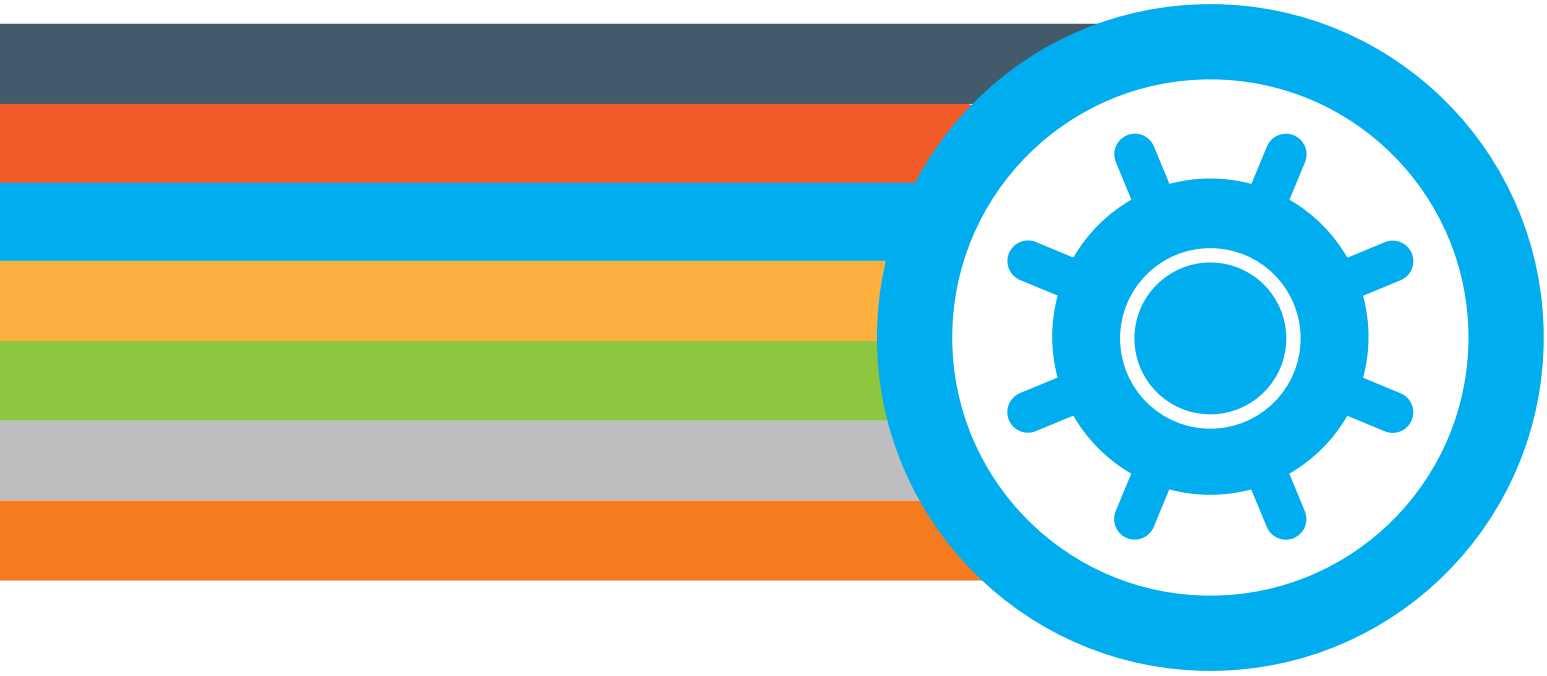


ADDRESSING ACCOUNTABILITY ISSUES INCLUDING COMPARABILITY

in the Design and Implementation of an Innovative Assessment and Accountability System

July 2016

KnowledgeWorks

Center for Assessment

nellie mae EDUCATION FOUNDATION

Acknowledgements

Thanks to generous support from the Nellie Mae Education Foundation, KnowledgeWorks, and the National Center for the Improvement of Educational Assessment (Center for Assessment) have partnered to help states better understand and leverage the new Innovative Assessment and Accountability Demonstration Authority authorized under the Every Student Succeeds Act (ESSA). The goal of this partnership is to help states identify and explore a set of readiness conditions that are critical to the development of a high quality application and implementation process under this new authority. While we share a history of advocacy for next generation assessments, our organizations each bring a unique perspective to this work. KnowledgeWorks focuses on policy development, partnering with states, districts, and educators to identify and remove policy barriers that inhibit the growth of personalized learning. The Center for Assessment specializes in the design of assessment and accountability systems, helping states, districts, and other entities improve the quality of these systems and maximize student success.

Lyons, S., Marion, S.F., Pace, L., & Williams, M. (2016). Addressing Accountability Issues including Comparability in the Design and Implementation of an Innovative Assessment and Accountability System. www.knowledgeworks.org and www.nciea.org.

