tasks serve as a signal for the types of instructional activities one hopes to see in everyday classroom practice.
- **Support Educator Agency** – Educators must develop a sense of ownership in the assessment and accountability systems and play an integral role in designing, administering, and scoring complex performance tasks. Additionally, because we recommend that the assessments used for accountability be embedded within classroom-based instructional units, the critical connection among curriculum, instruction, learning, and assessment must be apparent to students and teachers.
- **Increase Educator Capacity** – Educators benefit from extensive training and professional development opportunities to deepen assessment literacy. Assessment literacy translates to improved local assessment design, as well as improved understanding of how to use student work and other assessment data for increasing student learning. Strong alignment among assessment, curriculum, and instruction also helps educators design engaging instructional programs that prepare students for success on assessment tasks.
- **Support Reciprocal Accountability** – School accountability works best if the responsibility for design and implementation is shared by districts and the state. Balanced assessment systems of the kind we envision are designed to help states and communities transition away from externally-oriented accountability approaches that often disenfranchise local leaders and educators who feel little sense of ownership in the processes and metrics on which they are evaluated. Reciprocal accountability systems, in contrast, require that for every request that the state makes of the local

districts, the SEA provides the necessary and timely support that districts need to meet those requests. School districts then have the onus to show how the support has been leveraged to implement the accountability system with fidelity and ultimately improve student outcomes.

- **Reduce Concerns About Over-Testing** – Since assessments are embedded in curriculum and useful for informing instruction and academic planning, stakeholders can see the benefits of each assessment in real-time to support education improvement. This is in contrast to traditional standardized approaches where assessment data are generated after the fact. While the system of assessments envisioned here may not necessarily reduce the amount of assessment time per se, the assessments are seen as useful and relevant, so concerns about “testing time” evaporate. Because of the increased utility of the state-local system of assessments, the need for “extra” local district-administered assessments (e.g., commercial interim assessments) is likely reduced.
- **Increase Productivity of the Education System** - Access to greater evidence of student learning during the time when learning is occurring will help local and state educational leaders enhance the productivity of their education system and strengthen both the internal and external accountability processes, fostering continuous improvement of student and school performance.

This wide range of benefits represents the significant potential that exists with high-quality implementation of the Innovative Assessment and Accountability Demonstration Authority. In order to capitalize on these benefits, we encourage ED to incorporate the following recommendations into the design of its application process.

Section 1: Planning Process

While national interest in the demonstration authority is high, there are very few states that are ready to begin implementation. We strongly encourage ED to respond to this interest by allowing states to apply for a two-year planning process, which would not count against the statutory implementation time schedule, to enable them to put together a high-quality implementation plan. The benefits of establishing a planning process include:

- **Invest in Quality** – ED, the states, and philanthropic partners can dedicate resources to support interested states in the development of a high quality plan.
- **Establish Parameters Around a Quality Planning Process** – ED can ensure that applicants address key elements in their planning process that will be essential to high quality implementation.
- **Establish Communities of Practice** – ED can work with external partners to provide necessary supports to states to address specific areas of interest that arise in the planning process.
- **Establish a Transparent Progression to Implementation** – A planning process would bring greater transparency to the application process, establishing a pathway for ED and other interested stakeholders to develop a deeper understanding of each state’s vision and goals so they can support them along a progression to approval for implementation.

A planning process would provide ED with an opportunity to help establish a pipeline of states focused on developing high quality applications. [Appendix 1](#) provides specific planning recommendations for ED to consider as it explores this option. The chart in the appendix specifically outlines elements that ED should incorporate into a planning process to ensure states are on track for development of a high quality application to begin implementation. The elements encompass a state vision, eligibility criteria, competitive preference priorities, selection criteria, allowable activities, and evaluation.

Section 2: Implementation Process

In order to design a high quality application and monitoring process, ED must ensure that states have a compelling plan to address each of the guardrails included in the program. Based on our experiences working with states and other stakeholders interested in student-centered assessments, we believe there are four critical guardrails for effective implementation. As ED considers each of them, it is important to have a clear vision of how a high quality application would address these requirements. The following recommendations provides context to help ED better understand each of these guardrails. Please also reference [Appendix 2](#) for a summary tool that summarizes the key points from the recommendations below.

