

**Paul K. Leather**

Deputy Commissioner of Education

603-271-3801

**Virginia M. Barry, Ph.D.**

Commissioner of Education

Tel. 603-271-3144

## STATE OF NEW HAMPSHIRE

**DEPARTMENT OF EDUCATION**

**101 Pleasant Street**

**Concord, N.H. 03301-3860**

**FAX 603-271-1953**

[**www.education.nh.gov**](http://www.education.nh.gov)

# New Hampshire Performance Assessment of Competency Education: Waiver Extension Request to the United States Department of Education

## June 20, 2016

## Introduction

The New Hampshire Department of Education (NHDOE) is pleased to provide the following information in support of the State’s request to extend its waiver under the Elementary and Secondary Education Act. This memo is our formal request to allow NH DOE to continue with the Performance Assessment of Competency Education (PACE) pilot in nine (9) NH school districts. NH DOE has previously submitted many documents, starting with our original proposal in November 2015 and continuing with many of our progress reports through this past March. Therefore, we do not repeat that information here. Rather, we focus on the request from Ann Whalen, Senior Advisor to the Secretary, Delegated the Duties of the Assistant Secretary of Elementary and Secondary Education, to provide the following information:

* The Participating Districts and Schools planning to implement NH PACE in SY 2016-2017 (Tier 1), as well as those in preparation, (Tiers 2 and 3).
* A description of how New Hampshire will ensure that each measure of student achievement, including unit summative assessments, used to determine student proficiency on a PACE assessment either: (a) follows the PACE performance task model (i.e., development, administration and scoring processes and procedures), and/or (b) is vetted through the state's formal process for ensuring its technical quality for its intended purposes prior to use in a PACE assessment; and submit plans to evaluate and improve such quality controls during 2016-17.
* A plan for documenting the comparability of the meaning and interpretation of the PACE assessment results across PACE districts, including planned analyses comparing the results for individual students taking PACE and Smarter Balanced assessments within and across years and other analyses regarding topics construct validity and year-to-year comparability of results.

We discuss and provide evidence regarding each of these requests below.

**Participation**

In School Year 2016 – 2017, the following Tier 1 SAUs/Districts/Charter Schools will participate in implementing NH PACE:

* Rochester, Sanborn, Epping, Souhegan, Concord, Pittsfield, Monroe, Seacoast Charter School, and SAU 35 White Mountains

In addition, the following SAUs/Districts/Schools are in preparation to enter NH PACE, and are currently in either Tier 2 (primarily involved in professional development in quality performance assessment) or Tier 3 (initial preparation for NH PACE):

* Allenstown, Deerfield, Fall Mountain, Plymouth, SAU 23 North Haverhill, SAU 58 Groveton, Manchester, Rollinsford, Ashland, SAU 39 Amherst and Mont Vernon, Virtual Learning Academy Charter School, Newport, and SAU 1 Conval (Peterborough and surrounding towns)

Tier 1 SAUs/Districts/Schools represent 14,280 students, or approximately 8% of the NH student population. All three Tiers represent 37,935 students, 22 SAUs and two Charter Schools, or approximately 22% of the Supervisory Unions and 21% of the student population in the state.