Table of Contents

Introduction

This is the third in a series of policy and practice briefs produced by KnowledgeWorks and the National Center for the Improvement of Educational Assessment (Center for Assessment) designed to assist states in thinking through the opportunities and challenges associated with flexibility provided under the Every Student Succeeds Act (ESSA).[1] These briefs help define "Readiness Conditions" for states considering applying for and successfully implementing an innovative assessment and accountability system as defined by the Demonstration Authority opportunity under ESSA. In addition to those that have already been published, the following briefs will be released over the next few months:

Supporting Educators and Students through Implementation of an Innovative Assessment and Accountability System

Evaluating and Continuously Improving an Innovative Assessment and Accountability System

Establishing a Timeline and Budget for Design and Implementation of an Innovative Assessment and Accountability System

Building Capacity and Stakeholder Support for Scaling an Innovative Assessment and Accountability System

[1]Brief #3 in a series of policy and practice briefs designed to help states prepare for the ESSA Assessment and Accountability Demonstration Authority. We are grateful to the Nellie Mae Foundation for their generous support of this project.

Purpose

The Innovative Assessment and Accountability Demonstration Authority (hereafter known as the "innovative pilot" or the "Demonstration Authority") provides states with an opportunity to collaborate with a sample of local districts to pilot a new kind of assessment and accountability system within the state. This system does not have to rely on statewide, standardized assessments as the sole indicator of student achievement, but instead may pilot different types of non-standardized assessments (e.g., instructionally embedded assessments, performance tasks) that may provide for some degree of local flexibility. Because states must incorporate assessment results from the pilot districts into the state accountability system alongside the results generated from the non-pilot districts, the assessment system must meet all of the same technical requirements as the state standardized assessments—e.g., alignment, validity, reliability, accessibility.[2] Additionally, because the innovative pilot will take time to scale statewide, the state must ensure that the assessment systems are producing comparable results within pilot districts, among pilot districts, and importantly, across pilot and non-pilot districts.

The purpose of this brief is to support states in planning for a successful Demonstration Authority application by providing key conceptual and technical considerations related to promoting and evaluating comparability in an innovative assessment and accountability pilot. We begin with a discussion of alignment to the state's theory of action so the pilot focuses on the intended goals of the system. Next, we define comparability in the era of ESSA flexibility, and lay the groundwork for a common understanding of how evidence of comparability differs depending on the nature and use of the reported scores. We then delve deeply into how states could approach comparability from a design perspective, providing detailed examples of processes that states could use to support their intended comparability claims. We additionally provide descriptions of the state and local roles for ensuring comparability. Lastly, we provide a case study that details a key comparability practice from the innovative assessment and accountability system in New Hampshire.

[2]For detailed information regarding the technical quality considerations of an innovative pilot, please refer to Brief #2 Ensuring and Evaluating Assessment Quality in the Design and Implementation of an Innovative Assessment and Accountability System.

Alignment to Theory of Action

As emphasized in the Project Narrative brief, the importance of a clear articulation of the state vision and the associated theory of action for attaining that vision cannot be overemphasized. Comparability is a critical goal whenever assessments are being used for accountability. This is especially true when states have incorporated some degree of local flexibility into its assessment systems. Providing for comparability within the initial design conceptualization of the system will be crucial to the success of the pilot.

In order to design and administer meaningful assessment that will change the way instruction and learning occurs in the classroom, local educators will need to engage in rich discussions about what deep learning looks like for every grade level and content area. For example, defining the expectations for student performance in a competency-based education model requires that educators across the state have shared definitions about both the content standards and the required evidence for evaluating student competence relative to the content standards. Therefore, the beginnings of a comparability argument are baked into the learning system of the innovative pilot that the assessment and accountability systems must capture. This brief provides examples of how states can achieve the goal of comparability by planning for it in the pilot design and the processes and audits that comprise the new assessment and accountability system. Each of these design features should be born out of an alignment with the overall theory of action for how learning is changing within the state, and how the pilot will ultimately bring about that change.

Defining Comparability

In educational measurement, comparability is usually premised on the notion of score interchangeability. If scores can be used interchangeably, that means the scores support the same interpretations about what students know and can do relative to the assessed content. Comparability is an accumulation of evidence to support claims about the meaning of test scores and whether scores from two or more tests can be used to support the same inferences and uses. While it is typical in the United States to support comparability by standardizing testing conditions (e.g., administration, scoring), we must acknowledge that score comparability is not necessarily at odds with flexibility.[3] As an example, we provide accommodations for standardized assessments because we believe this type of "flexibility" actually *improves* our claims of score comparability by removing barriers to the assessed content. Just like changing the administration conditions for students with different abilities supports our notion of comparability, it could be argued that changing the mode of assessment (e.g., performance-based assessment) will provide better information about what students know and can do for students in different educational settings (e.g., competency-based), than we could glean from traditional standardized assessments.

