

## *The Process for...*

# *Assuring the quality of “Teacher-Developed” Student Performance Measures*

## **Introduction**

The purpose of this document is to provide guidance for educators when developing measures of student performance and for applying the three (3) strands outlined in the *Performance Measure Rubric for Teachers*. The rubric is designed as a self-assessment tool used in ascertaining the quality of the developed performance measure created by educators for local use. The process used to design, build, and review teacher-made performance measures is contained within the Assessment Literacy Series (ALS). The ALS provides a professional development process easily accessible through the *Homeroom* online learning platform.

## **Teacher-made Performance Measures: Rubric Strands**

- Strand 1: Design
- Strand 2: Build
- Strand 3: Review

## **Doing this Work**

- 1) Convene a group of professional educators who are familiar with the content and grade level to be assessed (e.g., Professional Learning Communities). A lead facilitator should be designated and agreed upon with the group.
- 2) The participants should decide how consensus will be reached (100% agreement, majority, etc.).
- 3) Develop a plan and timeline for completing the development and review of assessment measures.

## **STRAND 1: DESIGN**

Task ID	Descriptor	Rating	Evidence
1.1	The purpose of the performance measure is explicitly stated (who, what, why).		
1.2	The performance measure has targeted content standards representing a range of knowledge and skills students are expected to know and demonstrate.		
1.3	The performance measure's design is appropriate for the intended audience and reflects challenging material needed to develop higher-order thinking skills.		
1.4	Specification tables articulate the number of items/tasks; item/task types, passage readability, and other information about the performance measure – <b>OR</b> – Blueprints are used to align items/tasks to targeted content standards.		
1.5	Items/tasks are rigorous (designed to measure a range of cognitive demands/higher-order thinking skills at developmentally appropriate levels) and of sufficient quantities to measure the depth and breadth of the targeted content standards.		
	<i>Strand 1 Summary</i>	<i>out of 5</i>	

### **Guiding Questions: *Designing Phase***

- What is the performance measure intended to measure and at what grade?
- What are the developmental characteristics of the test-takers?
- Which content areas will be focused on among the various content standards?
- How will the results (overall score and “growth” inferences) be used?
- When will the performance measure be administered?
- Do the items/tasks capture the purpose of the standard?
- Are there enough items/tasks so that students at all levels can demonstrate their knowledge?
- What are the time constraints for both educators and students?
- Do the standards and items/tasks reflect the design framework?

### **Materials Needed: *Designing Phase***

- State Content Standards, National Standards, Common Core Standards
- Teacher-made items/tasks/projects
- Textbooks and other ancillary materials
- ALS Handout #1 – Purpose Statement Example

- ALS Handout #3 – Test Specifications and Blueprints
- ALS Handout #4 – Depth of Knowledge (DoK) chart

### **Goal for Strand 1: *Designing Phase***

- Develop a “purpose statement” that states the *why*, *what*, and *how* aspects of the measure being developed.
- Review standards and identify the initial set of standards for the measure. This is “*what*” is going to be measured. This focuses the performance measure to a specific set of standards.
- Determine “*how*” the standard will be measured. Determine design features such as item type, depth of knowledge, timeframe for administration, and use of test scores.
- Complete specification table **AND/OR** blueprint to identify:
  - Content standards/description of content standards
  - Item types
    - Multiple Choice (MC)
    - Short Constructed Response (SCR)
    - Extended Constructed Response (ECR)
  - Item weights – the point value given to each item
  - Depth of Knowledge (DoK) distributions
- Specification tables and/or blueprints must be completed for both the pre-test and post-test performance measure, when using this test design.

## **STRAND 2: BUILD**

Task ID	Descriptor	Rating	Evidence
2.1	Items/tasks and score keys are developed using standardized procedures, including scoring rubrics for human-scored, open-ended questions (e.g., short constructed response, writing prompts, performance tasks, etc.).		
2.2	Items/tasks are created and reviewed in terms of: (a) alignment to the targeted content standards, (b) content accuracy, (c) developmentally appropriateness, (d) cognitive demand, and (e) bias, sensitivity, and fairness.		
2.3	Administrative guidelines are developed that contain the step-by-step procedures used to administer the performance measure in a consistent manner, including scripts to orally communicate directions to students, day and time constraints, and allowable accommodations/adaptations.		
2.4	Scoring guidelines are developed for human-scored items/tasks to promote score consistency across items/tasks and among different scorers. These guidelines articulate point values for each item/task used to combine results into an overall score.		
2.5	Summary scores are reported using both raw score points and performance level. Performance levels reflect the range of scores possible on the assessment and use terms or symbols to denote performance levels.		
2.6	The total time to administer the performance measure is developmentally appropriate for the test-taker. Generally, this is 30 minutes or less for young students and up to 60 minutes per session for older students (high school).		
	<i>Strand2 Summary</i>	<i>__ out of 6</i>	

### **Guiding Questions: *Building Phase***

- Are the items aligned with targeted content standards?
- Do the selected items/tasks allow students to demonstrate content knowledge by:
  - Responding to questions/prompts
  - Performing tasks, actions, demonstrations
- Are the items/tasks measuring the content knowledge, skill, or process and not some external or environmental factor (e.g., guessing)?
- Are there sufficient items/tasks to sample the targeted content?
- Are the items/tasks developmentally appropriate for the intended test-takers?
- Are the correct answers and/or expected responses clearly identified?