## Supporting Local Assessment Quality

New Hampshire is committed to raising the bar for all students by defining college and career-readiness to encompass the knowledge, skills, and work-study practices that students need for post-secondary success including deeper learning skills such as critical thinking, problem-solving, persistence, communication, collaboration, academic mindset, and learning to learn. However, NH’s educational leaders recognize that the level of improvement required cannot occur with the same type of externally-oriented assessment model that has been employed for the past 12 years. In fact, the state argues that the current system is likely an impediment for moving from good to great. PACE is a shift to a model with significantly greater levels of local design and agency to facilitate transformational change in performance. As part of this shift in orientation, the state believes there are more effective ways to assess student learning for informing and improving students’ progress towards college and career ready outcomes. The State argues that a competency-based approach to instruction, learning, and assessment is philosophically and conceptually related to this internally-oriented approach to accountability and can best support the goal of significant improvements in college and career readiness. The information provided by performance-based assessments of competencies can be used to inform both instructional improvement and the annual determinations of achievement for accountability purposes. The PACE pilot is based on the belief that a rich system of local and common (across multiple districts) performance-based assessments is necessary for supporting deeper learning as well as allowing students to demonstrate their competency through multiple performance assessments in a variety of contexts. Accomplishing these learning goals requires an assessment system tailored to local contexts, but since these local assessments contribute to state accountability indicators, it is important to document that students are assessed on the required content standards and done so with relatively high quality assessments. Ensuring high quality assessments while supporting local agency is a needle that NH DOE is trying to thread. However, all PACE Tier I districts have agreed to continue to transition to relying on state-vetted performance assessments for annual determinations. That said, it is still impractical for NH DOE to evaluate the quality of every local assessment immediately. Further, such an effort would threaten the local agency and assessment literacy that NH DOE is trying hard to foster. Finally, we continue to emphasize the importance of the assessment system and not on every single assessment! For example, we know that the reliability of a single 10-item test is likely unacceptably low, but the reliability of ten, 10-item tests when taken together would be incredibly high, close to what we would expect from a 100-item test.

Given this context, NH DOE is employing a multi-faceted approach to supporting local assessment quality that we argue fully meets ED’s intentions, but does so in practical and sustainable ways. It is the state’s responsibility to ensure that the annual determinations resulting from the innovative system of assessments are aligned, generated as a result of high quality assessments, reliable, and valid for use in the accountability system. To that end, the state is taking the following steps in the 2016-2017 school year:

1. **Alignment.**

**Review of Local Assessment Maps**. NH DOE has collected assessment maps from all PACE districts for all grades and subjects covered under the PACE pilot as a way to document that all content standards are addressed in the assessment system, which is a window into the degree to which all students are provided a meaningful opportunity to learn the required standards for that grade level (see Appendix A for an example). In the 2016-2017 academic year, the layout and content of the assessment maps will be standardized across districts to provide for clear information about the types and number of assessments that measure each of the New Hampshire State Standards and district competencies within each district’s assessment system. It will be re-emphasized to the district leaders that the assessments included in the maps are to be all of the summative assessments that contribute to the individual student end-of-year competency scores. In submitting the district assessment maps, district leaders will understand that any assessment listed in the map may subject to state quality audits at any time. The assessment maps will be reviewed at the state level by assessment experts to evaluate the degree of coverage of the full breadth of the New Hampshire State Standards. Upon review, should any gaps in coverage arise, PACE leaders will work with individual district leaders to ensure those gaps are fully addressed before the end of the academic year.

1. **Technical Quality of the Local Tasks.**

**Local and Common Task Development.** Because new common performance assessments are developed each year and assess different competencies, “retired” common tasks are maintained within an assessment bank for teachers to continue to use to support their local assessment needs. With the expansion of PACE to eight districts in 2015-2016, more teachers were involved in the common task development work. It would have been inefficient to have all of the teachers involved in developing a single task. Therefore, in certain grade spans and subjects where there were enough teachers, we were able to develop extra tasks that will supply the local task bank with very high quality tasks. For 2016-2017, we will engage in a more systematic development process where several additional tasks will be developed for each grade and subject area to help populate the local task bank. These tasks are all evaluated using the Center for Assessment review process to ensure that they are high quality. While teachers have already begun using the task bank to support their local assessment needs, our goal is to have the bank populated with multiple high quality tasks for each competency within 2-3 years. The intention is for PACE school districts to draw on these high quality tasks for local use and thereby ensure that students are being provided opportunities to engage with pre-vetted high quality performance tasks multiple times throughout the year. These tasks may also be shared with the ILN/CCSSO Task Bank.

