Hawai'i Comprehensive Assessment Program

End-of-Unit Performance Task Administration Guide

Hawaiʻi Department of Education
Assessment Section
Office of Strategy, Innovation, and Performance
hsa@k12.hi.us

Last Updated March 2022

Version 1.0

TABLE OF CONTENTS

SECTION 1.0: INTRODUCTION TO THE GUIDE

1.1. Intended Audience

This Performance Assessment Administration Guide is intended for Hawai'i teachers and school/district leaders who play a role in the selection, interpretation, and use of high-quality, curriculum-embedded classroom performance-based instructional activities and assessments using Hawai'i's Performance Assessment Task Bank.

1.2. Purpose of the Guide

This guide provides administration instructions to support the appropriate use of Hawai'i's Performance Assessment Task Bank. The performance assessments included in the task bank were created by Hawai'i teachers for Hawai'i students during the Performance Assessment Development Initiative (PADI)--a federally-funded, state-sponsored initiative that started during the 2021-22 school year. The initial English language arts and mathematics performance assessments in the task bank were created during year-long development and review cycles supported by the Hawai'i Department of Education and the Center for Assessment. Each performance assessment in the task bank is aligned to a subset of Hawai'i's content standards and is intended to support standards-based classroom teaching and learning. Hawai'i's Performance Assessment Task Bank continues to grow each year, with more and more performance-based tasks and content areas added for use in classrooms across Hawai'i. Videos that accompany this guide can be accessed from the following video library.

1.3. Organization of the Guide

This guide is organized into the following sections:

- **Overview of Performance Assessments:** an overview of performance assessments–what they are, main reasons for using performance assessments, and how they relate to standards and performance outcomes.
- **Intended Uses of the Hawai'i Performance Assessments in the Task Bank**: explains the intended uses of the performance assessments in the task bank (classroom instructional purposes or classroom summative assessment purposes). This section also explains how the task bank is organized to support embedding the performance tasks within the curriculum and aligning the selected tasks to ongoing standards-based instruction. Instructions on how to submit additional performance assessments to the task bank are provided.
- **Guidelines for Administering the Performance-Based Tasks for Classroom Instructional Purposes:** provides guidelines for using the performance tasks included in the task bank as instructional tools rather than assessment tools.
- **Guidelines for Administering the Performance-Based Tasks for Classroom Summative Assessment Purposes:** provides guidelines for using the performance tasks included in the task bank for classroom summative assessment purposes. Guidelines are provided to

support successful administration, including specific instructions for what to consider before, during and after administration.

- **Glossary of Terms:** provides definitions of key terms related to performance assessment design, implementation, and scoring **bolded in blue font** throughout the document.
- **References:** provides a list of references based on in-text citations.

SECTION 2.0: OVERVIEW OF PERFORMANCE TASKS

2.1. Definition of Performance Task

Numerous definitions and understandings of the term **performance assessment** exist (Lane & Stone, 2006). The definition of performance assessment used in this guide is derived from assessment experts and scholars (Darling-Hammond & Adamson, 2010; Education Week, 2019; Lane, 2010; McTighe & Ferrara, 2021):

> Performance assessments measure how well students apply their knowledge, skills, and abilities to authentic problems. Performance assessments require the student to produce something (e.g., a report, product, experiment, or performance), which is scored against specific criteria. A performance assessment may be designed to occur over a period of hours, days or weeks depending on the range and complexity of skills to be assessed.

Although the terms performance assessment and **performance task** are often used interchangeably, in this guide we make a distinction (Education Week, 2019). A performance assessment refers to the entire assessment, which might include several performance tasks. When we use the term performance assessment, we are referring to the entire assessment, including the teacher instructions, student instructions/prompt, rubric, and other materials. A performance task is the activity that students will complete to demonstrate their knowledge and skills.

Additionally, another term often used alongside performance assessment is **authentic assessment** (Marion & Buckley, 2016; Wiggins, 1989). Assessments are authentic when they are designed for students to demonstrate knowledge and skills that are used in a specific field or discipline. For example, after a unit on chemical interactions, students might be asked to create their own chemical interaction for a specific application. This task is authentic because that is the kind of work that chemists do. Authentic assessment and performance assessment are related concepts. Performance assessments may be authentic, but they do not have to be. It is often harder to design performance assessments to be authentic with younger students as they may not have all the necessary prerequisite knowledge and skills to produce something like someone in a specific field or discipline.

2.2. Why Use Performance Tasks

Performance assessments have been used for decades in classrooms to gather evidence about students' ability to apply complex knowledge and skills to new or novel contexts. Performance assessments offer many benefits for students and teachers, if designed well. Specifically, performance assessments allow an educator to:

1. **Assess complex standards and skills:** supports the assessment of 21st century skills such as critical thinking, collaboration, communication, and problem solving (Center for Assessment, 2020; National Research Council, 2011).
2. **Engage students:** engages students as active participants in the assessment process (Tayler et al., 2016).
3. **Support personalization:** provides students with choices about how to show what they know and can do, and by providing individualized feedback (Stiggins, 1997).
4. **Honors students' culture, agency, and lived experiences**: supports **culturally relevant** and **culturally responsive** educational mindsets by drawing on students' cultural identities, assets, and engagement (Aronson & Laughter, 2016; Evans, 2021; Gay, 2018; Ladson-Billings, 1995; Stembridge, 2020).
5. **Build authentic experiences:** provides students with **authentic assessment** opportunities (Wiggins, 1989).
6. **Provide instructional information in addition to summative information:** provides a window into students' thinking in ways that selected-response items cannot (Marion & Buckley, 2016).

