



PERFORMANCE MEASURE RUBRIC FOR TEACHERS

General Purpose

The enclosed rubric is designed to examine the quality characteristics of teacher-made performance measures. The rubric is comprised of 18 technical aspects organized into three (3) strands. The rubric's purpose is to provide teachers with a self-assessment tool that assists in building "great" measures of student achievement.

Rating Tasks

- Step 1.** Review information, data, and documents associated with the design, development, and review of the selected performance measure.
- Step 2.** Assign a value in the "Rating" column for each aspect within a particular strand using the following scale:
- a. (1) = fully addressed
 - b. (.5) = partially addressed
 - c. (0) = not addressed
 - d. (N/A) = not applicable at this time
- Step 3.** Reference supporting information associated with each assigned rating in the "Evidence" column.
- Step 4.** Add any additional notations and/or comments that articulate any important nuances of the performance measure.
- Step 5.** Compile assigned values and place in the "Strand Summary" row.

Summary Matrix

Strand	Points Possible	Points Earned
Design	5	
Build	6	
Review	7	
<i>Summary</i>	<i>18</i>	

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 - OR - 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.		
	Strand 1 Summary	— out of 5	
Additional Comments/Notes			

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	Item/tasks are created and reviewed in terms of: (a) alignment to the targeted content standards, (b) content accuracy, (c) developmental 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>Strand 2 Summary</i>	<i>— out of 6</i>	
Additional Comments/Notes			

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"> • Items/tasks are distributed based upon the design properties found within the specification or blueprint documents. • Item/task and form statistics are used to examine levels of difficulty, complexity, distracter quality, and other properties. • Items/tasks and forms are rigorous and free of bias, sensitive, or unfair characteristics. 		
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"> • Identifies words, text, reading passages, and/or graphics that require copyright permission or acknowledgements • Applies Universal Design principles • Ensures linguistic demands and/or readability is developmentally appropriate 		
3.3	<p>The performance measures are reviewed in terms of alignment characteristics-</p> <ul style="list-style-type: none"> • Pattern consistency (within specifications and/or blueprints) • Matching the targeted content standards • Cognitive demand • Developmental appropriateness 		
3.4	<p>Cut scores are established for each performance level. Performance level descriptors describe the achievement continuum using content-based competencies for each assessed content area.</p>		
3.5	<p>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.</p>		

Task ID	Descriptor	Rating	Evidence
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>	
Additional Comments/Notes			