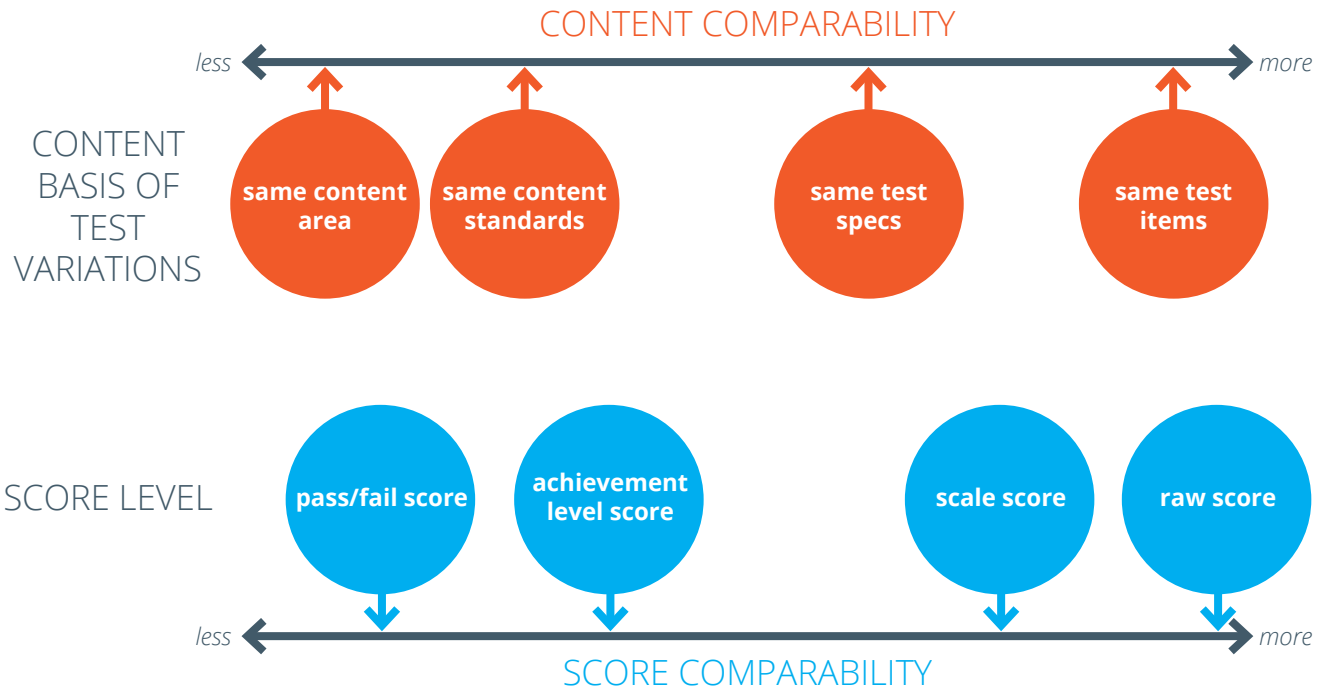
Comparability is…	Comparability is NOT…
An evidence-based claim	Necessarily at odds with flexibility
Score-based	A single number (e.g., an equating constant, or a linking error)
A continuum	The same for every assessment (e.g., the evidence required will differ)

Because claims of comparability are inherently tied to the interpretations and uses of the scores, comparability rests on what is being reported. This means that evidence used to support claims of comparability will differ depending on the nature (or grain-size) of the reported scores. For example, supporting claims of raw score interchangeability—the strongest form of comparability—would likely require the administration of a single assessment form with measurement properties that are the same across all respondents (i.e., measurement invariance). Any state assessment system with multiple assessment forms fails to meet this level of score interchangeability. Instead, the design of most state assessment systems aims to be comparable enough to support scale score interchangeability. This level of comparability typically requires that multiple test forms are designed to the same blueprint, administered under almost identical conditions, and scored using the same rules and procedures. Still, many states continue to struggle to meet this level of comparability (e.g., challenges with multiple modes of administration—paper-based, computer-based, and device-based). In this way, comparability is an evidence-based argument, and the strength of evidence needed will necessarily depend on the type of score being supported. As shown in Figure 1, comparability lies on a continuum that is based on both the degree of similarity in the assessed content and the granularity of the score being reported.[4]

[3]Gong, B., & DePascale, C. (2013). *Different but the same: Assessment "comparability" in the era of the Common Core State Standards.* Washington, DC: The Council of Chief State School Officers.

[4]Winter, P. (2010). Comparability and test variations. In P. Winter (Ed.), *Evaluating the comparability of scores from achievement test variations* (pp. 1–11). Washington, DC: Council of Chief State School Officers.

Figure 1. Comparability Continuum[5]



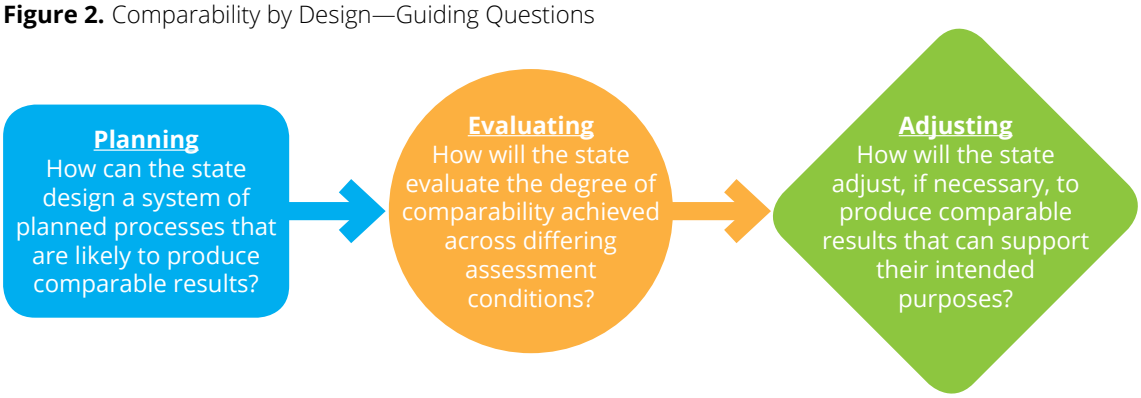
The Demonstration Authority requires states to ensure that summative "annual determinations" (e.g., performance levels) are comparable. Comparability, therefore, must exist at the level of the annual determinations. This means that if a student is determined to be "proficient" relative to the grade-level content standards in one district in the state, had that student been assigned to another district's assessment system (either pilot or non-pilot) he or she could expect to be proficient. To support claims of comparability at the annual determination level, any pilot program will need to build in a number of processes and auditing mechanisms to create a strong evidence base for supporting the claims of comparability within each pilot district, among pilot districts, and across pilot and non-pilot districts.