**Mandatory State Audits of Local Assessments**. In the 2016-2017 academic year, a proportion of each district’s summative assessments will be selected from the district assessment maps and reviewed for quality by assessment experts. These assessments will be reviewed for technical quality, and when quality can be improved, formative feedback will be provided to the districts. The most important review criteria will be the alignment to the standards and the depth at which the standards are being measured. Part of the theory of action of the PACE pilot is that by having students engage in rich, cognitively demanding assessment experiences, instruction and student achievement will improve. State audits of a proportion of local assessments will help to ensure that students have an opportunity to learn the content standards and they are being assessed at a high depth of knowledge. If the audit reveals any systematic problems in the quality of tasks being given at any of the districts, state leaders will work with those districts to ensure that further professional capacity building opportunities are provided to the educators and administrators who are engaging in local assessment design or selection.

**External Evaluation.** The HumRRO evaluation is largely focused on the development, quality, implementation, and scoring of common PACE performance tasks, but it will also document the degree to which the common task development processes have infiltrated the development of local assessments. This is a key part of the PACE theory of action and we look forward to seeing the results from HumRRO’s analyses.

**Local Task Review.** NH DOE has reached an agreement with the Council of Chief State School Officers (CCSSO) Innovative Lab Network (ILN) and the Stanford Center for Assessment, Learning, and Equity (SCALE) to have locally-developed performance tasks go through the ILN/SCALE Task Bank review process. All of the local performance assessment tasks submitted to the NH DOE are required to be in the format of the NH PACE Performance Task Template, including the same components of a quality performance assessment as the common PACE tasks. All submitted local tasks will be reviewed through the ILN/SCALE and the NH DOE process. Assessment tasks that meet the task bank criteria will be included as resources that can be used by anyone with access to the task bank. Assessments that do not initially meet the rigorous criteria will be returned with comments and suggestions for improving the quality, thereby increasing the assessment literacy of the local educators who developed and submitted the tasks. The communication and local review and revision will be managed by NH PACE office and the NH PACE Content Lead Teams. Further, NH DOE has negotiated with ILN/SCALE to have a separate section of the website devoted to PACE local tasks in the development and review process. Currently, NH has over 100 local tasks going through the review process this summer. NH PACE District Leads have agreed to submit locally developed non-common tasks for this review and we anticipate all districts will take advantage of this resource and thus, not only provide a window into the quality of local assessments, but a mechanism for improving local task quality.

1. **Reliability.**

**Generalizability Study.** The PACE Technical Advisory Committee (TAC) recommended conducting generalizability analyses to document the extent to which inferences about comparability and quality from the common task generalize to other tasks in the system. One of the technical challenges when students are assessed on a limited set of classroom assessment evidence is the generalizability of such decisions. For example, would students likely demonstrate similar levels of achievement had they been given a different set of assessment tasks? And how many classroom assessments are needed to provide a stable measure of student achievement? These questions can be evaluated using generalizability theory (Brennan, 1992; Cronbach, Linn, Brennan, & Haertel, 1997; Shavelson & Webb, 1991). The purpose of the preliminary generalizability study is to (1) examine the reliability of generalization from a collection of classroom assessments intended to measure student achievement to the universe of all possible assessments and (2) determine an efficient number of classroom assessments necessary to ensure high reliability of decisions about student achievement made in a school accountability context. Using electronic grade book data provided by one of the eight districts implementing NH’s PACE pilot in 2015-2016, we have begun to examine the generalizability of the individual scores that contribute to achievement estimates (e.g., summative tests, quizzes, projects, performance tasks) in six subject/grade combinations (Table 1). Generalizability theory provides a reliability coefficient called a generalizability (G) coefficient. As with other estimates of reliability, the G coefficient represents the proportion of variability in observed assessment scores attributable to systematic differences in student achievement. Preliminary findings show the reliability of the achievement estimates resulting from the PACE assessment system in the studied district are very high (>.9 for almost all subjects and grades) indicating highly stable student achievement results. The generalizability report from the Center for Assessment will be available by October 2016.