- 1) **Assessment Quality** - *The state needs to demonstrate that the system of assessments is comprised of high quality assessments that provide useful information to relevant stakeholders about what students know and can do relative to the content standards as well as supporting comparable annual determinations.*

Since the demonstration authority would permit a state to derive comparable annual determinations, at least in part, from local assessment information, a state and local focus on assessment quality is essential to the success of the demonstration authority. Further, the proposed system must attend explicitly to the assessment literacy of educators who will develop and administer the assessments by investing in training and building assessment capacity. Through the concept of a reciprocal accountability model we described earlier, the districts receiving state support must put forward resources to improve the quality of their locally developed and administered assessments to ensure that they are meeting standards for quality. Not only will this system ensure students are engaging in high quality assessment tasks, but as educator capacity increases, the use of local assessment information to improve instruction will proceed on a parallel path. In this way, an intense focus on building local capacity to improve and monitor the quality of local assessments will be essential for realizing the ultimate goal of improved student outcomes.

More specifically, we recommend that states and participating districts have a plan that addresses the following dimensions of assessment system quality:

- The state’s plan for **designing and developing high quality assessments** must employ methods based on up-to-date research and best practices such as the use of a “principled assessment design” process (e.g., Evidence-Centered Design (ECD)). Rigorously following a process such as ECD is generally beyond what is done for all state assessments, but the state and districts should use a process that builds off what we know about high quality assessment development. State plans must include a plan for building the capacity of local educators to improve their assessment literacy and expertise in assessment development and use.
- The assessments used for both instructional and accountability purposes must be **aligned to the content standards** students are expected to learn. In the case of the Demonstration Authority, the alignment focus needs to be on the system of assessments used to produce the annual determinations. This does not mean that individual assessments should not be aligned to the standards and competencies, but any single assessment within the system may measure (by design) only a portion of the content standards. The state should be expected to demonstrate that the assessment system used to produce the annual determinations covers the full range, breadth, and especially the depth of the required knowledge and skills.

- The latest version of the *Standards for Educational and Psychological Measurement* (AERA, APA, & NCME, 2014) elevates the importance of **fairness and accessibility** as does the latest iteration of the ED Standards and Assessment Peer Review Guidance. Therefore, states and districts must document that they are using an approach such as Universal Design for Learning (UDL) as part of their design framework and must also have a policy regarding the use of accommodations for students with identified disabilities and English learners.
 - Concerns are often raised about the quality and **consistency of scoring** when discussing performance assessments and other open-response tasks. While this is a legitimate concern, it is something that is well-understood and there are many well-developed tools and approaches for ensuring that teachers can score assessments accurately and consistently. That said, states and districts should be required to discuss how they will monitor the quality of the scoring rubrics, the within-district interrater reliability evidence, and the cross-district calibration and consistency.
 - One of the most important reasons for states to seek out the flexibility offered under the Demonstration Authority is because they want to move away from assessments that only provide a “look back” and are not useful for improving instruction. Therefore, **utility of assessment results** in terms of improving instruction is a crucial consideration for the quality of state assessment systems.
- 2) **Comparability of the Assessment System Results** - *The state needs to demonstrate that its innovative assessment system produces yearly, student-level annual determinations comparable across LEAs participating in the pilot and with those not participating in the pilot.*

Comparability of assessment results and annual determinations is a major policy goal and essential for building stakeholder trust in the quality of assessment data used to inform accountability determinations. While the term “comparability” appears simple, the specific meaning is interpreted quite differently depending on one’s technical background, the focus of the comparability inferences (e.g., specific assessment or assessment system), and the intended uses of the assessment system results. Psychometricians often speak of “interchangeability” of assessment scores such that a score of 240, for example, on an assessment in one year is interchangeable with a score of 240 on that assessment in a different year. This is the basis and purpose of test score equating. On the other hand, policymakers often want to know that students in their state (or district) are being held to expectations similar to those in other states (or districts). This is a more technically relaxed version of comparability. We assert that the comparability required for assessment systems as part of the demonstration authority falls between these two examples.

