**Facilitator Guide**

**Module 3 - Score**

To meet the following objectives, participants will use a pre-existing assessment or create a new assessment that was utilized in Module 1 and 2 – Design and Build

Participants will be able to:

* 1. Develop scoring keys and rubrics for assessment items and tasks.

**Module 3.1 Scoring Tools: Scoring Keys, Scoring Rubrics**

**SLIDE 11**

**3.1.1 Procedural Steps: Develop a Scoring Key (SR Items)**

1. Enter the assessment information at the top of the *Scoring Key Template*.
2. Record the item number, item tag (optional), item type, and point value.
3. Record the MC answers in the *Answer* column.
4. Repeat Steps 1-3 until all items on the *test specifications* are reflected within the *Scoring Key Template*.
5. Validate that each question-to-answer relationship is recorded correctly.

**3.1.1 Scoring Key-Social Studies Grade 4 (Example)**

**SLIDE 12**

|  |  |  |  |
| --- | --- | --- | --- |
| **Assessment Name** | **Grade/Course** | **Administration** | **Total Possible Points** |
| **Social Studies Final** | **Elementary Social Studies-Grade 4** | **Post-test** | **50** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item #** | **Item Tag** | **Item Type** | **Point Value** | **Answer** |
| 1 | 0001.SOS.GR4.POST.MC-LV1-GEO2.1 | MC | 1 | B |
| 2 | 0002.SOS.GR4.POST.MC-LV1-GEO2.4 | MC | 1 | A |

**3.1.1 Scoring Key: QA Checklist**

**SLIDE 13**

|  |  |
| --- | --- |
| **Task** | **Task Question** |
| **Articulation** | Are all items articulated in the test specifications represented within the Scoring Key? |
| **Validation** | Is there only one correct answer for every item? |
| **Queuing** | Are the answers arranged in a way that does not create a discernible pattern? |
| **Balance** | Are the answers “balanced” among the possible options? |
| **Revalidation** | Are the answers revalidated after the assessment has been reviewed? |

**Module 3.1.2 Scoring Rubrics**

**SLIDE 14**

Scoring Rubrics for use with:

Short Constructed Response (SCR)

SCR Stand-Alone Items

SCR Passage Based Items

Extended Constructed Response (ECR)

ECR Stand-Alone Items

Performance Task (PT) Items

Performance Task Items

**SLIDE 15**

***Scoring rubrics*** are tools used to measure and evaluate students’ achievement of a task. Rubrics are able to score items and tasks based on one or more dimensions. These dimensions are often called “criteria.”

There are two major types of scoring rubrics: **holistic** and **analytic.**

In **holistic** scoring, the scorer is combining multiple aspects or criteria with a particular classification scheme. This means that the performance is judged in its totality and assigned a point value based on a single dimension or criteria.

Conversely, **analytic** rubrics define key criteria either within or across multiple dimensions and assign point values. These point values are aggregated into the classification scheme, and aggregated points are assigned to determine “pass or fail” classifications.

**SLIDE 16**

**GUIDELINES scoring rubrics**

**Holistic/single-dimension rubrics:**

* are used when performance criteria cannot be separated clearly
* combine all performance criteria for simultaneous evaluation
* require the scorer to enter only a singular point value based on the student’s performance as a whole

**Analytic/multi-dimension rubrics:**

* are used for more complex tasks that measure many skills at once
* separate performance criteria into separate dimensions, and each criterion is scored individually
* provide every scoring level of each different criterion its own statement that specifies guidelines for attaining that specific level of achievement

**Example-SCR Holistic Rubric**

|  |
| --- |
| **SCR Passage-based Item**  **\*\*\*\* Sample #6 from Build Module 2.1.5\*\*\*\***  **In the Rainforest**  Life in the rainforest is wet and wild. Rainforests are jungle areas that have a large amount of rainfall each year. Rainforests cover a small part of this earth. They are home to over half of the different types of animals and plants. Rainforests are not only beautiful, but also very important to people. We get food from the rainforest such as chocolate and cinnamon. Rainforests also have ingredients for many medicines. Even though rainforests are important, many are being cut down. Many people are working to make sure that the rainforests are saved because we will need them in the future. |
| 1. In two complete sentences, explain why people should help save the rainforests. |

**3.1.2 SCR Holistic Rubric-Reading Grade 6 (Example) SLIDE 17**

|  |  |
| --- | --- |
| **Sample Response for: “In two complete sentences, explain why people should help save the rainforests.”** | |
| **2 points** | The student’s response is written in complete sentences and contains two valid reasons for saving the rainforest.  *“People must save the rainforest to save the animals’ homes. People need to save the rainforest because we get ingredients for many medicines from there.”* |
| **1 point** | The student’s response contains only one reason.  *“People should save the rainforest because it is important and because people and animals need it.”* |
| **0 points** | There is no response, or the response is completely incorrect or irrelevant. |