1. **Validity of the Assessment Results.**

**Review of “bodies of evidence” generated from local assessment systems**. As part of validating the annual determinations produced for the 2015-2016 school year, we have collected a “body of evidence” for a small sample of students from each participating district. Throughout the academic year we have asked that each district choose a sample of nine students, representing the range of performance in that district, for one content area per grade level. Teachers are asked to collect samples of student work from those nine students for each of the competencies. This summer, teachers have come together to review the portfolios of student work from districts that are not their own to and make judgments about student achievement relative to the Achievement Level Descriptors. These teacher judgments will be reconciled with reported student performance as an additional source of validity evidence to support the accuracy and comparability of the annual determinations across PACE districts. Additionally, these sets of evidence will be used to document the quality of the local assessments. Researchers at Rutgers University have been given access to the tasks and student work samples submitted for the body of evidence analysis. The intended research outcome is a measure of tasks quality that can be used as evidence of high quality implementation of the instructional and assessment practices embodied in the New Hampshire State Standards. This project will shed light on the fidelity with which the state standards are being taught and assessed in the PACE districts.

## The state believes that through the multi-faceted approach described above, the evidentiary basis for evaluating and supporting the validity of the PACE annual determinations is strong. Not only do the quality control processes outlined in detail above provide the necessary information for monitoring the quality of assessment results, but there are mechanisms in place for course correcting and improving assessment quality throughout.

##

## Comparability

New Hampshire’s Performance Assessment for Competency (PACE) program is currently the first and only assessment and accountability state-led pilot initiative. The New Hampshire Department of Education received a waiver from USED in March 2015 that allows 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. The annual determinations are used to inform parents and stakeholders of students’ knowledge and skills relative to the state-adopted content standards and are also used in the statewide school accountability system. Because the annual determinations are based on the results of a balance of common state assessments and local assessments, the PACE pilot gathers multiple sources of evidence to support the claims that the determinations are comparable across the different PACE districts, comparable with non-PACE districts, and comparable across time. This brief is intended to describe how the PACE pilot plans to continue to gather evidence to support these claims in the coming academic year.

### Defining Comparability

ESSA allows for multiple assessment systems within a state (e.g., advanced assessment option in eighth grade, the nationally-recognized high school assessment option, and the assessment and accountability demonstration authority) so it is useful to define 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 not determined by an equating coefficient or a linking error, but is instead an accumulation of evidence to support the claim that the scores carry the same meaning and can be used to support the same inferences and uses. 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 (Winter, 2010).



*Figure 1. Comparability Continuum (Winter, 2010, p. 5).*

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. 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 the multiple tests 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. The chosen approach for establishing comparability is dependent upon the nature of the assessments and the intended interpretations and use of the reported scores

For the PACE pilot, the “annual determinations” (e.g., performance levels) are the key reported “score.” 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 randomly assigned to another district’s assessment system (PACE or non-PACE) he or she could expect to be proficient. To support claims of comparability at the annual determination level, the PACE pilot program has built in a number of processes and auditing mechanisms. Due to the planned and purposeful nature of these processes and audits, and their integration into the structure of the PACE pilot, we believe—and the initial pilot year work supports—the accumulated evidence-base is strong for supporting claims of score comparability across PACE districts, across PACE and non-PACE districts, and across years.

### Comparability by Design

The integration of comparability supports and audits throughout the design of the PACE pilot is a strength of the innovative assessment and accountability system. We provide mechanisms and analyses to support the validity and comparability of the annual determinations and ensure system coherence by checking the assumptions that underlie those determinations. Comparability at the level of the annual determinations rests on the comparability of local scoring **within districts**, the comparability in the evaluation of student work **across districts**, and finally, the comparability of annual determinations **across PACE districts and across PACE and non-PACE districts**. Gathering evidence of comparability at every level has been essential for supporting the claims of comparability, and ultimately the validity of the system as a whole. Each of the activities and audits that occur at the three levels is described in detail below.