[5]Winter, P. (2010). Comparability and test variations. In P. Winter (Ed.), Evaluating the comparability of scores from achievement test variations (p. 5). Washington, DC: Council of Chief State School Officers.

Comparability by Design

The methods for gathering evidence to support a comparability claim are not a series of analyses, but rather begin with the design of the innovative assessment and accountability pilot itself. In traditional standardized assessment programs, comparability is generally established by **planning** for it in the assessment system design (e.g., addressing the same learning targets in the same ways, embedding items), **evaluating** the degree of comparability achieved (e.g., analyses of differential item functioning), and then, if necessary, **adjusting** the measurement scales to account for differences (e.g., equating). Providing evidence of comparability for the innovative assessment system will require discussion related to each of these steps, even if the methods related to each step are necessarily different. Three key questions shown in Figure 2 below can guide the process of designing a pilot to produce comparability results—*comparability by design*.

Figure 2. Comparability by Design—Guiding Questions



The order of these guiding questions is exceedingly important. It will not be possible to evaluate the degree of comparability that these scores produce under different assessment systems if comparability has not been carefully planned for (e.g., through common items or tasks). Similarly, it will not be possible to calibrate results if the nature and magnitude of the adjustments are not known through a careful evaluation of the degree of comparability achieved across assessment systems. No amount of evaluation and calibration can fix a system that has not been carefully designed to produce scores that are likely to be comparable. Thus, garnering evidence to support comparability of the assessment system results will require thoughtful planning of the program processes that will promote comparability, and the program monitoring mechanisms that will evaluate comparability. Examples of how this could be done to support claims of comparability of results within pilot districts, among pilot districts, and across pilot and non-pilot districts are provided on the next few pages.

Methods for Establishing a Strong Evidence Base to Support Claims of Comparability

The integration of comparability supports and audits throughout the design of the pilot is a sign of strength in any innovative assessment and accountability system. The pilot must be designed to support the validity and comparability of the annual determinations.

For a system that relies on local flexibility in the assessments administered to support annual determinations, comparability will rest on evidence regarding the local scoring ***within districts****, the performance standards for student achievement* ***among pilot districts****, and finally, the annual determinations* ***across pilot districts*** *and* ***non-pilot districts****. Gathering evidence at each of these levels will be essential for supporting the claims of comparability, and ultimately supporting the validity of the system as a whole.*

Examples of the activities and audits that could occur at the three levels are summarized in Figure 3 and described in detail below.

Within-District Comparability in Expectations for Student Performance
States must plan for efforts to improve and monitor within-district comparability. Promoting and evaluating consistency in educator scoring of student work within districts should be accomplished using multiple methods, and may include one or more of the following three example methodologies:

1) Within-district calibration sessions resulting in annotated anchor papers.
   Providing training and resources for participating districts to hold grade-level calibration sessions for the scoring of common or local assessments is the first step for within-district calibration. Teachers would bring samples of their student work from one or more assessments that represent the range of achievement in their classrooms and will then come to a common understanding about how to use the rubrics to score papers and identify prototypical examples of student work for each score point on each rubric dimension. The educators annotate each of the anchor papers documenting the groups' rationale for the given score-point decision. These annotated anchor papers are then distributed throughout the district to help improve within-district consistency in scoring. Additionally, if this work is done using an assessment that is common across districts, the anchor papers could be vetted and shared across districts to simultaneously improve cross-district calibration in scoring.

2) Within-district estimates of inter-rater reliability.
   External audits of the consistency in scoring could be achieved by asking each district to submit a sample of papers from each assessment (or a sample of assessments) that have been double-blind scored by teachers. The collection of double scores could then be analyzed using a variety of traditional inter-rater reliability techniques for estimating rater scoring consistency within-districts (e.g., percent exact and adjacent agreement, Cohen's Kappa,[6] intraclass correlations[7]).

3) Testing the generalizability of the local assessment systems.
   If the design of the innovative pilot involves at least some common tasks and some local tasks for generating annual summative scores, much of the work for gathering evidence of comparability will rely on the use of the common tasks as calibration tools. However, the utility of the common items or tasks for judging the degree of comparability across districts rests heavily on the assumption that within-district or local scoring on the common tasks is representative of local scoring on the local tasks. This assumption requires that findings associated with the common tasks are generalizable across the entire assessment system within each participating district. Therefore, it will be necessary to test this assumption by running generalizability analyses using all of the assessment scores (local and common) within a given district's assessment system. Conducting these analyses has the added benefit of providing an index of score reliability that can be used to support technical quality of the assessment results.