**Change Holistic Rubric to an Analytic Rubric**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sentences** | **There are 2 complete sentences.** | **There is only one complete sentence.** | **There are no complete sentences.** |
| **Explanation** | **Student response contains at least one valid reason for saving the rainforest in each sentence.** | **Student response contains at least one valid reason for saving the rainforest.** | **Student response does not contain a valid reason for saving the rainforest.** |

**SLIDE 18**

**Example-ECR/PT analytic rubric**

By contrast, the analytic rubric is scored on two separate dimensions:

1. Sketch a Face and

2. Color Scheme.

Each dimension is scored separately, and all are worth up to four (4) points each. After assigning a score to each dimension, the two separate scores are added for a total score out of eight possible points.

**3.1.2 ECR Analytic Rubric-Art (Example)**

| **Dimension** | **Advanced**  **(4 pts.)** | **Proficient**  **(3 pts.)** | **Basic**  **(2 pts.)** | **Below Basic**  **(1 pt.)** |
| --- | --- | --- | --- | --- |
| **Art Elements and Principles**  **- Sketch a Face** | The drawing of the face shows a breakdown of the proportions of the facial features with a lot of attention paid to details, such as strands of hair, details in eyes. The facial expression establishes a mood. | The drawing of the face shows a breakdown of the face with facial features placed in proportion to one another. The portrait lacks facial expression, thus no mood is shown in the drawing. | The drawing of the face shows some breakdown of the face but following through with other visual cues, such as eyebrows lining up with tops of the ears, are not evident. No attention is paid to facial expression to establish mood. | The drawing of the face is not broken down into proportional segments, so the placement of the facial features is random and abstracts the realism of the portrait. No attention is paid to details, such as strands of hair, eyebrows, eyelashes, etc. |
| **Art Elements and Principles**  **- Color Scheme** | Background collage uses either a monochromatic or analogous color scheme with all ends of the pictures glued down. The arrangement of the collage is unique using either space of transition of color in a planned way. The color mixing and transition of the paint in the portrait shows a broad range of tones, and the application contours the shape of the face. | Background collage uses either a monochromatic or analogous color scheme with some ends of the pictures sticking up or small portions of the oak tag showing through. The color mixing with the paint in the portrait shows a range of either a monochromatic or analogous color scheme, but no awareness of contour color application. | Background collage has at least 60% of the paper fitting into either a monochromatic or analogous color scheme. Color mixing in the portrait is blotchy and not blended well; each color is separate with no attempt to transition one color to the next. | Background collage and color mixing show no identification to either a monochromatic or an analogous color scheme. |
| **Art Elements and Principles**  **Point of View**  **- Original Work** | Portrait and background show a full understanding of monochromatic and analogous color schemes. Paint mixing is gradual using a range of hues that have more than eight tones, and the tones are applied enhancing the contour and roundness of the face. | Portrait and background show understanding of monochromatic and analogous color schemes. Some edges of the collage are not glued down. The color portrait is painted with a range of six or more hues of a monochromatic or analogous color scheme. The transitions of color are smooth and gradual. | Portrait and background show some understanding of monochromatic and analogous color schemes with only one or two colors that do not fit into the selected schemes. Half of the color mixing with paint shows gradual transitions of color. | Portrait and background show no knowledge of monochromatic or analogous color schemes. There is no range of color mixing within either color scheme with no gradual transitions of color. |

**SLIDE 22**

**3.1.4-3.1.8 Procedural Steps: Develop *Assessment Scoring Rubrics* (SCR/ECR items, PT tasks)**

1. Review the *Item Framework* and the criteria articulated in the stem/directions. (From the Design and Build modules)
2. Select a rubric structure (holistic or analytic) based upon scoring criteria and the number of dimensions being measured.
3. Modify the rubric language, using the specific criteria expected in the response to award the maximum number of points.

Ensure that the criteria/guidelines articulated in the rubric are clear enough for students and other teachers to understand.

1. Determine how much the response can deviate from “fully correct” in order to earn the next (lower) point value. [Continue until the full range of possible scores is described.] For an analytic rubric, this step must be completed for each dimension/criterion being measured.
2. During the quality assurance review, ensure the response expectation, scoring rubric, and test specifications are fully aligned.