1. **Within-District Comparability in Expectations for Student Performance**

Cross-district comparability is a key goal of PACE, but cross-district comparability is not meaningful if within-district comparability is not achieved. Therefore, we first discuss efforts to improve and monitor within district comparability. First and foremost, the innovation occurring in the PACE pilot happens at the local school district level. Within each district, rich discussions among educators about what competency looks like for every grade level and content area occurs in the adoption of a competency-based instructional model and the development of aligned performance assessments. Defining the expectations for student performance in a competency-based education model requires that the educators have shared definitions about both the content standards and the required evidence for evaluating student competence relative to the content standards. Therefore, baked within this model is within-district comparability in expectations for student performance.

The PACE pilot encourages and audits the degree of consistency in educator scores of student work through the use of common performance tasks in two main ways: 1) within district calibration sessions resulting in annotated anchor papers, and 2) within district estimates of inter-rater reliability. First, all PACE districts hold grade-level calibration sessions for the scoring of the common task. Teachers bring samples of their student work from the common performance task representing the range of achievement in their classrooms. Teachers work together to 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.[[1]](#footnote-1) Second, we externally audit the consistency in scoring by asking each district to submit a sample of papers from each common performance task that have been double-blind scored by teachers. The collection of double scores is then analyzed using a variety of inter-rater reliability methods to estimate within-district scoring consistency (e.g., percent exact and adjacent agreement, Cohen’s Kappa, intraclass correlations).

1. **Cross-District Comparability in Evaluating Student Work**

As discussed in the March 2016 progress report, there are three main purposes for the common tasks across districts: 1) to serve as models of high quality tasks to support local capacity building, 2) to contribute to the long-term goal of building a large task bank from which districts can draw for local assessment purposes, and finally 3) to provide evidence of cross-district comparability in the evaluation of student work relative to the content standards. While the March report explicates the first two, the purpose of this brief is to discuss the use of common performance tasks to estimate cross-district comparability in scoring. Because all students in the participating PACE districts are taking at least one common performance task, 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. The method of estimating comparability using a moderation audit with the common performance tasks was heavily-based on international examples drawn from Queensland, Australia and the United Kingdom. The primary goal of NH PACE’s moderation 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. Beyond providing quality assurance about the comparability of annual determinations among PACE districts, the moderation audit also provides data to inform conversations about leniency or stringency of scoring at the school and district level. These conversations are particularly useful because they should lead to improvements in teacher assessment literacy as well as improving how teachers judge the quality of student work relative to the state competencies thereby improving cross-district comparability over time.

The moderation audit in 2016 was closely modeled on the same process conducted in the summer of 2015[[2]](#footnote-2) with incremental improvements based on lessons learned (e.g., the evaluation of student work and scoring will all occur online rather than paper-based). This audit is heavily based on methods that have been successful in Queensland, Australia for decades (see Queenland Studies Authority, 2014). The moderation audit involves pairing teachers, each representing different districts, to score student work samples. After training and practice, both judges within the pairs are asked to individually score their assigned samples of student work and record their 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 analogous to a “true score.” 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.

In addition to the moderation audit with the common performance tasks, in 2015 we piloted a rank-ordering method for evaluating the comparability of scores for local performance tasks that were distinct, but assessed the same content area. This methodology had been previously used the United Kingdom to ensure comparability in the written examinations used for admissions to universities that are developed, administered, and scored by a number of different boards across England (see Bramley, 2005). The rank ordering method—tested last year using the high school life science content area—showed success in detecting differences in both scoring and task quality.[[3]](#footnote-3) We plan to continue to test this method in additional content areas for potential use in future years to monitor the comparability in scoring of local (uncommon) performance assessments.