Cross-District Comparability in Evaluating Student Work

The primary goal of a cross-district comparability audit is quality control: to gather evidence of the degree to which there are systematic differences in the stringency or leniency of scoring across participating districts. Depending on the design of the pilot, there are methods for evaluating the degree in comparability in scoring across districts for common assessments and for local assessments. Additionally, the comparability of the results of the assessment system can be evaluated by critically examining bodies of evidence (student work) generated by a cross-district sample of students participating in the innovative assessment system. An example of each of these types of methods is provided below:

1) Social moderation audit with common tasks.
   The design of social moderation audits can be modeled after a number of international examples; one that may be particularly useful is Queensland, Australia where externally-moderated school-based assessments replaced external standardized assessments.[8] If all students in the participating pilot districts are taking at least one common performance task, then student scores on these tasks can be used to determine the degree of comparability of teacher judgments about the quality of student work across districts. A consensus scoring social moderation method could involve pairing teachers together; each representing different districts, to score student work samples from yet a third district. After training and practice, both judges within the pairs are asked to individually score their assigned samples of student work and record their

[6]Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.

[7]Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420.

[8]Queensland Studies Authority. (2014). School-based assessment: The Queensland assessment. Queensland, Australia: The State of Queensland (Queensland Studies Authority).

scores. Working through the work samples one at a time, the judges discuss their individual scores and then come to an agreement on a "consensus score." The purpose of collecting consensus score data is to estimate what might be considered analogous to a "true score," which is used as a calibration weight. These consensus scores are then used in follow-up analyses to detect any systematic, cross-district differences in the stringency of standards used for local scoring. If systematic differences are detected, the project leaders can make defensible decisions about calibrating (or making adjustments to) the district-specific performance standards.

2) Social moderation audit with local tasks.

The comparability of local tasks measuring the same or similar knowledge and skills can be evaluated using a rank-ordering social moderation technique. In the United Kingdom (UK), the results of written exams are used to inform decisions about post-secondary job and university placements. However, across the UK, different awarding bodies (or examination boards) are responsible for creating their own written examinations. Therefore, social moderation audits are used to ensure the standard for post-secondary placements is comparable across awarding bodies. One approach for ensuring comparability is a rank-ordering social moderation method.[9] The rank ordering method involves asking trained judges to rank-order samples of student work within a number of pre-designed packets. The packets are grouped by similar overall score, which is blind to the reviewers. The work within the packets is arranged and distributed across judges in a way that allows for each sample of work to be compared with all other student work receiving similar scores and ranked by more than one judge. The rank-order data resulting from the judges can then be transformed into paired-comparison data that can be used to estimate a Thurstone scale. An indicator of relative district stringency and leniency in scoring can be derived from comparing the Thurstone scale scores with the local scores of each sample of student work.

3) Validating the performance standards with a body-of-work method.

Cross-district comparability rests on the notion that the results of the assessment system in one district carry the same meaning and can be interpreted and used in the same way as results of the assessment system in another district. Since the innovative pilot will likely involve a degree of local variability across the assessment systems in the pilot districts, the assumption of comparable results must be verified. One way to validate the district standards is to engage in a student work-based standard setting method such as the Body of Work method or some variation thereof.[10] The body of work method requires teachers or other judges to review the portfolios of student work and make judgments about student achievement relative to the achievement level descriptors. These teacher judgments are then reconciled with the reported achievement levels as an additional source of validity evidence to support the comparability of the annual determinations across pilot districts.

[9]Bramley, T. (2005). A rank-ordering method for equating tests by expert judgment. *Journal of Applied Measurement*, 6(2), 202–223.

[8]Kahl, S.R., Crockett, T.J., DePascale, C.A., & Rindfleisch, S.L. (1995, June). Setting standards for performance levels using the student-based constructed-response method. Paper presented at the annual meeting of the American Research Association, San Francisco, CA.