(Criteria Discovery Task: Clean the Car)

**SLIDE 23 Rubric Guidelines 2A**

Selecting Criteria

Choose criteria that assess intended learning outcomes of the standards or instructional goals you intend to assess. Effective rubrics do not list all possible criteria; they list the right criteria for the assessment’s purpose. Ask yourself this question: “What characteristics of student work would give evidence for student learning of the knowledge or skills specified in the standard or instructional goal? “

Criterion selected should demonstrate the following characteristics:

* Appropriate. Each criterion represents an aspect of a standard, curricular goal or instructional goal that students are intended to learn.
* Definable. Each criterion has a clear, agreed-upon meaning that both students and teachers understand.
* Observable. Each criterion describes a quality in the performance that can be perceived by someone other than the person performing.
* Distinct from one another. Each criterion identifies a separate aspect of the learning outcome the performance is intended to assess.
* Complete. All the criteria together describe the whole of the learning outcomes the performance is intended to assess.
* Able to support descriptions along a continuum of quality. Each criterion can be described over a range of performance levels.

**SLIDE 24 Rubric Guidelines 2B Describing Criteria Performance**

1. Describe a continuum of levels of performance quality for each criterion.

* Describe what is observed
* Clear
* Cover the whole range of performance
* Distinguish among levels
* Center the target performance (acceptable, mastery, passing) at the appropriate level
* Feature parallel descriptions from level to level

2. Avoid odd numbers of criteria descriptor levels.

3. Identify and describe the Proficient Level first.

4. Describe student performance in terms that allow for many different paths to success.

Facilitator Notes:

Once criteria have been selected, the range of performance levels must be described. Ask yourself this question: “What does student work look like at each level of quality, from high to low, on this criterion?”

Rubrics should describe a continuum of levels of performance quality for each criterion. While evaluative terms like “excellent, good, fair and poor” may be used as rating categories, they should be avoided in describing performance. The descriptions should be clear and based on the same elements of performance from level to level. Consider these characteristics for descriptions of levels of performance:

* The performance is described in terms of what is observed in the work.
* Both students and teachers clearly understand what the descriptions mean.
* The performance is described from one extreme of the continuum of quality to another for each criterion
* The performance descriptions are different enough from level to level that work can be categorized unambiguously. It should be possible to match examples of work to performance descriptions at each level.
* The description of performance at the level expected by the standard or curriculum goal is placed at the intended level on the rubric.
* Performance descriptions at each level of the continuum for a given standard describe different quality levels for the same aspects, or criterion, of the work.

Avoid odd numbers of criteria descriptor levels. Three and five tiered rubrics cause language in the middle tier to be unclear as to whether or not proficiency has been achieved. Additionally, raters tend to drift to this tier when they are unsure about the response’s relationship to the descriptor language.

Identify and describe the proficient level first, and build the rest of the rubric around proficiency. In a four-tier rubric, a “3” or proficient level is an acceptable score and shows proficiency at performing the task or understanding the content. From this point, building the remainder of the rubric is more readily defined: a “1” or failing level shows minimal understanding or performance; a “2” or needs improvement level shows some understanding or performance, but with significant gaps; a “4” or distinguished level shows advanced understanding or performance.

When complete, descriptor language should describe student performance in terms that allow for many different paths to success.

**SLIDE 25 Rubric Guidelines 2C Designing Rubrics**

**Top Down Approach**

1. Create (or adapt from existing) a conceptual framework for achievement.
2. Write general scoring rubrics fusing these dimensions and performance levels.
3. Adapt general scoring rubrics for the specific learning goal of the performance being scored.
4. Use the rubrics to assess several student performances and adapt as needed for final use.

**Bottom Up Approach**

1. Get a dozen or more copies of student work.
2. Sort into three piles: high, medium and low quality work
3. Write specific descriptions of why each piece of work is categorized as it is.
4. Compare and contrast the descriptions and extract criteria/dimensions.
5. For each criterion identified in step 4, write descriptions of quality for that criterion.

**Facilitator Notes:**

Designing rubrics can be a daunting task. The task can be simplified, however, by considering one of two different approaches:

The “top down” approach begins with a model rubric framework in mind, and encourages adaptation of that rubric to meet the demands of the content area, standards and anticipated student achievement as requested in the assessment task.

The “bottom up” approach begins with student achievement samples and develops a process where criteria and descriptions of student work emerges from those samples. This approach is similar an “identifying exemplars” and “human scoring calibration process” that will be presented later in this module component and in module 3.3.

While neither process is perfect, both have merits toward understanding the need to carefully “wordsmith” rubric language to guarantee that this type of scoring tool accurately measures student achievement of content standards as identified in the assessment’s design and clarified in the task’s Item Framework.