2.3. When to Use Performance Tasks

Performance assessment includes tasks or activities that result in students producing tangible products or performances to serve as evidence of their learning. Well-designed performance assessments require students to apply complex knowledge and skills–rather than simply to recall and recognize. Since performance assessments call for application, they are well suited to assess students' application of conceptual understandings, disciplinary practices, integration of knowledge across content areas, and 21st century skills (Darling-Hammond & Adamson, 2010; McTighe & Ferrara, 2021).

Performance assessment may not be necessary or the most appropriate method when the skills being assessed involve memorization of facts, routine procedures, or discrete skills (McTighe & Ferrara, 2021). For example, a performance assessment is not necessary to gather evidence about whether a student can correctly read a thermometer–a quick selected-response item could be used. However, a performance assessment could be used to gather evidence about students' ability to predict the weather using complex knowledge and skills involving application of mathematical equations using observational data where reading a thermometer is one component of the task. Additionally, performance assessments are typically open-ended and best suited to situations where the task or question has multiple possible solution paths and

answers–rather than one correct answer. The quality of student thinking can then be judged qualitatively using a rubric instead of correct or incorrect.

SECTION 3.0: INTENDED USES OF THE HAWAIʻI PERFORMANCE TASKS IN THE BANK

3.1. Intended Uses of the Hawaiʻi Performance Tasks in the Bank

The performance assessments included in the **task bank** are intended to be used for classroom assessment purposes only–not for any high-stakes use such as within the school accountability system or determining student achievement. Since curriculum, instruction, and assessment should be designed as a coordinated and coherent whole, the performance assessments are intended to be embedded within high-quality curriculum. Each performance assessment was designed to assess specific Hawaiʻi content standards to support standards-based teaching and learning.

There are two primary uses of the performance tasks in the task bank: classroom instructional purposes and classroom **summative assessment** purposes.

3.2. Classroom Instructional Purposes

Assessment should be coherent with curriculum and instruction. In order for that to be true, high-quality performance assessments used for classroom summative assessment and grading purposes (see next section) should be indistinguishable from the rich performance-based tasks used for classroom instructional purposes. In other words, *the performance tasks in the task bank can be used as an instructional teaching and modeling tool rather than an assessment tool* (see Section 4.2 for more specificity on instructional use). Doing so allows students to practice applying complex knowledge and skills in new or novel situations, which supports the ultimate goal of teaching and learning: transfer (Baker & Gordon, 2014; National Research Council, 2012). See Section 4.0 for guidelines on using the performance-based tasks for instructional purposes.

3.3. Classroom Summative Assessment Purposes

The most common use for performance assessment is to assess student application of knowledge and skills towards the end of a unit or course of study. This summative classroom assessment purpose documents student achievement of Hawaiʻi's content standards at a point in time–sometimes referred to as assessment of learning. These assessments may be graded and reported to parents and students. Because there are rubrics associated with performance assessments, more specific feedback on performance can be provided along with additional teacher notes in addition to a grade or performance level. See Section 5.0 for guidelines on using performance assessments for classroom summative assessment and grading purposes.

3.4. Task Bank Organization

The **task bank** is organized to support standards-based instruction and assessment, as well as embedding the performance-based tasks within high-quality curriculum. As such, the task bank lists the performance assessment name along with the aligned grade/content area standards and big idea. Curriculum connections can be found in the performance assessment booklet. The search feature in the upper right hand corner of the website can be used to search for key topics (e.g., fractions, informational writing, algebra) or standards (e.g., 4.OA.1). There is an explanatory video in the video library (video #3) that illustrates in detail how to navigate and search the task bank to select performance assessments that align to specific standards and units of study.

SECTION 4.0: GUIDELINES FOR ADMINISTERING THE PERFORMANCE TASKS FOR CLASSROOM INSTRUCTIONAL PURPOSES

4.1. Purpose of Performance-Based Instructional Tasks

As stated previously, the performance tasks included in the task bank can be used as an instructional tool rather than an assessment tool. **Performance-based instructional tasks** are an entirely necessary and appropriate tool for teachers to use in their regular instruction with students. A performance-based instructional task is simply a performance task used for teaching rather than evaluative purposes. The purpose for using the performance tasks during instruction is to support high-quality teaching and learning. Performance tasks require students to apply complex knowledge and skills, which is a key component of addressing the depth of the state's content standards. For example, some state content standards require students to move beyond discrete skills to application, analysis, evaluation, and synthesis.

4.2. Various Ways to Use the Performance-Based Instructional Tasks

There are various ways a teacher can use the performance tasks in the task bank for instructional purposes. There is no one right time to use a performance-based instruction task. It all depends upon the unit of instruction, standards, and learning targets. Some teachers may choose to use a performance-based instructional task at the beginning, middle, or end of a unit or course of study. Specifying the intended purpose and use for the performance-based instructional task will clarify timing.

Below are some selected examples that show how teachers may use performance-based instructional tasks. For example, teachers may use performance tasks during instruction to:

- Model and scaffold for students how to approach a complex problem by thinking aloud with their students. For example, a teacher could project the performance task and verbally share with students how they read the prompt to understand the task, how they break the complex task into multiple parts, and how they pull on their knowledge

and skills to produce a product or perform what is expected. Over different instructional tasks the teacher could gradually pull back and have students take a more active role in producing the product or performance (i.e., I do, we do, you do method).