### **Guiding Questions: *Building Phase (cont.)***

- Do the performance measures' directions specify:
  - What the test-taker is supposed to do, read, analyze, etc.?
  - Where and how the test-taker is supposed to respond, demonstrate the task, etc.?
  - How many points a correct/complete response is worth towards the overall score?
- Are there directions for different item/task types?

### **Materials Needed: *Building Phase***

- Teacher-made items/tasks/projects
- Textbooks and other ancillary materials
- Formative assessment materials, curriculum-based measures
- ALS Handout #5 – Item Examples
- ALS Handout #6 – Scoring Key Example
- ALS Handout #7 – Rubric Examples
- ALS Handout #8 – Operational Form Example
- ALS Handout #9 – Administrative Guidelines Example

### **Goal for Strand 2: *Building Phase***

- Review specification tables and/or blueprints and targeted standards.
- Select and/or write the item/task passage.
- Select and/or write the response options.
  - Multiple choice (MC)
  - Short Constructed Response (SCR)
  - Extended Constructed Response (ECR)
- Specify DoKs.
- Develop test guidelines for test-takers.
- Develop test guidelines for educators.
- Develop scoring sheets (matrices) and scoring rubrics.

### **STRAND 3: REVIEW**

Task ID	Descriptor	Rating	Evidence
3.1	<p>The performance measures are reviewed in terms of design fidelity –</p> <ul style="list-style-type: none"> <li>Items/tasks are distributed based upon the design properties found within the specification or blueprint documents.</li> <li>Item/task and form statistics are used to examine levels of difficulty, complexity, distracter quality, and other properties.</li> <li>Items/tasks and forms are rigorous and free of bias, sensitive, or unfair characteristics.</li> </ul>		
3.2	<p>The performance measures are reviewed in terms of editorial soundness, while ensuring consistency and accuracy of other documents (e.g., administration) –</p> <ul style="list-style-type: none"> <li>Identifies words, text, reading passages, and/or graphics that require copyright permission or acknowledgements</li> <li>Applies Universal Design principles</li> <li>Ensures linguistic demands and/or readability is developmentally appropriate</li> </ul>		
3.3	<p>The performance measures are reviewed in terms of alignment characteristics –</p> <ul style="list-style-type: none"> <li>Pattern consistency (within specifications and/or blueprints)</li> <li>Matching the targeted content standards</li> <li>Cognitive demand</li> <li>Developmental appropriateness</li> </ul>		
3.4	Cut scores are established for each performance level. Performance level descriptors describe the achievement continuum using content-based competencies for each assessed content area.		
3.5	As part of the assessment cycle, post administration analyses are conducted to examine such aspects as items/tasks performance, scale functioning, overall score distribution, rater drift, content alignment, etc.		
3.6	The performance measure has score validity evidence that demonstrated item responses were consistent with content specifications. Data suggest the scores represent the intended construct by using an adequate sample of items/tasks within the targeted content standards. Other sources of validity evidence such as the interrelationship of items/tasks and alignment characteristics of the performance measure are collected.		
3.7	Reliability coefficients are reported for the performance measure, which includes estimating internal consistency. Standard errors are reported for summary scores. When applicable, other reliability statistics such as classification accuracy, rater reliabilities, and others are calculated and reviewed.		
<i>Strand 3 Summary</i>		<i>__ out of 7</i>	

### **Guiding Questions: *Reviewing Phase***

- Does the question clearly address the standard?
- Is the reading difficulty and vocabulary appropriate?
- Is the language clear, consistent, and understandable?
- Are charts, tables, graphs, and diagrams clear and understandable?
- Is there only one (1) correct answer?
- Have the items been reviewed for bias and sensitivity?
  - Items provide an equal opportunity for all students to demonstrate their knowledge and skills. The question and any stimulus material (e.g., reading passage, artwork, and diagram) do not raise bias and/or sensitivity concerns that would interfere with the performance of a particular group of students.
- Are the items developmentally appropriate for test-takers?
- Do the specifications and/or blueprints reflect the test forms?
- Do the scoring rubrics provide detailed scoring information?

### **Materials Needed: *Reviewing Phase***

- Specifications and/or blueprints
- Draft operational form(s)
- Draft scoring sheet and rubric(s)
- Handout #10 – Quality Assurance Checklist

### **Goal for Strand 3: *Reviewing Phase***

- Review of all required documents to include:
  - Purpose statement
  - Specifications **OR/AND** blueprint
  - Operational form(s)
  - Scoring sheets
  - Administrative guidelines