**SLIDE 26 Quality Assurance Checklist**

**3.1.2 Scoring Rubric: QA Checklist**

|  |  |
| --- | --- |
| **Task** | **Task Question** |
| **Targeted Content Standards** | Does the rubric reflect a performance continuum? |
| **Developmentally Appropriate** | Is the rubric clear and concise? |
| **Aligned to Task** | Does the rubric provide all dimensions (components) of the task? |
| **Criteria** | Does the rubric include expectations for a “fully correct” response? |
| **Potential Bias** | Does the rubric omit non-cognitive attributes (e.g., motivation, timeliness)? |
| **Editing** | Have editorial correctness and Universal Design principles been applied? |

****

**Design Activity to support analytic vs holistic rubrics and how to determine which to use**

**MODULE 3.2 SCORING GUIDES**

**SLIDE 27**

Scoring guides are collections of guidelines and instructions used to ensure understanding of how to use a particular item’s scoring rubric. Module 3.2 will provide guidelines for all eight types of items presented in Module 2.

**SLIDE 29 Scoring Guide Protocols** for

1. SR Stand-Alone items

2. SR Passage Based items

3. SR Evidence Based items

Answers are scored by an **answer key,** since the type and style of the correct response has been pre-selected.

Scoring guide protocols:

* Word questions positively,
* include one correct answer with plausible distractors,
* create three options for grades K-2 and four for grades 3-12,
* keep answer options of the same length and structure,
* use consistent grammar for all options,
* avoid “all of the above” and “none of the above,”
* arrange the answer options in a numeric or chronological order
* place charts-tables-and graphs within the item as much as possible.
* avoid overlapping alternatives, where one answer is part of another. This frequently happens with answers that are numbers.
* Avoid combining options, as in “both A and B.”

**SLIDE 30 Scoring Guide Protocols** for

4. SCR Stand-alone Items  
5. SCR Passage Based Items

* Presents a prompt, scenario, or passage using content-specific terminology, along with developmentally appropriate references.
* Requires a brief (2-5 minute) response.
* Typically solved in one or two steps
* Focused on DoK Levels 1 or 2.
* Ensure the prompt, passage, or scenario is concise, free from jargon, and grammatically correct.
* Needs rubrics/human scorers to evaluate answer.
* Use appropriate verbs to communicate expectations.
* Articulate response instructions using a clear sentence structure.
* Communicate in clear, unambiguous terms the extent of the expected answer.

**Facilitator Notes:**

Constructed Response items, unlike selected response items, eliminate the option of guessing. Constructed response items, as compared to selected response items, are often seen by teachers as a quick and easy to develop. This perception can quickly change, however, once it is time to score constructed response items.

Short constructed response items present a prompt, scenario, or passage using content-specific terminology, along with developmentally appropriate references that require a brief—no more than two to five minutes of the student’s test-taking time--one or two step response. These items generally assess learning at DoK levels 1 and 2. When writing short constructed response items, ensure that the prompt, passage, or scenario is concise, free from jargon, and grammatically correct.

The scoring of these types of questions, however, should be done with a rubric, so time and energy saved in item stem development is reallocated to preparing and implementing the scoring process. Simple expression of “right or wrong” may seem intuitive and uncomplicated, but does not provide students with appropriate feedback as to ways in which achievement has or has not been demonstrated. Students should be presented with rubrics so that they understand the extent of the expected answer, and human scorers should have clear understanding of the ways in which the rubric is to be utilized toward scoring student responses.

Short constructed response items are often scored using two- to four-tiered holistic rubrics. When writing these rubrics, use appropriate verbs to communicate expectations. Articulate response instructions using a clear sentence structure, and communicate in clear, unambiguous terms the extent of the expected answer.

**SLIDE 31 Scoring Guide Protocols** for   
 6. ECR Stand-alone Items

* Can assess student learning at DoK level 3
* Solved using multiple steps and often organized and presented to the test-taker in parts (e.g., Part A, Part B)
* Worth four or more points towards the overall score
* Requires 5-10 minutes to answer
* Requires significant human scorer time and effort
* Typically requires exemplars and scoring calibration efforts when multiple scorers are used.

**Facilitator Notes:**

Rubrics for extended constructed response items typically require exemplars of student work as well as rater scoring calibration efforts when multiple scorers are used.

**SLIDE 32 Scoring Guide Protocols** for   
 7. Extended Constructed Response (ECR) TDA Items

* Often DoK Level 3 or 4.
* Requires reading and analyzing a text in order to respond to a given prompt or scenario.
* Requires students to create informed claims or arguments about a text, and to support these claims with evidence found specifically within the same text
* Goes beyond literal understanding, and often focuses upon a small portion (sometimes even a word or phrase) of a text.