One of the ways to achieve comparability, particularly in the U.S., is through standardization. The more standardized we make the various aspects of the assessment from the design as evidenced by such things as test blueprints to administration, scoring, and cutscores, among other dimensions, the more evidence we are able to garner to support comparability claims. It is important to recognize that comparability exists on a continuum and is not a yes/no determination. For example, the National Assessment of Educational Progress (NAEP) employs a degree of standardization far beyond any state or consortium assessment. We do not say that the assessment results in State X are not comparable; rather we say that they are comparable enough for the identified purposes and uses.

As ED implements the Innovative Assessment and Accountability Demonstration Authority, it should ensure that states provide sound evidence that their assessment systems maintain levels of comparability at the system level rather than on any single assessment within that system. While the annual determinations would support the same interpretations about what students know and can do relative to the grade level content standards, flexibility at the assessment level will allow for the use of assessments that can be more directly embedded within classroom learning.

Comparability can be achieved by ensuring that local assessments provide accurate information about student achievement for all of the grade level content (e.g., the assessments measure the full range of the content standards). Quantitative and qualitative auditing techniques can be used to monitor the degree of comparability of the annual determinations (see New Hampshire's October and March PACE reports to ED), while still allowing for local control over the assessments that comprise the annual determination. This balance between flexibility and standardization must be reflected within a high quality submission for the demonstration authority.

ED should look for the following types of evidence and documentation from applicants for supporting claims of comparability:

- **The state includes a plan for developing achievement level descriptors.** High quality plans would include a process by which content experts (e.g., teachers, instructional coaches) are brought together to define the expectations at each level of performance. The resulting achievement level descriptors should undergo a review and assurance that they provide for consistent interpretations of what students know and can do at each level.
 - The state includes a defensible plan **for setting standards (cutscores) that are aligned with the achievement level descriptors** to provide for valid inferences.
 - The state includes a plan to **evaluate claims of comparability using one or more auditing techniques.** Though we argue that claims of comparability should be made at the annual determination level, the annual determinations will likely be made on the basis of information gathered using multiple local assessments. Auditing the comparability of scores on those local assessments will contribute to the evidence supporting the comparability of annual determinations. In order to audit the comparability of local assessments the state may want to plan to include one or more common assessments across pilot districts.
- 3) Statewide System** - *If the state is proposing to administer the innovative assessment system initially in a subset of local educational agencies, the state must have a plan to scale up the innovative assessment system statewide in the State's proposed demonstration authority period (which cannot exceed 5 years) or at the end of an additional 2-year extension period.*

A successful state submission for the Innovative Assessment and Accountability Demonstration Authority will contain a coherent and actionable plan for sustainability. Any high quality plan will include an acknowledgement that the initially proposed system should evolve as the state learns about implementation from a sample of pilot districts. As the state is building a set of processes and procedures that define the assessment and accountability systems for the pilot districts, the consideration of sustainability for statewide implementation is a necessary lens through which to evaluate these programmatic decisions. The statewide sustainability plan should also include a

mechanism for assessing district readiness to join the innovative assessment and accountability system and approving district entry into that system. Most importantly, a high quality and thoughtful state sustainability plan will build in the allowance for re-evaluation and course correction at key junctures throughout the timeframe. Giving states the freedom to reconsider their plans for scaling the system after they have had the benefit of experience with pilot districts will be crucial for the success of the innovation.

- 4) Demographic Similarity** - *The state can describe how the inclusion of additional local educational agencies will help the state make progress toward achieving high-quality and consistent implementation across demographically diverse local educational agencies. The state can also describe how it will ensure that the participating local educational agencies, as a group, will be demographically similar to the state as a whole by the end of the state's demonstration authority period.*

There is no question that states should select districts to participate in the pilot that represent diverse geographic regions and the demographics of the state as a whole. Because of the necessary “proof of concept” period required for all states participating in the demonstration authority, we urge ED to allow for flexibility in interpreting “demographically diverse” in the first few years of the state’s demonstration and “demographically similar” towards the end of the demonstration period as long as the state has a clear plan for meeting the demographic similarity requirement. In fact, we urge ED to allow states additional time in meeting these requirements as long as states have a high quality plan and can demonstrate that they are making progress towards these important goals. Further, even when evaluating the demographic similarity of a subset of districts to the state as a whole, we urge ED to consider reasonable tolerance levels (e.g., $\pm 10\%$) when evaluating the representation of specific student groups.