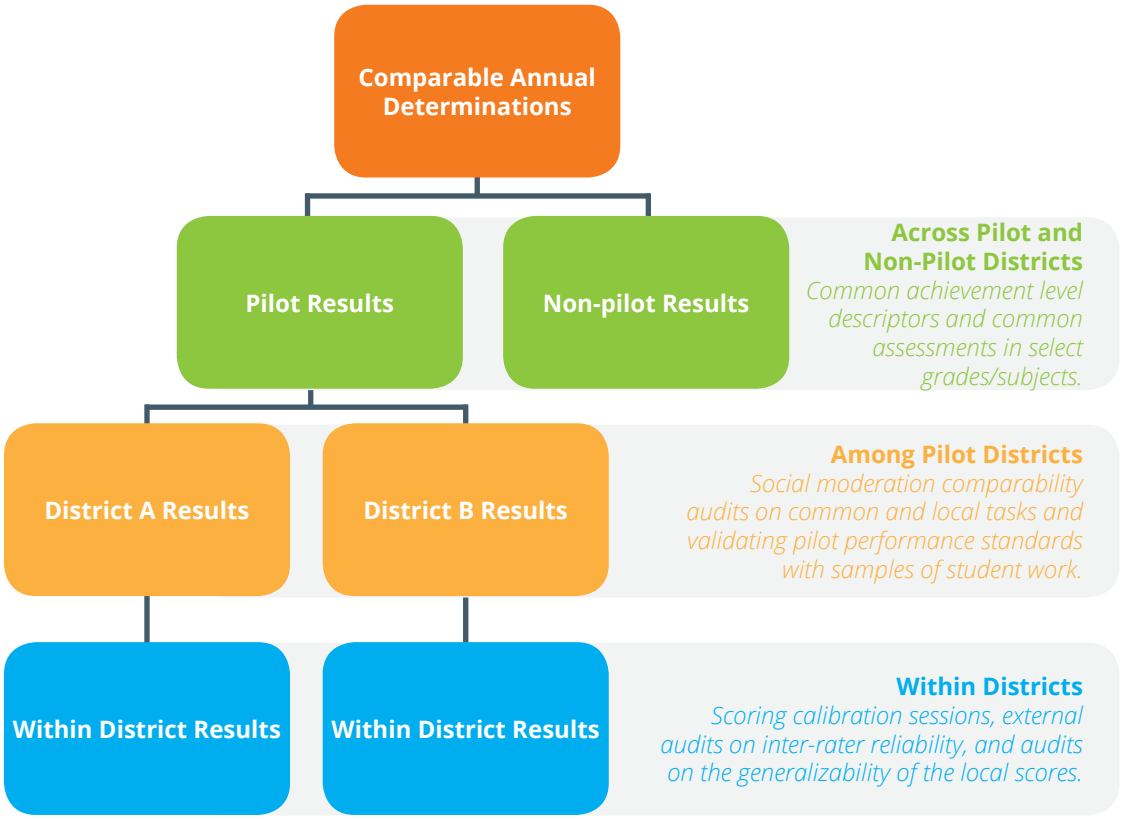
Comparability of Annual Determinations Across Pilot Districts and Non-Pilot Districts
The accountability uses for the assessment system results rests on the comparability of annual determinations. Therefore, the comparability claims for the innovative pilot will apply to the reported performance levels (as opposed to scale scores for more traditional assessment models). The comparability processes and audits that occur at both the local, within-district level, and the cross-district level are all in an effort to support the claim of comparability in the annual determinations. However, if the pilot is not statewide, a major ESSA comparability requirement is that the pilot system results are comparable with the non-pilot district results. The following are examples of procedures that could be used to formally promote and evaluate the comparability of the annual determinations across both pilot and non-pilot districts:

1) Setting standards using common achievement level descriptors (ALDs).
   Achievement level descriptors are exhaustive, content-based descriptions that illustrate and define student achievement at each of the reported performance levels. Detailed ALDs are typically developed by teams of content experts and educators to be used for the purposes of setting criterion-referenced performance standards (i.e., cutscores) for an assessment program. The use of common ALDs across the pilot and non-pilot assessment systems will support shared interpretations of performance relative to the content standards; and ultimately, through the chosen standard setting procedures, provide evidence for the comparability of the performance standards across the two assessment systems. If the selected standard setting methods across the two programs rest heavily on common ALDs, then having common ALDs will serve as a foundation for the inference that the resulting achievement levels carry the same meaning and can be used to support the same purposes (i.e., accountability and reporting).

2) Administering a common standardized assessment to a sample of students in both pilot and non-pilot districts.
   Importantly, the degree of comparability of the annual determinations across the two assessment systems within the state can be directly evaluated by administering an assessment that is common across the two programs to a sample of students. For example, a state could administer the statewide standardized assessment to students in select grade levels and subjects within the pilot districts. The comparability of the annual determinations between pilot and non-pilot districts could then by evaluated by directly comparing annual determinations for the students that participated in both assessment systems. By calculating two sets of annual determinations for these students, the state will have both traditional and innovative data points for some of the students in each pilot district. The degree of agreement between the two sets of annual determinations could then be analyzed to provide further evidence regarding the comparability of the interpretations of the reported achievement levels, or if systematic differences are detected, inform decisions about calibrating results to provide for comparability when appropriate.

We note however, that just because we have two sets of data to evaluate the performance of students across different settings, it does not mean that the results should be equivalent. For example, if approximately 55% of the students were scoring in Levels 3 and 4 on the state standardized assessment, that does not mean we should expect exactly 55% of the students to be classified in Levels 3 and 4 on the innovative pilot assessments. There could be very good reasons why the results would differ in either direction. For example, if a state is using an innovative performance assessment model in the pilot districts, these assessments may be capturing additional information relative to real-world application and knowledge transfer that provides for more valid representations of the construct than possible with traditional standardized assessments. That said, states should be able to explain these discrepancies in terms of their theories of action. Further, it would be hard to explain significant variations between the two sets of results, especially if such variability was found in only a subset of the pilot districts.

Figure 3. Establishing an Evidence-Base for Comparable Annual Determinations



State and District Roles

Because districts are likely accustomed to having complete authority over their local assessment systems, and states too are accustomed to garnering sole responsibility for the state assessment and accountability system, navigating a new partnership to balance the needs of both parties must be carefully planned with newly established lines for open communication. The state needs to clearly articulate the full scope of the pilot expectations including the evidence and data collection protocols that will be necessary to support the comparability of the assessment system. While the state's innovative pilot is still new and in the process of improving and scaling, state and district leaders must work together to find a balance between the need to collect the right evidence, and the reality of the local burden for gathering and organizing the necessary documentation. Adaptability will be a key characteristic for success in the state and district leadership as the nature and scope of the data collection will likely evolve over the course of the first few implementation years. Key comparability considerations for both state and local officials are provided in Figure 4 below.