**Facilitator Notes:**

Scoring guidelines for text dependent analysis items can evaluate student learning at DoK levels 3 and 4. The item prompt requires that students read and analyze a text and create informed claims or arguments about the text, supporting their claims with evidence found in the text. This analysis, and subsequent scoring of the student response, goes beyond literal extraction of facts from the text.

**SLIDE 33 Scoring Guide Protocols** for   
 8. Performance Task Items

* Measure high levels of DoK (e.g., Level 4).
* Aligned to multiple tasks for several targeted content standards.
* Administered over an extended time period.
* Typically a culminating event/project (e.g., final portfolio, performance, or project).
* Often used in conjunction with “on-demand” items/tasks (i.e., SR, SCR, ECR) at key phases/milestones.

**Facilitator Notes:**

Performance Task items often measure learning at DoK level 4.

Frequently, multiple tasks are used and several targeted content standards are assessed.

These tasks are administered over an extended period of time and often culminate in the form of a final portfolio, project or performance.

It can also be a common practice to incorporate other types of items at key phases and milestones toward completing a performance task.

* **SLIDE 34** **Optional: for reinforcing skill.**
* Improve the following multiple response choices based on the selected response scoring guidelines presented in Slide 25. *(As a follow-up to module 2, it would be appropriate to improve the item stems as well!)*

**OPTIONAL:**

Improve the following multiple response choices based on the selected response scoring guidelines presented in Slide 29. Make recommendations for additions or deletions to meet the criteria of the scoring guidelines. *(As a follow-up to module 2, it would be appropriate to improve the item stems as well!)*

(The questions come from an online “American Trivia” source, [*http://www.triviacountry.com/M1-Multiple-Choice-Trivia-Questions.htm*](http://www.triviacountry.com/M1-Multiple-Choice-Trivia-Questions.htm) , so are not intended to be linked to any specific standards or grade levels. )

1. In the year 1900 in the U.S. what were the most popular first names given to boy and girl babies?  
A. William and Elizabeth  
B. Joseph & Catherine  
C. **John and Mary**  
D. George/Anne

2. When did the Liberty Bell get its name?  
A. when it was made, in 1701  
B. when it rang on July 4, 1776  
C. **in the 19th century, when it became a symbol of the abolition of slavery**  
D. none of the above

3. In 1985, five percent of U.S. households had telephone answering machines. By 1990 what percentage of homes had answering machines?  
A. 15 percent  
B. 10 percent  
C. **31 percent**  
D. 51 percent

4. Which of these characters turned 50 years old in 2000?  
A. **Charlie Brown**  
B. Bugs Bunny  
C. Mickey Mouse  
D. Goofy

5. Before becoming George Bush's Secretary of Defense, what was Dick Cheney's position?  
**A. congressman from Wyoming**  
B. governor of New Hampshire  
C. secretary of defense under Ronald Reagan

**SLIDES 37-38**

**Scoring Guide Procedural Guidelines  
For SCR, ECR and PT items (Scoring Rubrics)**

1. Review the *Item, TDA or Performance Task Framework* and apply the Quality Control checklist.

2. a. Develop a clear statement that articulates specific criteria for the test-taker to provide.

b. Develop a clear statement for each subordinate task that articulates specific criteria for the test-taker to provide.

c. Develop a clear statement for the final product and for each subordinate task that articulates specific criteria for the test-taker to provide.

d. Select and develop an appropriate rubric type.

|  |  |
| --- | --- |
| **SCR Stand-Alone** | SCR Holistic |
| **SCR Passage Based** | SCR Holistic |
| **ECR Stand-Alone** | ECR Holistic, ECR Analytic |
| **ECR Text Dependent Analysis** | ECR Analytic TDA |
| **Performance Task** | ECR Analytic) |

3. Give clear, detailed instructions about how the item is to be scored through use of the rubric.

Include sample responses/exemplars.

4. Give further instructions for calculating final scores of items (e.g., multiply the raw score by 2 for the final score) when applicable. If there are special conditions for scoring or awarding point values (e.g., when to award partial credit/half of a point for a certain performance criteria), ensure that they are included in the scoring instruction and are easy for the scorer to see and comprehend.

5. Review the item and answer options for grammatical soundness.

**SLIDE 40**

Connecting Scoring Tools and Guides to Item Types  
  
**Examples**

|  |  |
| --- | --- |
| **Item Type** | **Scoring Tool** |
| **Selected Response (SR) Stand –Alone** | Scoring Key |
| **Selected Response (SR) Passage Based** | Scoring Key |
| **Selected Response (SR) Evidence Based** | Scoring Key |
| **Short Constructed Response (SCR) Stand-Alone** | SCR Holistic |
| **Short Constructed Response (SCR) Passage** | SCR Holistic |
| **Extended Constructed Response (ECR) Stand-Alone** | ECR Holistic, ECR Analytic |
| **Extended Constructed Response (ECR)** | ECR Analytic |
| **Performance Task** | ECR Analytic |

**SLIDE 41 Group Work**

Using protocol from slides 30-31 designing rubrics, adapt the following oral presentation holistic rubric to become a visual presentation (poster, PowerPoint , YouTube, science fair project, etc.) rubric.