Conclusion

We are excited about the promise for states and school districts to work together to implement innovative assessment and accountability systems to deepen student learning. We recognize and support that ED is charged with striking a delicate balance between innovation and rigor, but we recommend ensuring that innovation is not stifled in the push for technical quality. We urge ED to move quickly to produce applications and associated guidance for the Demonstration Authority, because we know that several states are eager to apply for the flexibility. Our two organizations are poised to support ED in any way that we can to ensure the success of the Demonstration Authority. For more information, please contact Lillian Pace of KnowledgeWorks (PaceL@knowledgeworks.org) or Scott Marion of the Center for Assessment (smarion@ncea.org).

Appendix 1: Guidance for Development of a Planning Process

An application for a planning process should incorporate all of the following elements.

Element	Evidence
State Vision	<p>The state must describe its vision to improve teaching and learning with a framework that aligns a new innovative assessment system to the state’s academic standards, accountability system, and school, teacher, and leader supports. Importantly, the state must describe how it plans to partner with local districts to enact this vision. For example, many states may choose to implement a version of reciprocal accountability (e.g., Elmore, 2002) to address how the two entities will work collaboratively to meet the goals of the pilot.</p>
Eligibility Criteria	<p>The state must describe how it will address each of the following criteria to ensure eligibility for the planning process.</p> <ul style="list-style-type: none"> ➤ If developing a competency-based approach, develop a coherent set of K-12 course and grade competencies aligned to the state’s challenging academic standards. ➤ Build leadership and educator buy-in and capacity to participate effectively in the pilot and identify initial group of participating districts. ➤ Establish a partnership with individuals and entities with experience in student-centered assessment design. ➤ Establish a process for designing a comprehensive assessment system that will appropriately measure student mastery of standards and aligned competencies as applicable. ➤ Establish a process for identifying quality review procedures to ensure technical quality of the assessment system. ➤ Develop a plan to build state and district capacity to scale the innovative assessment system statewide by the end of the demonstration period including a high quality plan to engage demographically diverse districts. ➤ Engage district leaders, educators, parents, students, institutions of higher education, civil rights organizations, business leaders, community leaders, and other stakeholders in the state throughout the planning process. ➤ Identify and remove state policy barriers that would impede the design or implementation phase.
Competitive Preference Priorities	<ul style="list-style-type: none"> ➤ ED will give competitive preference to states that propose and have taken steps to enact a compelling vision for building an integrated learning, assessment, and accountability education system. Such steps would include authorizing legislation (e.g., requirements for all districts to enact a competency-based graduation system), a successful history of enacting at least some of the reforms envisioned (e.g., performance-based assessments), and documentation of successful efforts for building the capacity of local educators for activities in the proposed Demonstration Authority or related initiative.
Selection Criteria	<ul style="list-style-type: none"> ➤ Quality of the proposed state vision. ➤ Level of stakeholder buy-in. ➤ A defensible plan for developing the capacity to implement the proposed project. ➤ Evidence that the state has an aligned state and local policy framework or has a defensible plan and capacity to develop such an aligned system.

<p>Allowable Activities (if federal funding provided to states as part of planning process)</p>	<ul style="list-style-type: none"> ➤ Partner with technical experts to support the state in the planning and design process for the assessment system including the development of procedures to establish, monitor, and evaluate technical quality of the system. ➤ Facilitate communication with a wide range of stakeholders to increase buy-in for the new assessments. ➤ Conduct professional development activities for educators who will be participating in the assessments. ➤ Make improvements to state and district technology infrastructure to ensure compatibility with the new assessment system.
<p>Evaluation</p>	<p>The state must describe the goals and benchmarks the state will establish to evaluate readiness for transitioning from the planning process to the early stages of design and implementation.</p>

Appendix 2: Recommendations for Implementation Applications

The following chart provides guidance for ED on four guardrail requirements included in statute to help ED identify states with a high quality proposal to begin implementation. For more explanation of each of these elements, please refer to Section 2 in the body of the memo.