- Elicit evidence of student learning in order to adjust teaching and learning to better meet students' needs. For example, while using a performance-based task during instruction, a teacher may realize that students have many misconceptions about a topic that need to be addressed or retaught. Or a teacher may realize that a specific group of students need additional instruction and feedback in order to move their learning forward.
- Engage students in a cognitively rigorous performance task/activity that attempts to simulate real-world experiences and authentic assessments relevant to the student and the discipline. Student engagement in learning is crucial. Without engagement, students are not attentive or involved as active agents in their own learning. Performance-based instructional tasks provide teachers with a way to engage students because they can promote instruction that is more relevant to students' lived experiences, interests, or prior knowledge; fosters student autonomy; allows for collaboration; and promotes authenticity (Tayler et al., 2016).
- Prepare students to successfully engage and successfully complete a classroom summative performance assessment. Students need to have had an opportunity to learn, practice, and receive formative feedback on how to deconstruct the expectations of a prompt, know how to apply their knowledge and skills in a new or novel situation rather than just selecting a response or solving an algorithm. They need to have had practice explaining their thinking or justifying a response. Oftentimes when teachers do not include rich performance-based tasks and activities within instruction, students may be able to perform simple procedures and discrete skills, but they are not prepared to apply complex knowledge and skills independently.
- Provide opportunities for student self-assessment and peer feedback. Teachers can model for students how to use the rubric to self-assessment progress towards learning goals. The teacher can also group students to provide formative feedback to one another on the quality of student work products or performances in ways that constructively further learning and growth.

SECTION 5.0: GUIDELINES FOR ADMINISTERING THE PERFORMANCE TASKS FOR CLASSROOM SUMMATIVE ASSESSMENT & GRADING PURPOSES

5.1. Preparing to Administer

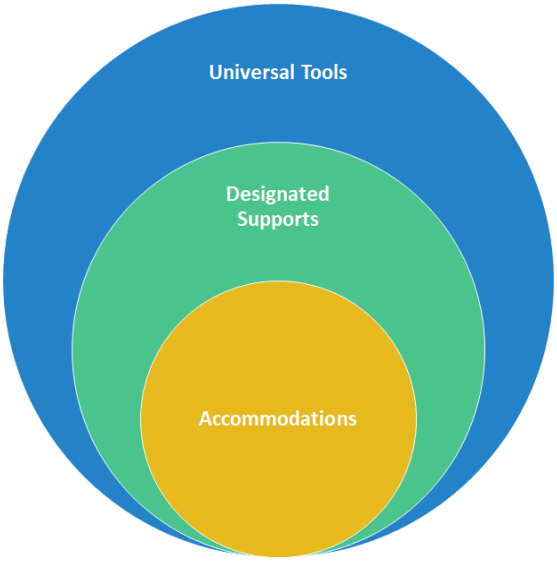
A teacher should complete the following steps prior to administering a performance assessment from the task bank for classroom summative assessment and grading purposes:

1. Thoroughly read the performance assessment materials. The materials contain important information about the standards and big idea(s) assessed; general description of the performance task; teacher directions including estimated amount of time to administer, resources/materials needed, accessibility strategies (see Section 5.2 on Accessibility Strategies below); student directions, which includes the student booklet and rubric; and scoring guidance.
2. Decide where the performance assessment fits best within your curricular scope and sequence. The performance assessments in the task bank are intended to be implemented in ways that support high-quality curriculum implementation. Use the standards specified in the performance assessment materials to identify curriculum connections such as possible units of instruction where the performance tasks could be embedded.
3. Complete the performance task yourself. This is an oftentimes overlooked and yet critical aspect of making sure you have prepared students to perform well on the performance task and that you identify likely questions or barriers students might face.
4. Decide if the performance assessment context, text materials, or scenario should be altered to make the assessment experience more engaging and culturally relevant/responsive for students in your classroom. The student directions portion of the PADI performance assessment booklet clarifies which aspects of the student booklet can be changed without changing what is being assessed. See the video library for a short video with examples.
5. Make copies of the student booklet/directions/prompt and any other necessary materials, or make sure the student booklet/directions/prompt is accessible to students to complete online.
6. Practice a performance-based instructional task with a different context, text materials, or scenario with students. Model and use guided practice to familiarize students with the type of thinking and responses required on a performance assessment.

5.2. Preparing Accessibility Strategies

Fairness is a key consideration in educational assessment design and validation. Fairness means that an assessment does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct (or what is being assessed)(AERA, APA, & NCME, 2014). **Accessibility** is a necessary condition for fairness and means that all students have an unobstructed opportunity to demonstrate their understanding of the standards that are being assessed without being unduly advantaged or disadvantaged by individual characteristics (AERA, APA, & NCME, 2014).

Fairness and accessibility not only relate to how performance assessments are designed and validated, but also to how performance assessments are administered. **Accessibility strategies** is an umbrella term used to describe multi-tiered strategies to support all students individually based on their unique needs and preferences. These multi-tiered strategies include **universal tools, designated supports**, and **accommodations**.



As shown in the table below, universal tools are a set of tools available to all students, designated supports are provided based upon teacher discretion, and accommodations are legally documented and mandated. The PADI performance assessments in the task bank contain accessibility strategies information in the teacher directions section of the task template. Smarter Balanced provides detailed read aloud guidelines and scribing protocols (see Appendices D & E in the hyperlinked document) which should be used to inform how those accommodations are implemented. Only those guidelines relevant to a classroom performance assessment should be followed.