*This learning exercise gives practice in taking an existing rubric and applying modifications that meet criteria for a specific item.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Using protocol from previous slides in designing rubrics, adapt the following oral presentation holistic rubric to become a visual presentation (poster, PowerPoint , YouTube, science fair project, etc.) rubric.  Upon completion, a colleague should review for clarity   |  |  |  | | --- | --- | --- | | ***Score*** | ***Oral Presentation Holistic Rubric*** | ***Visual Presentation Holistic Rubric*** | | 4 | The topic is addressed clearly  Speech is loud enough and easy to understand  Good eye contact  Visual aid is used effectively  Well organized |  | | 3 | The topic is addressed adequately  Speech has appropriate volume  Eye contact is intermittent  Visual aid helps presentation  Good organization |  | | 2 | The topic is addressed adequately  Speech volume is not consistent  Student reads notes-erratic eye contact  Visual aid does not enhance speech  Organization falters occasionally |  | | 1 | The topic needs more explanation  Speech is difficult to understand at times  Lack of adequate eye contact  Poor visual aid-does not contribute to understanding  Lack of organization |  | | 0 | The topic is not addressed  Speech cannot be heard or understood  No eye contact-reads entire speech  No visual aid  No evidence of organization |  | |

[](http://www.media-management.eu/wp-content/uploads/individual.png)**SLIDE 42 Individual Work**

OBJECTIVE: This learning exercise provides application experience in terminology and content from this module

Complete the following charts listing the Advantages and Disadvantages of each

item and the scoring tool.

|  |  |  |
| --- | --- | --- |
|  | **Advantages** | **Disadvantages** |
| Holistic | 1. Scoring is faster than with analytic rubrics. 2. Requires less time to achieve inter-rater reliability. 3. Good for summative assessment | 1. Single overall score does not communicate information about what to do to improve. 2. Not good for formative assessment. |
| Analytic | 1. Gives diagnostic information to teacher. 2. Gives formative feedback to students. 3. Easier to link to instruction than holistic rubrics. 4. Good for formative assessment; adaptable for summative assessment; if you need an overall score for grading, you can combine the scores. | 1. Takes more time to score than holistic rubrics. 2. Takes more time to achieve inter-rater reliability than with holistic rubrics. |

|  |  |  |
| --- | --- | --- |
| SCR Stand-Alone | Advantages | Disadvantages |
| Holistic |  |  |
| Analytic |  |  |

|  |  |  |
| --- | --- | --- |
| SCR Passage-Based | Advantages | Disadvantages |
| Holistic |  |  |
| Analytic |  |  |

|  |  |  |
| --- | --- | --- |
| ECR Stand-Alone | Advantages | Disadvantages |
| Holistic |  |  |
| Analytic |  |  |

|  |  |  |
| --- | --- | --- |
| Performance Task | Advantages | Disadvantages |
| Holistic |  |  |
| Analytic |  |  |

**SLIDE 47 Example of Rubric**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item #1 Sample Response for: *In the letter, Adams informs her daughter about the experience of moving to a new city and living in a new home. Write an essay analyzing how Adams responds to her new surroundings. Use evidence from the letter to support your response.* | | | | |
| Dimension | Advanced  (4 points) | Proficient  (3 points) | Basic  (2 points) | Below Basic  (1 point) |
| Response to Prompt  (4 points) | Central claim is clearly articulated and completely addresses the writing prompt | Central claim is clearly articulated and mostly addresses the writing prompt | Central claim is ambiguous/unclear and partially addresses the writing prompt | An effort was made to create a central claim, but it does not address the writing prompt |
| Textual Evidence  (4 points) | Student uses sufficient (i.e., at least one instance per body paragraph) and relevant textual evidence to support the claim | Student uses relevant evidence to support the claim, but does not address all parts of the prompt with supporting textual evidence (i.e., one or more body paragraphs are missing textual evidence) | Student attempts to use sufficient (i.e., at least one instance per body paragraph) evidence to support the claim, but the evidence presented is irrelevant to the claim | Student attempts to use insufficient (i.e., one or more body paragraphs are missing textual evidence) evidence to support the claim, and the evidence presented is irrelevant to the claim |
| Structure  (4 points) | Response contains a clear beginning, middle, and end, and uses transitions to ensure these parts flow together and stay on-topic; textual evidence is integrated smoothly and its relevance is justified in the response | Student uses relevant evidence to support the claim, but does not address all parts of the prompt with supporting textual evidence (i.e., one or more body paragraphs are missing textual evidence) | Student attempts to use sufficient (i.e., at least one instance per body paragraph) evidence to support the claim, but the evidence presented is irrelevant to the claim | Student attempts to use insufficient (i.e., one or more body paragraphs are missing textual evidence) evidence to support the claim, and the evidence presented is irrelevant to the claim |
| Conventions  (4 points) | Student’s language is clear and appropriate for the task; there are up to two minor errors in spelling, grammar, and/or usage | Student’s language is clear and appropriate for the task; there is a major error and other minor errors in spelling, grammar, and/or usage | Student’s language is either clear OR appropriate for the task; there are up to three major errors in spelling, grammar, and/or usage | Student’s language is neither clear nor appropriate; there are up to five major errors in spelling, grammar, and/or usage |