1. **Comparability of Annual Determinations**

The accountability uses for the assessment system results lie at the level of the annual determinations, the PACE pilot comparability claims 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 PACE, cross-district level are all in an effort to support the claim of comparability in the annual determinations. PACE procedures to formally promote and evaluate the comparability of PACE annual determinations across both PACE and non-PACE districts occurs in three primary ways: 1) setting performance standards using common achievement level descriptors, 2) validating the performance standards with a body-of-work method, and 3) administering an external statewide standardized assessment once per grade span (i.e., Smarter Balanced and SAT).

First, the process used to set performance standards on the competency score scales supports the cross-district comparability in that the method is based on common achievement level descriptors (ALDs). The PACE ALDs support shared interpretations of performance relative to the content standards and ultimately, through the standard setting procedure, a common meaning for the performance standards (i.e. cut scores) for each of the PACE districts. The ALDs were developed using the Smarter Balanced (SBAC) ALDs as a starting point with the NH College and Career-Ready Standards as the primary reference. Per recommendation from the PACE Technical Advisory Committee, teachers from the participating districts modified the Smarter Balanced ALDs for two purposes; 1) to make them more closely reflect the intended prioritization of New Hampshire educators, and 2) to increase their interpretability by synthesizing the SBAC range ALDs into more holistic descriptors of student achievement. Therefore, the PACE ALDs support a shared definition of student achievement relative to the content standards across all participating districts. The Contrasting Groups standard setting method is designed to identify cut scores that carry the same meaning across districts based on the ALDs. This method therefore is part of the PACE processes to support comparability in the annual determinations (i.e., comparability by design). The degree to which this method truly results in comparable annual determinations is evaluated by comparing the distributions of student placement into achievement levels within and across districts (e.g., looking at predicted achievement based on external assessments).

In addition to comparing the distributions of student placement into achievement levels within and across districts, for 2016 the comparability of the performance standards will be additionally validated using a body-of-work method. Throughout the academic year we have asked that each district choose a sample of nine students, representing the range of performance in that district, for one content area per grade level. Teachers are asked to collect samples of student work from those nine students for each of the competencies. This summer, we have brought teachers together to review the portfolios of student work from districts that are not their own to and make judgments about student achievement relative to the ALDs. These teacher judgments will be reconciled with reported student performance as an additional source of validity evidence to support the comparability of the annual determinations across PACE districts.

Lastly, we will evaluate the comparability of the annual determinations between PACE and non-PACE districts by comparing performance across years and calculating PACE annual determinations for the grades that are currently taking SBAC and the SAT. First, since students participate in SBAC once per grade span, we will compare last years’ performance on SBAC with this years’ performance on PACE for students in grade 4 ELA, grade 5 math, and grade 9 math and ELA. Even more powerfully, by calculating PACE annual determinations for the students taking SBAC this year, the state will have both SBAC and PACE 2015-2106 data points for students in grade 3 ELA, grade 4 math, grade 8 ELA and math, and grade 11 ELA and math. The degree of agreement between the two sets of annual determinations will be analyzed to further support the comparability of the interpretations of the reported achievement levels. The results of all of the analyses described in this paragraph will be reported for all students and for each subgroup (when the minimum n size in met) for each of the four performance levels. In addition to t

## Summary and Discussion

This request provided information regarding participation of schools and districts for the coming school year. It also outlined the ways in which the NH DOE will ensure that locally-developed tasks and local assessment systems meet standards for high quality. We outlined six steps for documenting and supporting the quality of local tasks and systems. We are confident that this multi-faceted approach will provide the evidence necessary to document and improve the quality of the local assessment system.

The third part of this request outlined how the processes for promoting and evaluating the comparability of the inferences drawn from the PACE annual determinations are integrated at every level of the PACE innovative assessment and accountability design. Comparability is not determined by a single analysis, but is rather an evidence-based argument that is logically based in thoughtful system design and evaluation. Programmatic processes and audits are present throughout the design of the PACE pilot in order support the validity and cross-district comparability of the annual determinations. As discussed in this brief, those processes and data collection efforts begin within each district, are carried across the PACE network, and ultimately span both PACE and non-PACE districts to product comparable annual determinations that provide valid interpretations about student achievement and can be used the NH school accountability system.