Universal Tools	Designated Supports	Accommodations
A set of tools available to all students.	*A set of tools available to students for whom a need has been identified by school personnel familiar with each student's needs.*	*A set of resources available to students with a documented need noted in an Individualized Education Program (IEP) or 504 plan.*
• English Dictionary* • Highlighter • Math Tools (calculator, ruler, protractor, multiplication table, number line, 100s chart, etc.)* • Scratch Paper • Spell Check* • Thesaurus*	• Bilingual Dictionary* • Color Overlays • Illustration Glossaries* • Magnification • Noise Buffers • Read Aloud* • Read Aloud in Spanish* • Scribe* • Simplified Student Directions • Text-to-Speech* • Translated Student Directions* • Translations (Glossary)*	• American Sign Language • Braille • Braille Transcript • Closed Captioning* • Math Tools (calculator, ruler, protractor, multiplication table, number line, 100s chart, etc.)* • Read Aloud* • Scribe* • Speech-to-Text* • Text-to-Speech*

*Depends on measurement construct. For example, if the performance assessment is intended to measure reading comprehension, then reading aloud would not be an appropriate designated support.

NOTE: This is not a complete list of all possible universal tools, designated supports, and/or accommodations that could be used to reduce barriers and promote accessibility for classroom performance assessments. This list was taken from Smarter Balanced Usability, Accessibility, and Accommodations Guidelines.

5.3. Starting to Administer

Teachers should ensure that students understand what the performance task is asking them to produce or perform. Teachers can read the student directions/prompt with students and clarify any questions prior to students completing the work independently or collaboratively. Any universal tools available to all students should be discussed, as well as designated supports and accommodations clarified with students at the outset. See Section 5.2 on Accessibility Strategies for more details.

Teachers will also want to preview the scoring rubric with students so that students understand the criteria for success prior to starting.

5.4. Responding to Student Questions During Administration

Students often have questions while taking performance assessments. Some questions are easily answered because it relates to clarifying the student directions/prompt. Other questions though may suggest that students are struggling. It may be the case that students are struggling to figure out how to approach the problem, how to break the problem into steps, or how to apply what they know and can do to the new or novel situation.

Our advice for responding to student questions during administration is simple to state, but much more difficult to ascertain. There is a difference between productive struggle and unproductive struggle. No classroom assessment should place undue burden, anxiety, or stress on students. This is not a **high-stakes test**! Teachers will want to use their best judgment with respect to when students need a simple clarification about the directions or vocabulary, encouragement to continue to do their best and productively struggle, and when a student may need specific prompting or other scaffolding support. Teachers can make a note or write on the students' paper where they provided help/assistance, if it warrants noting.

Additionally, if a teacher notices that many students are struggling, it may make sense to pause, providing instruction on concepts that students are struggling with and/or modeling on application, and then consider resuming the administration of the performance assessment.

5.5. Scoring, Grading, & Reporting After Administration

The performance assessments in the Hawai'i Performance Assessment Task Bank are intended to be used for classroom instruction or assessment purposes only. Teachers may decide to use the associated **rubric** to score student work and provide a grade that reports on student progress towards proficiency to parents and students. Below are recommendations related to scoring, grading, and reporting **student work** resulting from performance tasks.

Scoring: There is an explanatory video in the video library that explains how to score and calibrate scores for performance tasks. It illustrates the bulleted steps below.

- Read the *Nine Principles of Scoring Student Work* in the Appendix.
- Review scoring rubrics, **anchor papers**, and **scoring annotations** in the performance assessment materials.
- Consider scoring an initial set of papers with colleagues to **calibrate scoring**. See *Scoring Calibration Protocols* in the Appendix for two methods: **double-blind scoring** and **anchor paper scoring.**

Grading:

There is an explanatory video in the video library that explains how to grade performance tasks.

We start first with what not to do: Add all the points in the rubric together and divide by the total number of points possible. For example, if using an **analytic rubric** with five criteria each worth a maximum of 4-points, there would be a total of 20-points possible on the performance task. If level 3 on the rubric represents proficiency, then a student who meets the state standards would earn 15 out of the total 20 points possible. However, simply dividing 15 by 20 results in a 75% or 'C' grade in most traditional grading systems–yet this student demonstrated proficiency.

If you are interested in a method of converting the rubric score to a grade, we have provided a suggested approach below. We also recommend you consult with your school leadership and other teachers. If rubric scores do not have to be converted to traditional letter grades or percentages, then we recommend avoiding the practice and skipping to the next paragraph.

- Convert rubric scores to numerical values out of 100 (e.g., Level 4/Above grade level expectations =100; Level 3/meets grade level expectations = 90; Level 2/making progress towards grade level expectations = 80; Level 1/well below grade level expectations = 70), add them together, then divide by the number of points possible and multiply by 100. For example, for a student who scores level 3 on five rubric criteria, they would score 450/500 which is 90% or an 'A-' in most traditional grading systems.

For those teachers who do not need to convert rubric scores to grades or percentages, but can report to parents and students using the scale of the rubric (e.g., 4 levels or 4 score points) then there are at least a couple ways to approach how to combine score information across criteria/dimensions on an analytic rubric. For example, on a rubric with 4 levels and 5 criteria (or dimensions), either the mode or the mean can be used to summarize the quality of the student's response, product, or performance.

- The **mode** is most similar to a preponderance of evidence approach where the pattern of student scores might be 3, 2, 3, 3, 2 and so the student earns a score of 3 because that is the mode.

- The **mean** (or average) could be used such that the student with the same score profile (3, 2, 3, 3, 2) would earn a score of 2.6.

In order to decide which method to use (mode vs. mean), create scores for a sample of student work from a class or entire school and then evaluate which score (the mode or the mean) best represents the quality of students' responses, products, or performances overall. Then stick with this method for grading all performance tasks.

In all cases, strive for consistent grading practices across a complex area, schools, and classrooms.