**Developing Scoring Frameworks Task: Draw a Snowman**

**SLIDE 49 Grouping the samples into three performance groups:**

**High, mid and low level.**

Once the task have been assigned a score, the responses should be separated into three groups, or buckets, each bucket identifying a performance level that the scorer believes the response has demonstrated.

If the set of responses reflects the normal curve, a larger percent of responses will be in the “middle” bucket.

**SLIDE 52** Module 3.2 QA checklist

|  |  |
| --- | --- |
| **Task** | **Task Question** |
| **Targeted Content Standards** | **To what degree does this item match the targeted standards?** |
| **Cognitive Level** | **To what degree does this item match the DoK expressed**  **in the standards?** |
| **Developmentally Appropriate** | **Are the readability and task requirements appropriate**  **for the test-takers?** |
| **Sensitive Material** | **Is there sensitive content with references to drugs,**  **death, suicide, etc.?** |
| **Potential Bias** | **Are there contextual, gender, or cultural assumptions?** |
| **Fairness** | **Has the test-taker had the opportunity to learn the**  **content within the item?** |
| **Editing** | **Have editorial correctness and Universal Design**  **principles been applied?** |

**Facillitator note:** A quality assurance review is appropriate for scoring keys and scoring rubrics. The concepts under scrutiny are similar to those for item stems, prompts, and task descriptions. The Quality Assurance checklist

1 Ensures targeted standards are consistent between the item and its selected response or rubric. The standard being evaluated should be present in the item, and the test-taker’s response should be evaluated by the selected response or the item’s rubric.

2 Ensures the DoK is consistent between the item and its selected response or rubric, and that the response or rubric assesses the item in a way that reflects the desired DoK of the item.

3 Creates scoring guidelines that the test-taker and assigned human scorers can understand.

4 Excludes sensitive material and/or criteria in the selected response or rubric that can compromise a student’s performance.

5 Excludes criteria in the selected response or rubric that is based on non-cognitive attributes and that may be biased or may favor certain populations.

6 Excludes performance criteria that are dependent upon knowledge that the test-taker has not been given a chance to learn or, in items that involve reading passages, provides the test taker the opportunity to memorize or rehearse a response that is unrelated to the construct being assessed.

7 Reviews the scoring guide and ensures it is grammatically sound.

**SLIDE 56 Sharing with Colleagues**

* Assessment design, build and score should be considered a team effort as the learning occurs.
* By growing together, the skills become sharper and assessed content/skills becomes more accurate.

**SLIDE 57 Sharing with students**

* Please make clear to students how and when they will be assessed, what they will be required to do, and when and how they will be evaluated.
* Directions to students, assessment tasks, and the scoring rubrics should be understandable to all students.

**SLIDE 58 Putting it all Together**

**SLIDE 58 Module 1 Design**

* Depth of Knowledge
* Content to be assessed
* Purpose of Assessment

**SLIDE 59 Module 2 Build**

1. Develop assessment items and tasks.

2. Construct an operational test form.

Learned vocabulary of Assessment construction:

**SR-**Selected Response: items that are multiple-choice, true-false, matching, any item where the test-taker has a bank of answers from which they choose an answer.

Selected response items may be:

* + - * Stand-Alone
      * Passage-Based
      * Evidence-Based

**SCR-**Short Constructed Response/Short Answer: the test-taker provides a one word or short phrase answer

Short Constructed Response may be:

* + - * Stand-Alone
      * Passage-Based

**ECR-**Extended Constructed Response: the test-taker writes a paragraph or two, or develops an authentic response within a class period’s time length

Extended Constructed Response may be:

* + - * Stand-Alone

**PT-** Performance Task: the test-taker develops a response that takes several class periods or the response is gathered over a period of time, as in a portfolio assessment.

