Decision Guidance and Interrater Consistency

The scoring rubric is designed to support consistent evaluation of optional measure submissions. Final approval decisions should be informed by rubric scores, but should not be based on score totals alone. Reviewers should use professional judgment in combination with rubric results to determine the most appropriate decision. Score ranges should be used to support consistency across reviewers.

Use of Score Ranges

Total scores provide general guidance for decision-making:

- Higher scores typically indicate that a measure is well-developed and suitable for use with minimal or no revisions
- Moderate scores may indicate that a measure is generally appropriate but requires clarification, refinement, or additional evidence
- Lower scores may indicate that significant revisions are needed before the measure can be considered for approval

Critical Criteria Considerations

Certain criteria are foundational to approval and should be considered carefully regardless of total score. If a measure demonstrates significant concerns in any of these areas, reviewers should consider whether approval is appropriate, even if the total score is relatively high.

- Technical Fit with the AEC SPF
  Measures must align to a defined indicator, produce quantitative data, and integrate into the Data Pipeline. Measures that do not meet these requirements are not appropriate for approval without substantial revision.
- Student Outcome Integrity
  Measures must reflect meaningful student outcomes that are attributable to the AEC. Measures that rely on prior or external data, or that do not clearly represent student performance, may not be appropriate for approval.
- Rigor and Validity
  Measures must demonstrate sufficient technical quality, including defined cut points, clear methodology, and evidence of validity and reliability.
- Sufficient Variability
  Measures must be designed to produce meaningful differentiation in performance across students. Measures that do not support all performance levels (including ‘does not meet’), or rely on cut points that do not differentiate performance may not be appropriate for approval.

Consistency in Reviewer Judgment

To support consistency across reviewers:

- Decisions should be grounded in evidence provided in the submission
- Reviewer justifications should clearly reference specific elements of the submission and rubric criteria
- Similar types of measures should be evaluated using consistent expectations across submissions
- When uncertainty exists, reviewers should prioritize clarity, defensibility, and alignment to the guiding principles

Use of Approval Categories

Final determinations should reflect both the overall strength of the measure and any required revisions:

- Approved - The measure meets expectations and can be used within the AEC SPF
- Approved with Conditions - The measure is appropriate for use but requires specific revisions, clarifications, or additional documentation
- Revise and Resubmit - The measure shows potential but requires substantial changes before it can be considered for approval
- Not Approved - The measure does not meet the necessary criteria for use within the AEC SPF

Rubric for New Optional Measures

Criteria	Description	Exemplary (4)	Proficient (3)	Basic (2)	Needs Improvement (1)
Technical Fit	The measure aligns well with the established AEC SPF and integrates seamlessly within the accountability system.	Clearly aligned to a specific AEC SPF indicator	Aligned to an AEC SPF indicator with minor clarification needed	Partial or unclear alignment to an AEC SPF indicator	Does not align to an AEC SPF indicator
		Produces quantitative data that is fully compatible with Data Pipeline	Quantitative and likely compatible with Data Pipeline	Quantitative data present but unclear submission feasibility	Not quantitative OR cannot be submitted through Data Pipeline
		Demonstrates clear, non-duplicative value (not already captured by required or approved measures)	Adds generally new information, but some overlap may exist	Potential duplication with existing measures	Clearly duplicates a state-required measures OR an existing optional measure
		Data source is clearly defined and not reused in conflicting ways	Data source is defined but could be clearer	Data source or usage is unclear or inconsistently described	Data source unclear or inappropriate
		Submission shows strong understanding of framework integration	No major duplication issues	May require moderate revisions to fit framework	Cannot be integrated into framework without major redesign
Use of Stude	The measure relies on reliable student outcome or performance	Clearly measures a defined student outcome	Measures a student outcome, though definition may need minor clarification	Student outcome is unclear, indirect, or partially defined	Does not measure a student outcome (e.g., input, perception, activity)

Criteria	Description	Exemplary (4)	Proficient (3)	Basic (2)	Needs Improvement (1)
	data to assess effectiveness and student success.	Outcomes are fully attributable to the AEC; Data is entirely from students enrolled at the AEC with no reliance on prior school or external data	Mostly attributable to AEC (minor ambiguity possible)	Attribution to AEC is weak or partially unsupported	Not attributable to AEC (uses prior or external data)
		Includes a strong majority of students (e.g., ≥75% of students eligible for the measure)	Includes a majority of students, but not fully representative	Limited population coverage OR unclear rationale	Covers a small or non-representative subset of students
		Participation rate is high (≥75%) OR there is a clear plan to achieve ≥75% participation and is structurally embedded (e.g., required administration), feasible given schedule and population, and barriers identified and mitigated	Reasonable plan for strong participation; Some details missing, but generally feasible	Participation plan is unclear or weak; Risks to achieving sufficient participation	No clear plan for participation; Likely low participation; Significant barriers not addressed
		Population is clearly defined and representative	Minor gaps in population clarity or inclusion	Some inclusion of data outside AEC context possible	Data source undermines validity of outcome
Rigor and	The measure demonstrates the necessary rigor and	Measure is state or nationally normed, or strongly justified if locally normed	Norming approach is defined but may require clarification	Norming unclear or weakly justified	No clear norming approach

Criteria	Description	Exemplary (4)	Proficient (3)	Basic (2)	Needs Improvement (1)
	validity to be used for accountability purposes.	Academic measures are clearly aligned to Colorado Academic Standards (if applicable)	Standards alignment is present but not fully detailed	Limited or partial standards alignment	No standards alignment (if required)
		Full technical specifications provided including calculation method, administration process, and scoring	Technical specifications are mostly complete	Technical specifications incomplete	Missing or unclear technical specifications
		Strong evidence of validity either external validation or well-supported internal validation	Validity evidence is present but limited	Minimal or unclear validity evidence	No evidence of validity
		Reliability is clearly addressed through standardized scoring or inter-rater reliability processes	Reliability is addressed but not fully developed	Reliability not clearly addressed	No reliability safeguards
		Cut point methodology is clearly defined and well-justified, and rationale for thresholds is supported by data or established benchmarks	Cut point methodology is defined but rationale could be stronger	Cut point methodology present but weak or poorly justified	Cut point methodology not defined or not usable
Sufficient Varia	The measure captures enough	Clear evidence or strong expectation of variation in student performance	Adequate variability demonstrated	Limited variability in data	Little to no variability in results

Criteria	Description	Exemplary (4)	Proficient (3)	Basic (2)	Needs Improvement (1)
	variability in performance to justify use in a system that compares performance across different individuals or groups.	Students and schools can perform at all levels, including "does not meet"			Not possible for students/schools to fail
		All four performance categories are defined and expected to be represented in data	Most performance levels represented	Some performance levels rarely represented	One or more performance levels missing
		Expected or observed distribution demonstrates meaningful spread across categories	Expected or observed distribution shows some differentiation	Distribution is clustered or uneven	Distribution is flat or uniform
		Cut points create clear differentiation	Cut point differentiation generally functional but may need refinement	Cut points do not clearly differentiate performance	Cut points do not support meaningful differentiation in performance
		No evidence of ceiling effects or artificial inflation	Minor concerns about clustering or spread	Risk of inflated or compressed results	
		Measure supports cross-school comparison and/or longitudinal comparison			Measure not appropriate for normative accountability