Federal Guardrail Requirements	Recommendations for State Applications	Recommendations for High-quality Implementation
1. Assessment Quality		
<p>The state needs to demonstrate that the system of assessments is comprised of high quality assessments that support the calculation of valid, reliable, and comparable annual determinations as well as provide useful information to relevant stakeholders about what students know and can do relative to the learning targets (e.g., content standards, competency statements).</p>	<p>States and participating districts should have a plan that addresses the following dimensions of assessment system quality:</p> <ul style="list-style-type: none"> ➤ Assessment design processes using research-based and best practice approaches that include a plan for building the capacity of local educators. ➤ Evidence that the system of assessments aligns to state academic standards. ➤ Approaches for promoting the fairness and accessibility of assessments for all students. ➤ Plans for monitoring the within-district interrater reliability evidence and the cross-district calibration and consistency. ➤ Utility of assessment results. 	<p>States and participating districts must have a defensible and logical plan for how they will ensure that the assessments used in the system for making annual determinations are of high quality. This plan must describe the ways (criteria) in which the state and districts are establishing, monitoring, and evaluating assessment and system quality, how the state and districts are supporting the development of assessment literacy among participating educators, and the approach for sustainably developing the assessments necessary to support the system’s needs.</p>
2. Comparability of the Assessment System		
<p>The state needs to demonstrate that its innovative assessment system produces yearly, student-level annual determinations that are comparable across LEAs and to the federally required statewide assessments and for each subgroup of students as compared to the results for such students on federally required state</p>	<p>States should provide sound evidence of comparability at the assessment system level that reflects an appropriate balance between flexibility and standardization. Some examples of the types of evidence and documentation that states should discuss regarding their claims of comparability include:</p> <ul style="list-style-type: none"> ➤ A plan for developing achievement level descriptors. ➤ A defensible plan for setting standards 	<p>The state must first articulate and provide a defensible rationale for its explicit definition and target(s) of comparability. The state must detail the methods and analytic techniques by which it will evaluate the intended comparability claims and how it will use these results for continuously improving its system. The state’s approach must address the multiple levels of the system from comparability of scoring specific assessments to comparability of annual determinations (e.g.,</p>

<p>assessments.</p>	<p>(cutscores) that are aligned with the achievement level descriptors.</p> <ul style="list-style-type: none"> ➤ A plan to evaluate claims of comparability using one or more auditing techniques. 	<p>proficiency) both within the pilot districts and between pilot and non-pilot districts.</p>
<p>3. Statewide System – Scalability</p>		
<p>If the state is proposing to administer the innovative assessment system initially in a subset of local educational agencies, the state must have a logical plan to scale up the innovative assessment system statewide in the State’s proposed demonstration authority period (which cannot exceed 5 years) or at the end of an additional 2-year extension period.</p>	<p>The statewide sustainability plan must include a mechanism for assessing district readiness to join the innovative assessment and accountability system and approving district entry into that system. A high quality and thoughtful state sustainability plan will build in the allowance for re-evaluation and course correction at key junctures throughout the timeframe. Giving states the freedom to reconsider their plans for scaling the system after they have had the benefit of experience with pilot districts will be crucial for the success of the innovation.</p>	<p>States will have to develop a strategy for moving beyond the initial set of pilot districts. States should approach scalability very deliberately and provide a clear plan for ensuring that all districts are able to acquire the knowledge and skills necessary for implementing the instructional and assessment system envisioned in the pilot. Scalability of any major reform effort is challenging and susceptible to failure. States should consult with the literature on school and organizational reform to develop their plans for scaling the pilot.</p>
<p>4. Demographic Similarity</p>		
<p>The state can describe how the inclusion of additional local educational agencies will help the state make progress toward achieving high-quality and consistent implementation across demographically diverse local educational agencies. The state can also describe how it will ensure that the participating local educational agencies, as a group, will be demographically similar to the state as a whole by the end of the state’s demonstration authority period.</p>	<p>States must submit information describing the demographic characteristics of its pilot districts and the state. Further, states must describe the differences in the demographics between the pilot districts (collectively) and the state and provide a plan for adding any major gaps as part of the state’s plan for recruiting additional districts to participate in the pilot. When evaluating the demographic similarity of a subset of districts to the state as a whole, we urge states and ED to employ reasonable tolerance levels (e.g., $\pm 10\%$) when evaluating the representation of specific student groups.</p>	<p>Related to its planning for scaling the initiative, the state must consider the demographic and geographic representation of districts participating in the pilot. The state should define its criteria for determining when the pilot districts meet the demographic diversity and similarity targets and when they would fall short.</p>