Collectively, NH DOE believes that this document fulfills the request for information necessary for Secretary King to grant an extension of New Hampshire’s waiver under NCLB and now ESSA.

## References

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

Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice*, (Winter), 27–34.

Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, *57*(3), 373–399.

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.

Queenland Studies Authority. (2014). School-based assessment: The Queensland assessment.

Queensland, Australia: The State of Queensland (Queensland Studies Authority).

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, *30*, 215–232.

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.

## Appendix A: Sample Course Assessment Plans for District Competencies

***Rochester Competency Assessment Matrix***

***Grade 3 Math***

|  |
| --- |
| **C1~Operations & Algebraic Thinking:** Students will demonstrate the ability to compute accurately, make reasonable estimates, understand meanings of operations and use algebraic notation to represent and analyze patterns and relationships. |
| **Performance Indicator(s)** | **RCC** | **CCSS** | **U1** | **U2** | **U3** | **U4** | **U5** | **U6** | **U7** | **U8** | **U9** | **Beginning** | **Mid** | **End** |
| Students will be able to represent and solve problems involving multiplication and division. | C1PI1a | OA. 1, OA.2 | Ch | W Ch | W Ch | C | W |   | W |   |   | W | W |   |
| C1PI1b | OA. 3 | S W Ch | S W Ch |   |   |   | S W Ch |   | S W Ch C | W | W | W | W |
| C1PI1c | OA.4, OA.6, OA.7 | W Ch | S W Ch C | S W Ch | C | S W | W C |   | W Ch C | S W |   | W | W |
| Students will understand properties of multiplication and the relationship between multiplication and division. | C1PI2 | OA.5 |   |   | S W Ch |   | W Ch O | S W Ch C |   | C | W Ch |   |   | W |
| Students will be able to solve problems involving the four operations, and identify and explain patterns in arithmetic. | C1PI3a | OA.8 |   | S W Ch | W |   |   | S W Ch C |   | C | W | W | W | W |
| C1PI3b | OA.9 |   |   |   |   | S W |   |   |   | O | W | W |   |
| C1PI3c |   |   |   |   |   |   | S W |   | C |   |   |   | W |
| **C2~:** **Number and Numeration in Base Ten/Fractions:** Students will demonstrate the ability to understand the meanings, uses, and representations of numbers, as well as equivalent names for numbers. |
| **Performance Indicator(s)** | **RCC** | **CCSS** | **U1** | **U2** | **U3** | **U4** | **U5** | **U6** | **U7** | **U8** | **U9** | **Beginning** | **Mid** | **End** |
| Students will be able to use place value understanding and properties of operations to perform multi-digit arithmetic. | C2PI1a | NBT.1 | S W | C | W | C |   | C |   | C |   |   | W | W |
| C2PI1b | NBT.2 | S W Ch | S W Ch | S W O | C |   | S W Ch |   | C |   | W | W | W |
| C2PI1c | NBT.3 |   |   |   |   |   |   |   | S W | S |   |   | W |
| Students will develop understanding of fractions as numbers. | C2PI2a | NF.1 |   |   |   |   | S W |   | W | C |   |   |   | W |
| C2PI2b | NF.2a, NF.2b |   |   |   |   |   |   | S W Ch |   |   |   |   | W |
| Students will be able to explain equivalence of fractions in special cases, and compare fractions by reasoning about their size. | C2PI3a | NF.3a |   |   |   |   | W | C | W |   |   |   |   |   |
| C2PI3b | NF.3b |   |   |   |   | S W | C | S Ch | C |   |   |   | W |
| C2PI3c | NF.3c |   |   |   |   |   |   |   |   |   |   |   | W |
| C2PI3d | NF.3d |   |   |   |   |   |   | S W Ch O | C |   |   |   | W |