Reporting:
Simply reporting a score, percentage, or grade to a student or parent may hold some value in terms of communicating student achievement. However, the primary purpose of grades is to communicate (or report) student achievement based on clearly articulated learning goals and established criteria. Therefore, one of the benefits of returning the scored rubric to students and parents is that the scores are contextualized next to the criteria.

Additional verbal or written feedback added to the rubric scores can also promote students' ability to learn and grow from the assessment experience. Specific and meaningful feedback can help students understand what they did well and what they could do differently next time to improve the quality of their response, product, or performance.

5.6. Student Work Analysis After Administration

Student work provides a window into how students construct meaning of key concepts and skills. By analyzing and interpreting student work through a clear and systematic process, teachers can improve instructional decisions for individuals and groups of students, and ultimately impact student achievement (see *Student Work Analysis Protocol for Instructional Purposes* in the Appendix). Although teachers certainly review student work in order to provide a grade and perhaps to determine students' understanding of specific content standards, a systematic diagnostic analysis that allows for determining instructional next steps is often missing. We want teachers to make a shift from only *scoring* student work to also *diagnosing* student performance for instructional purposes.

In addition, analyzing student work can provide a lens in which to determine the quality of assignments and the unit as a whole. We know that a clear and cohesive unit is created when there is a progression of learning that advances knowledge and skills over time and assignments allow students to demonstrate this understanding with appropriate scaffolds and independence. High quality assignments and units allow authentic learning to be demonstrated in ways that provide for student choice and interest. It is only through this diagnostic work that

teachers can make thoughtful instructional revisions and make appropriate instructional decisions that will improve student learning.

Based upon findings from the student work analysis, teachers can use the process to inform next instructional steps in at least two ways: looking backwards and looking forwards (see the table below). **Looking backwards** involves the teacher (or team of teachers) reflecting back on the previous unit and examining what they could do differently in their curriculum, instruction, or assessment practices with the next cohort of students to better prepare students to meet the expectations in the performance task. **Looking forward** involves the teacher (or team of teachers) projecting forward to upcoming units of instruction and examining how they can weave in disciplinary practices or concepts that may need additional instruction because these practices or concepts repeat over the course of the year. In both of these situations–looking backwards and looking forwards–the student work analysis information is used to inform subsequent instruction with a different cohort of students or with the same students but in future units.

Students can also be involved in this instructional process by **looking back** and reflecting on the success of their learning strategies and efforts in relation to the learning goals. **Looking forward** students can use teacher feedback on their performance to adjust future learning strategies and goal setting for the next unit or future units of instruction.

	Teacher	Student
Looking Back	**Instructional effectiveness *evaluation* use:** teacher examines the effectiveness of instruction in helping all students demonstrate learning goals. The teacher could use this feedback to adjust curriculum, instruction, or assessment strategies next year with a different group of students.	**Learning strategies *evaluation* use:** student reflects on the effectiveness of their learning strategies and efforts applied in relation to successful demonstration of learning goals.
Looking Forward	**Instructional *planning* use:** teacher uses assessment results related to what students need to re-learn or disciplinary practices and concepts that repeat over the course of the year to inform future instructional planning. The teacher could use this feedback to adjust curriculum, instruction, or assessment strategies for the next unit or future units of instruction.	**Learning strategies *planning* use:** student uses self-assessment and teacher feedback on their performance to adjust future learning strategies and goal setting for the next unit or future units of instruction.

SECTION 6.0. GLOSSARY OF TERMS

Some of the definitions below are adapted from the glossary in McTighe & Ferrara, 2021.

- **Accessibility:** A necessary condition for fairness. Accessibility means that all students have an unobstructed opportunity to demonstrate their understanding of the standards that are being assessed without being unduly advantaged or disadvantaged by individual characteristics.
- **Accessibility strategies:** An umbrella term used to describe multi-tiered strategies to support all students individually based on their unique needs and preferences. These multi-tiered strategies include universal tools, designated supports, and accommodations.
- **Accommodations:** A set of resources available to students during summative assessment administration with a documented need noted in an Individualized Education Plan (IEP) or 504 plan.
- **Alignment:** An assessment is aligned with the standards it is designed to measure if there is a content match between the standards and evidence collected and if the level of cognitive complexity of the standards matches the evidence collected.
- **Analytic rubric:** A rubric that identifies several criteria with specific qualitative descriptions of performance at each level (or score point) which evaluators score separately. For example, a piece of writing may be evaluated on several criteria, such as idea development, organization, use of details, attention to audience, and language usage and mechanics. Analytic scores may be weighted and totaled.
- **Anchor papers:** representative responses, products, or performances used to illustrate either each level (or point) on a scoring scale. They are also sometimes referred to as models or benchmark papers. Anchors for the highest score point are sometimes referred to as exemplars.
- **Anchor paper scoring:** A process of scoring calibration where anchor papers (sometimes referred to as benchmark papers) are selected and then used to promote scoring accuracy and consistency across teachers and time.
- **Authentic assessment:** students demonstrate knowledge and skills that are used in a specific field or discipline.
- **Calibrate scoring** (see scoring calibration).
- **Culturally sensitive:** Awareness that cultural differences and similarities between people exist without assigning value to them.
- **Culturally relevant:** Intentionally linkages between students' heritage and community cultural practices and the learning that takes place in schools. Students' cultural identities are linked with their academic identities when teachers act as cultural bridge builders and translators between students' everyday lived cultural experiences and the intended learning targets.