Figure 4. Key Comparability Considerations for State and Local Officials

State Considerations	Local Considerations
How will the state support districts in establishing strategies for promoting within-district comparability?	How will districts ensure that their local assessment systems produce results that are internally comparable within and across grade levels?
What design features and processes will the state need to put into place to ensure that comparability is promoted and evaluated? What resources will the state provide to help districts successfully engage in the comparability processes and audits?	What resources do local districts need to dedicate to assessment design, delivery, and reporting to ensure they can participate in the comparability processes and audits within the pilot design?
In what ways will the state ensure the results of the innovative assessment system carry the same meaning and can be used in the same way as the results of the non-pilot assessment system within the state?	In preparing to administer a new, innovative assessment system, how will the local districts plan for the sample of students that will be expected to continue to take the statewide standardized assessment as an external assessment audit?
How will common understandings about the nature of the pilot and the purposes of the extensive data collection efforts be clearly communicated to the local district leaders, principals, teachers, students, and parents?	
What external support may be necessary at the state and local levels to ensure the design and implementation of the pilot is establishing a strong evidence base for promoting, evaluating, and calibrating/establishing comparability?	

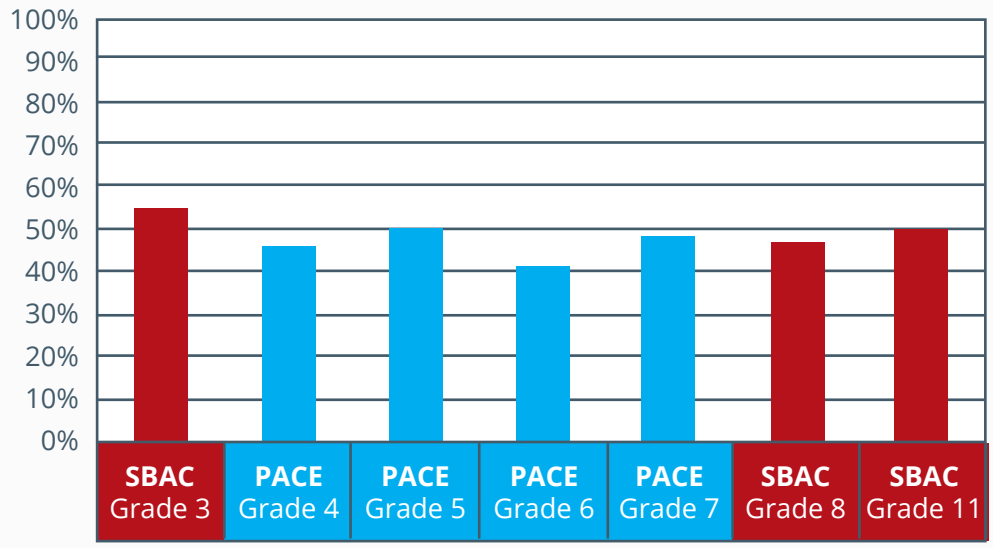
STATE EXAMPLE

The New Hampshire Department of Education received a waiver from the United States Department of Education in March 2015 that allowed the use of a combination of local and common performance assessments in lieu of a statewide standardized assessment to make annual determinations of student proficiency. This pilot is called the Performance Assessment of Competency Education or PACE. The annual determinations are used to inform parents and stakeholders of students' knowledge and skills relative to the state-adopted competencies and are also used in the statewide school accountability system. The PACE pilot gathers multiple sources of evidence to support the claims that the determinations are comparable within each district, across the different PACE districts, and with non-PACE districts. While the PACE model engages in many of the processes and audits discussed throughout this brief to establish comparability, the contrasting groups standard setting process is an additional method of achieving comparability used within PACE.

The standard setting method is used to determine the location in the score distributions of the appropriate "cut points" for establishing achievement levels. New Hampshire needed to choose a standard setting method that relied heavily on the common achievement level descriptors to ensure that the standards set across districts (both PACE and non-PACE) were comparable. An examinee-centered judgmental method called contrasting groups was used to establish cut points. This standard setting method involves judgments from panelists about the qualifications of the examinees based on prior knowledge of the examinee. PACE teachers were asked to make judgments about the achievement level that best described each of their students from the previous year. The contrasting groups standard setting methodology then involves comparing the PACE competency scores with student placements into achievement levels in order to determine cut scores that would accurately classify the highest percentage of students.

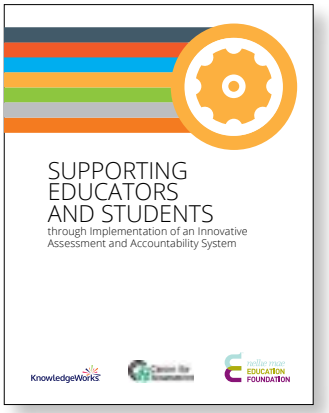
The results of the standard setting method provided for comparable annual determinations by a number of measures, including a comparison with the distribution of achievement for those grade levels tested using the Smarter Balanced statewide assessment. Figure 5 below shows the percent of students deemed proficient in ELA for the PACE and Smarter Balanced grade levels for the pilot districts. While there are some small fluctuations in performance across the grade levels, they do not appear to be primarily a function of the assessment system and are not atypical from those fluctuations seen in districts administering the same assessment system across grade levels. New Hampshire is committed to ensuring comparability and continuous improvement, so trends in student achievement will be closely monitored when results from Year 2 are available.