| Students will be able to read and write whole numbers up to 1,000,000 and decimals through the 100ths place. | C2PI1a | NBT.1 | S W | C | W | C |   | C |   | C |   |   | W | W |
| **C3~Measurement:**  Students will demonstrate the ability to understand the systems and processes of measurement; use appropriate techniques, tools, units, and formulas in making measurements. |
| **Performance Indicator(s)** | **RCC** | **CCSS** | **U1** | **U2** | **U3** | **U4** | **U5** | **U6** | **U7** | **U8** | **U9** | **Beginning** | **Mid** | **End** |
| Students will be able to solve problems involving measurement and estimation of intervals of time, liquid volumes, and masses of objects. | C3PI1a | MD.1 | S W Ch O | C |   | C |   |   |   | C | S W Ch | W | W | W |
| C3PI1b | MD.2 | W |   |   |   |   |   | S W | C | W Ch |   | W | W |
| Students will understand concepts of area and relate area to multiplication and division, (geometric measurement). | C3PI2a | MD.5a MD.5b MD.6 |   |   |   | S W |   | C |   |   |   |   |   |   |
| C3PI2b | MD.7a |   |   |   | W |   | C |   | C |   |   | W |   |
| C3PI2c | MD.7b |   |   |   | Ch |   |   |   | C |   |   |   | W |
| C3PI2d | MD.7c |   |   |   |   | S W |   |   |   |   |   |   |   |
| C3PI2e | MD.7d |   |   |   | S W | Ch |   |   | C |   |   |   | W |
| Students will be able to recognize perimeter as an attribute of plane figures and distinguish between linear and area measures (geometric measurement). | C3PI3 | MD.8 |   |   |   | S W Ch |   | C |   | C |   |   |   | W |
| **C4~Data:** Students will demonstrate the ability to represent and analyze data.  |
| **Performance Indicator(s)** | **RCC** | **CCSS** | **U1** | **U2** | **U3** | **U4** | **U5** | **U6** | **U7** | **U8** | **U9** | **Beginning** | **Mid** | **End** |
| Students will represent and interpret data. | C4PI1a | MD.3 | S W |   | W |   |   | C |   |   |   | W | W |   |
| C4PI1b | MD.3 | W O | C |   | C |   | C |   |   |   | W | W |   |
| C4PI1c | MD.4 |   |   |   | S W Ch |   | C |   | S W |   |   |   | W |
| **C5~Geometry:** Students will demonstrate the ability to investigate characteristics and properties of two and three dimensional and geometrical shapes and apply transformations and symmetry in geometric situations.  |
| **Performance Indicator(s)** | **RCC** | **CCSS** | **U1** | **U2** | **U3** | **U4** | **U5** | **U6** | **U7** | **U8** | **U9** | **Beginning** | **Mid** | **End** |
| Students will be able to reason with shapes and their attributes. | C5PI1a | G.1 |   |   |   | S W Ch |   | C |   | W Ch |   |   | W | W |
| C5PI1b | G.2 |   |   |   |   | W | C | W | S C |   | W |   | W |

1. Beginning AY 2016-2017 the plan is to have a centralized state-level review process for the annotated anchor papers so they can be distributed across all PACE districts for each of the common performance tasks. Not only will this increase cross-district comparability in scoring for the common tasks in 2016-2017, but the anchor papers will be included in the performance task bank along with the retired common task for future years. This will help increase comparability in the scoring of local tasks as the task bank is increasingly relied on to support local assessment needs. [↑](#footnote-ref-1)
2. Detailed descriptions of the methods and analyses can be found in the October 2015 and December 2015 PACE reports. [↑](#footnote-ref-2)
3. See the October 2015 PACE report for a detailed explanation of the analysis and results. [↑](#footnote-ref-3)