- **Culturally responsive:** Schools adapt to students. Students have many assets that can be leveraged, and school can be adapted to the students who walk through the classroom doors.
- **Culturally sustaining:** Students' heritage and community cultural practices are resources to honor, explore, *and extend*. Youth are producers of culture as well as consumers.
- **Designated supports:** A set of tools available to students during summative assessment administration for whom a need has been identified by school personnel familiar with each student's needs.
- **Double-blind scoring:** A process of scoring calibration where at least two teachers independently score the same piece of student work using a common rubric without sharing scores until after both have completed. Teachers then discuss independent ratings and any adjacent or discrepant scores until they come to a common understanding of proficiency and a common interpretation of the rubric. Double-blind scoring promotes consistency in scoring across teachers.
- **Fairness:** A key consideration in all phases of assessment construction, administration, and scoring. Fairness means an assessment does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct (or what is being assessed).
- **Formative assessment:** Ongoing process used by teachers and students during instruction that provides feedback to adjust teaching and learning to improve students' achievement of intended instructional outcomes. Formative assessment processes are for the purpose of informing teachers and students, not for evaluating–which is why formative assessment is often referred to as assessment *for* learning. Thus, results of these assessment processes should not be graded.
- **High-stakes test:** A test may be labeled "high stakes" if there are notable consequences based on the results. Standardized accountability tests are often considered high stakes for districts and schools if their results influence ratings, educational funding, accreditation, property values in a community, etc. For students, their performance on a high-stakes test could impact consequential decisions such as promotion, graduation, admission, certification, evaluation, or awards.
- **Holistic rubric:** A rubric that integrates all criteria into one qualitative description of student performance at each level (or score point) which is used to provide a single score or judgment about the overall quality of the student response, product, or performance.
- **Performance assessment (or performance-based assessment):** students apply complex knowledge and skills to construct an answer, produce a product, or perform an activity.
- **Performance task (or performance-based task):** An assessment activity, or set of activities, that elicits one or more responses to a question or problem.
- **Performance-based instructional task:** A performance task used for instructional purposes.
- **Rubric:** A scoring tool used to evaluate a student's performance in a content area. Rubrics consiste of a fixed scale (e.g., a 4-point scale) and a list of criteria that describes the

characteristics of products or performances for each score point. Rubrics are sometimes accompanied by examples (anchor papers) of responses, products, or performances to illustrate each of the points on the scale.

- **Scoring annotations:** Notes or annotations added to anchor papers or student work samples to explain rubric scores. Evidence from the student work should be explicitly connected to descriptors in the rubric to justify scores.
- **Scoring calibration:** Calibration is a process that allows multiple scorers to come to a shared understanding of how to interpret student work relative to scoring guides or rubrics. Calibration can lead to improved score quality: defined as scoring accuracy and consistency across teachers/scorers and time. Double-blind scoring and anchor paper scoring are two methods for calibrating scores.
- **Standardized assessment**: An assessment that uses a set of consistent procedures for constructing, administering, and scoring. The goal of standardization is that all students are assessed under uniform conditions so that interpretation of their performance is comparable and not influenced by differing conditions.
- **Student work (or student work sample):** A student's response to the performance task(s).
- **Summative classroom assessment** (see summative assessment)
- **Summative assessment:** Culminating assessment for a unit, grade level, or course of study used for evaluating the degree of mastery or proficiency according to identified content standards. Summative assessment documents student achievement of identified content standards at a point in time: sometimes referred to as assessment *of* learning. Unlike formative assessments, the results of summative assessments are typically evaluated and reported (e.g., as a grade, proficiency level, or percentage). High-stakes tests and summative classroom assessments are two examples of summative purposes, yet with different purposes, consequences and associated designs.
- **Task bank:** A collection (or bank) of performance assessments in different content areas and grades.
- **Universal tools:** A set of tools available to all students during summative assessment administration.
- **Universal design for assessment:** Universally designed assessments are designed and developed from the beginning to allow participation of the widest possible range of students, and to result in valid inferences about performance for all students who participate in the assessment (see Johnstone et al., 2006; Thompson et al., 2002).

SECTION 7.0: REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Education Research Association and National Academy of Education.

Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. Review of Educational Research, 86(1), 163–206.

Baker, E. L., & Gordon, E. W. (2014). From the assessment OF education to the assessment FOR education: Policy and futures. Teachers College Record, 116(11).

Center for Assessment. (2020). Assessing 21st century skills resources. Dover, NH: Author. Retrieved from https://www.nciea.org/library/assessing-21st-century-skills-resources

Darling-Hammond, L., & Adamson, F. (2010). Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Education Week. (2019). What is performance assessment? Retrieved from https://www.edweek.org/teaching-learning/what-is-performance-assessment/2019/02#:~:text=Performance task%3A A piece of,consists of several performance tasks.

Evans, C. M. (2021). Culturally sensitive, relevant, responsive, and sustaining assessment: Are there limits to making large-scale standardized testing culturally responsive? (Blog). Retrieved from https://www.nciea.org/blog/culturally-responsive/culturally-sensitive-relevant-responsive-and-sustaining-assessment

Gay, G. (2018). Culturally responsive teaching: Theory, research, and practice (3rd ed.). New York, NY: Teachers College Press.

Johnstone, C. J., Altman, J., Thurlow, M., & Moore, M. (2006). Universal design online manual. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.info/Resources/publications/UDmanual/UDmanualPrint.htm

Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. Theory Into Practice, 34(3), 159–165.

Lane, S. (2010). Performance assessment: The state of the art. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Lane, S., & Stone, C. A. (2006). Performance assessment. In R. L. Brennan (Ed.), Educational Measurement (4th ed, pp. 387–431). Westport, CT: American Council on Education and Praeger Publishers.