A Performance Task may be:

* Multi-Day Task

**SLIDE 60 Module 3 Score**

**Participants will be able to:**

1. Develop scoring keys and rubrics for assessment items and tasks.

**SLIDE 61 NEXT STEPS**

Each educator will apply and implement new skills throughout the remainder of their career.

Assessment building is a skill. It is through designing, building and scoring each assessment that is given throughout the year that the skill becomes easier and greater success in purpose is achieved.

**PRACTICE MAKES PERMANENT**

Thank you for your dedication and support of the education of the students in the Commonwealth of Pennsylvania.

Pennsylvania Department of Education

The Bureau of Curriculum, Assessment and Instruction

Division of Instructional Quality

Content Advisors:

* + David Deitz, Arts and Humanities
  + Kevin Mauro, Math
  + Dr. David Bauman, Science
  + Sally Flaherty, Social Studies

**SAMPLES**

**Extended Constructed Response Rubric**

|  |  |
| --- | --- |
| 3 points | The response demonstrates a comprehensive understanding and analysis of a problem.  • Application of a reasonable strategy in the context of the problem is indicated.  • Explanation of and/or justification for the mathematical process(es) used to solve a problem is clear, fully developed, and logical.  • Connections and/or extensions made within mathematics or outside of mathematics are clear and stated explicitly.  • Supportive information and/or numbers are provided as appropriate. |
| 2 points | The response demonstrates a general understanding and analysis of a problem.  • Application of a reasonable strategy in the context of the problem is indicated.  • Explanation of and/or justification for the mathematical process(es) used to solve a problem is feasible, but may be only partially developed.  • Connections and/or extensions made within mathematics or outside of mathematics are partial or overly general, or may be implied.  • Supportive information and/or numbers are provided as appropriate. |
| 1 point | The response demonstrates a minimal understanding and analysis of a problem.  • Partial application of a strategy in the context of the problem is indicated.  • Explanation of and/or justification for the mathematical process(es) used to solve a problem is logically flawed or missing.  •Connections and/or extensions made within mathematics or outside of mathematics are flawed or missing.  •Supportive information and/or numbers may or may not be provided as appropriate. |
| 0 points | The response is completely incorrect, irrelevant to the problem, or missing. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item #1 Sample Response for: *Text Analysis*** | | | | |
| **Dimension** | **Advanced**  **(4 points)** | **Proficient**  **(3 points)** | **Basic**  **(2 points)** | **Below Basic**  **(1 point)** |
| **Response to Prompt**  **(4 points)** | Central claim is clearly articulated and completely addresses the writing prompt | Central claim is clearly articulated and mostly addresses the writing prompt | Central claim is ambiguous/unclear and partially addresses the writing prompt | An effort was made to create a central claim, but it does not address the writing prompt |
| **Textual Evidence**  **(4 points)** | Student uses sufficient (i.e., at least one instance per body paragraph) and relevant textual evidence to support the claim | Student uses relevant evidence to support the claim, but does not address all parts of the prompt with supporting textual evidence (i.e., one or more body paragraphs are missing textual evidence) | Student attempts to use sufficient (i.e., at least one instance per body paragraph) evidence to support the claim, but the evidence presented is irrelevant to the claim | Student attempts to use insufficient (i.e., one or more body paragraphs are missing textual evidence) evidence to support the claim, and the evidence presented is irrelevant to the claim |
| **Structure**  **(4 points)** | Response contains a clear beginning, middle, and end, and uses transitions to ensure these parts flow together and stay on-topic; textual evidence is integrated smoothly and its relevance is justified in the response | Student uses relevant evidence to support the claim, but does not address all parts of the prompt with supporting textual evidence (i.e., one or more body paragraphs are missing textual evidence) | Student attempts to use sufficient (i.e., at least one instance per body paragraph) evidence to support the claim, but the evidence presented is irrelevant to the claim | Student attempts to use insufficient (i.e., one or more body paragraphs are missing textual evidence) evidence to support the claim, and the evidence presented is irrelevant to the claim |
| **Conventions**  **(4 points)** | Student’s language is clear and appropriate for the task; there are up to two minor errors in spelling, grammar, and/or usage | Student’s language is clear and appropriate for the task; there is a major error and other minor errors in spelling, grammar, and/or usage | Student’s language is either clear OR appropriate for the task; there are up to three major errors in spelling, grammar, and/or usage | Student’s language is neither clear nor appropriate; there are up to five major errors in spelling, grammar, and/or usage |