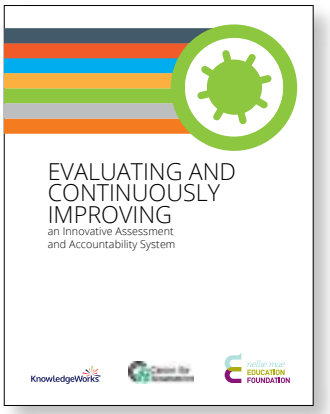
Figure 5. PACE Districts Percent Scoring at Levels 3 & 4 in ELA

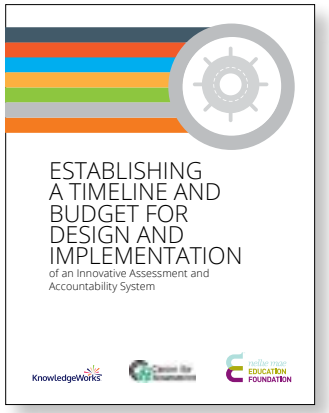


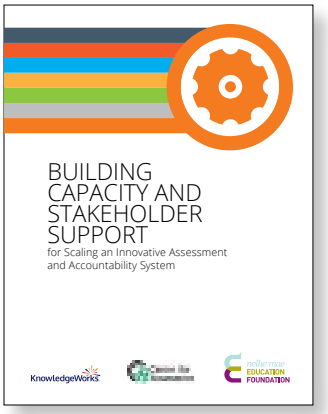
Summary

This third brief in our series of State Readiness Conditions publications is designed to help state leaders recognize the unique comparability opportunities associated with the Innovative Assessment and Accountability Demonstration Authority flexibility afforded under the recently passed ESSA. We also present a realistic picture of the challenge associated with planning for and gathering comparability evidence within an innovation pilot. This brief clarifies the definition of comparability and provides in-depth examples of the design features and implementation processes that would support claims of comparability under a Demonstration Authority. KnowledgeWorks and the Center for Assessment will continue to support states through the summer and fall with additional briefs on topics related to fleshing out the design of a Demonstration Authority application, including:









Additional Support

KnowledgeWorks and the Center for Assessment are available to help states as they explore, design, and implement next generation assessment systems. Contact information for our organizations is listed below.

KnowledgeWorks can help states, districts, and other interested stakeholders establish the policy environments to support personalized learning at scale. The organization's expertise spans the federal, state, and district levels, supporting states with strategies to leverage current policy opportunities, remove existing policy barriers, and develop new policies that will help states create an aligned policy environment to support personalized learning. To learn more, contact the following people:

For State Policy and Alignment:
Matt Williams
Vice President of Policy and Advocacy
Williamsm@knowledgeWorks.org

For Federal Policy and Alignment:
Lillian Pace
Senior Director of National Policy
pacel@knowledgeworks.org

The **Center for Assessment** strives to increase student learning through more meaningful educational assessment and accountability practices. We engage in deep partnerships with state and district education leaders to design, implement, and evaluate assessment and accountability policies and programs. We strive to design technically sound policy solutions to support important educational goals. The Center for Assessment's professionals have deep expertise in educational measurement, assessment, and accountability and have applied this expertise to assessment challenges ranging from improving the quality of classroom assessments to ensuring the technical quality of state's large-scale achievement tests and ultimately to designing coherent assessment and accountability systems.

For Assessment and Accountability System Design and Strategic Implementation:
Scott Marion, Ph.D.
Executive Director
smarion@nciea.org

For Technical Quality and Comparability Design and Analyses:
Susan Lyons, Ph.D.
Associate
slyons@nciea.org

For Assessment Quality and Performance Assessment Development:
Jeri Thompson, Ed.D.
Senior Associate
jthompson@nciea.org

About Us

KnowledgeWorks is a nonprofit organization dedicated to advancing personalized learning that empowers every child to take ownership of their success. With nearly 20 years of experience exploring the future of learning, growing educator impact and working with state and federal policymakers, our passionate team partners with schools and communities to grow a system-wide approach to sustain student-centered practices so that every child graduates ready for what's next. www.knowledgeworks.org

The National Center for the Improvement of Educational Assessment, Inc. (Center for Assessment) is a Dover, NH based not-for-profit (501(c)(3)) corporation that seeks to improve the educational achievement of students by promoting enhanced practices in educational assessment and accountability. The Center for Assessment does this by providing services directly to states, school districts, and other organizations regarding the design, implementation, and evaluation of assessment and accountability systems. As a non-profit organization committed to the improvement of student learning, the Center for Assessment maintains a strong "open-source" ethic in terms of distributing its many creations and inventions. For example, the Center has developed many tools related to alignment methodology, student growth analyses, student learning objectives, comparability methods for innovative assessment systems, and validity evaluation that it provides freely to its clients and other non-commercial entities. www.nciea.org

The Nellie Mae Education Foundation is the largest philanthropic organization in New England that focuses exclusively on education. The Foundation supports the promotion and integration of student-centered approaches to learning at the middle and high school levels across New England—where learning is personalized; learning is competency-based; learning takes place anytime, anywhere; and students exert ownership over their own learning. To elevate student-centered approaches, the Foundation utilizes a four-part strategy that focuses on: building educator ownership, understanding and capacity; advancing quality and rigor of SCL practices; developing effective systems designs; and building public understanding and demand. Since 1998, the Foundation has distributed over $180 million in grants. For more information about the Nellie Mae Education Foundation, visit www.nmefoundation.org.