Marion, S. F., & Buckley, K. (2016). Design and implementation considerations of performance-based and authentic assessments for use in accountability systems. In H. Braun (Ed.), Meeting challenges to measurement in an era of accountability (pp. 49–76). New York, NY: Routledge.

McTighe, J., & Ferrara, S. (2021). Assessing student learning by design: Principles and practices for teachers and school leaders. New York, NY: Teachers College Press.

National Research Council. (2011). Assessing 21st century skills: Summary of a Workshop. Washington, DC: The National Academies Press. https://doi.org/10.17226/13215

National Research Council. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. Washington, DC: National Academies Press. https://doi.org/10.17226/13398

Stiggins, R. (1997). Student-centered classroom assessment (2nd ed.). Macmillan.

Stembridge, A. (2020). Culturally responsive education in the classroom: An equity framework for pedagogy. New York, NY: Routledge.

Taylor, C., Kokka, K., Darling-Hammond, L., Dieckmann, J., Santana Pacheco, V., Sandler, S., & Bae, S. (2016). Student engagement: A framework for on-demand performance assessment tasks. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved from https://edpolicy.stanford.edu/sites/default/files/publications/student-engagement-framework.pdf

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). Universal design applied to large scale assessments (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.info/Resources/publications/OnlinePubs/Synthesis44.html

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. The Phi Delta Kappan, 70(9), 703–713.

SECTION 8.0: APPENDICES

8.1. Nine Principles of Scoring Student Work

Adapted from a tool developed for the Literacy Design Collaborative by Measured Progress and the Stanford Center for Assessment, Learning, and Equity.

1. Know and stick to the rubric. Stick to the text. Every score must be an attempt to apply the rubric's language and meaning.

2. Trust evidence, not intuition. Calibration with other scorers requires us to base our judgments on the evidence that everyone can see, not on what a particular person feels or thinks the student might know even if he/she hasn't shown it.

3. Match evidence to language in the rubric and to the anchor papers. Try to find the direct correspondence between the rubric and the student work product itself. Further, it is important that you try to make sure that the score you give to the particular paper has the features that closely match one or more of the anchor papers for that score point.

4. Weigh evidence carefully; base judgments on the preponderance of evidence. Within each scoring dimension, the score must be based on the overall performance as evidenced throughout the work product. Therefore, the score is not based on the student's best or worst moment; rather, the score reflects what is generally true, and best matches the anchor paper, about the student's overall performance within each of the scoring dimensions.

5. Know your biases; leave them at the door. Scorers must be mindful of how biases can trigger first impressions that can color all judgments that follow. The violation of a cherished grammar rule, for example, must not blind you to all other grammatical aspects the student handled correctly.

6. Focus on what the student does, not on what the student does not do. Scorers who attend to what is in the student work product, rather than what is not or what is missing, tend to score more accurately. That shouldn't surprise us: It is easier to agree on what is than on what could be. A score is always based on what is.

7. Isolate your judgment: One bad element does not equal a bad paper. Problems in student work samples often affect the overall evaluation experience. But an analytic rubric is not designed to assess the overall experience. Rather, it is isolating variables, distinguishing between relative strengths and weaknesses. Certain work products will require that you invest more cognitive work into their scoring. Be sure not to be overly punitive in scoring those products, and be mindful that a student's poor performance in one scoring dimension does not cloud your judgment on the scoring of other dimensions.

8. Resist seduction: One good element does not equal a good paper. It also works the other way. You read an insightful and fluidly written introduction, and after that the writer can do no wrong. (This is known as the "halo effect.") One exceptional insight does not cancel out the many vague points the student does not develop.

9. Recognize direct copy or plagiarism. Be sure to distinguish between the use of quotes in support of the student's ideas and what may be intentional copying of the author's words.

8.2. Scoring Calibration Protocols

See the video library for an explanation of these protocols.

Double Blind Scoring

- Teacher 1 scores a student work sample using a rubric. Writes scores on a post-it note and then folds it in half so the scores are hidden.
- Teacher 2 scores the same student work sample using the same rubric. Writes scores on the outside of the post-it note.
- Teachers 1 & 2 discuss scoring differences using evidence from the student work and the rubric. Keep double scoring until scores are consistent. Additional teachers can be added to this protocol, as desired.

Anchor Paper Scoring

Step 1: Identification of Benchmark/Anchor Papers

Purpose: This step is designed to help make the rubrics "real" by identifying papers that can serve as benchmarks or anchors for scoring this year and in subsequent years.

Process:

- Content or grade-level teams within each school/district select approximately 20-30 student work samples from the performance assessment that represent the range of possible scores and that generally represent the distribution of student scores (e.g., if there are twice as many 3s as 2s, the sample should include twice as many 3s as 2s). This process can be initiated by having each teacher select a handful of papers representing the distribution in their class.
- The teachers review each of the papers to try to identify and agree upon prototypical (benchmark) papers at each score point (4, 3, 2, & 1). It is helpful to have more than one benchmark paper for each score point, otherwise future scorers might think the only way to get that score is to do the exact things done in the benchmark paper.
- Recognizing that most of the rubrics contain multiple dimensions, it would be ideal to have benchmark papers identified for each score point for each dimension. This might not require unique papers for each point because certain papers might serve as benchmarks for multiple dimensions.

Products: Copies of the one or more benchmark papers for each score point by dimension.

Step 2: Scoring Practice and Qualification in Teams

Purpose: This step is designed to ensure that all scorers receive appropriate training and practice before they begin scoring actual papers. Further, this step can be used to document that scorers are certified prior to scoring student papers.

Process:

- Content or grade-level teams score five performance assessments together, using the benchmark papers produced in Step 1.
- The group of scorers should discuss why they scored each paper the way they did to gain practice in scoring including striving toward agreement on scores. This should even be done for papers on which they agreed to make sure they agreed for the same reasons.
- If the team is not reaching consensus on scoring, then the team should score an additional five work products until consensus is reached consistently.

Products: Copies of the five (or more) performance assessments scored together with the scores and any notes on scoring decisions.

Step 3: Individual Teacher Scoring

Purpose: This step is the major work of scoring for classroom and school uses of the scores.

Process:

- Read the Nine Principles of Scoring Student Work (below).
- After the teachers have demonstrated that they can score to consensus with their peers, each teacher then scores his/her student responses for each of their students who completed the task.
- Teachers should use the anchor papers to match student work to score points by rubric dimension. The anchor papers can be used to help decide between adjacent score points. For example, teachers can ask themselves, "Does this work look more like the anchor paper for score 2 or score 3 for this rubric dimension?" This step will help ensure that teachers' scores are consistent within schools/districts.

Products: Score data (by rubric criteria) for each student response.

Step 4: Teacher Re-Calibration (Optional)

Purpose: This is an optional, but recommended, step that is used to help ensure that scorers do not "drift" over time, something that is very common in scoring. This step allows schools/districts to document that scorers have not drifted in their interpretation of the rubric over time.

Process:

- After each teacher scores about one-half of their papers, they should get together with at least one other grade-level teacher (more is better).
- The teachers should select at least 5 papers, some from each participating teacher, and all participating teachers should score the full set of papers.
- The teachers should be able to score to consensus on this calibration check. If not, they should select another small set of papers (e.g., 4-5) until they can reach agreement.

8.3. Student Work Analysis Protocol for Instructional Purposes

Goals:

- Analyze student work to diagnose student strengths and needs
- Based on student work determine instructional needs of students
- Use evidence from student work to inform revision of lessons for next year or inform upcoming unit of instruction

General instructions:

This protocol provides a *suggested* order for analyzing student work to diagnose student strengths and weaknesses for instructional purposes. The protocol could be used based on one class of students or by pooling all students together from a grade level/course team. We recommend teachers use this protocol with other grade level or grade span teachers, though an individual teacher may also modify this protocol to complete the process independently.

Materials needed:

- Class set of student work and copies for sharing
- Assessment Task/Prompt
- Rubric

1. Review Assessment and Identify Expectations (approximately 5 minutes)

- What texts or other materials were students expected to read and to draw evidence from, if applicable?
- Did the assessment prompt provide students an opportunity to demonstrate what they know, can do, and understand about the concepts and skills?

2. Reach Consensus about Proficiency (approximately 5 minutes)

- Describe what you consider to be a proficient response to this task.
- Exactly what do students need to demonstrate for you to consider their work proficient?

3. Student Work: Sorting Student Work (approximately 30 minutes)

- Individually, read the student work samples and without scoring, do a "quick sort" of students' work by the general degree of the **high, average, low**. A "not sure" pile may be needed. After sorting, any papers in the "not sure" pile should be matched with the typical papers in one of the other existing piles. Student work in the "high" pile may not constitute proficiency, but rather demonstrate the strongest response in the class.

- The teachers should compare which piles they placed the students' work. The presenting teacher should discuss the rationale used for placing the student work in a pile when there is disagreement, providing evidence from the student work to justify the thinking. Teachers should refer to the rubric or scoring criteria for expectations of the task when unsure.
- Consensus should be reached at this time and student names should be recorded in the columns below in order for the teacher to monitor his/her own students' progress over time.

HIGH	AVERAGE	LOW
_____% OF CLASS	_____% OF CLASS	_____% OF CLASS

- Compare the students at each level to where they began the year. Discuss the students' progress: Why do you think students are making progress? Why do you think they are not making progress?

4. Diagnosing Student Strengths (approximately 7 minutes)

Review multiple samples (approximately 4 papers) from each level (high, average, low) to discuss. Identify the prerequisite knowledge that students demonstrated about the expectations found in the standards. Record the students' strengths in the chart below – be specific.

HIGH	AVERAGE	LOW

5. Diagnosing Student Needs (approximately 7 minutes)

Using the reviewed samples from each level, discuss and identify the misconceptions, wrong information, and what students did not demonstrate that was expected. Record the students' needs in the chart below – be specific. This is not intended to be a laundry list of everything students did not do, but rather what needs they have within their zone of proximal development or what you would consider to be the next set of instructional needs.

HIGH	AVERAGE	LOW

6. Identifying Instructional Next Steps (approximately 10 minutes)

After diagnosing what the student knows and still needs to learn, discuss the learning needs for the students at each level considering the following questions.

- What patterns are noted for the whole class?

_____

_____

_____

_____

- What strategies will be beneficial for the whole class?

_____

_____

_____

_____

Based on the group's diagnosis of student responses at the high, average, and low levels, what specific strategies will be beneficial for students at each level?

HIGH	AVERAGE	LOW

7. Reflection on Task (approximately 15 minutes)

- Were there any unexpected student responses that should be considered that haven't been discussed?
- Did the student work demonstrate what was expected? If not, are there any recommended changes to the assignment and/or the prompt?

8. Whole Group Debrief (approximately 15 minutes)

- Did the student work demonstrate what was expected? If not, why do you think this occurred?
- Were there any unexpected student responses that should be discussed for the good of the whole group?
- How can the information gained from this Student Work Analysis inform your overall instructional practice?