

# Appendices

## Appendix A: CAEP Criteria for Evaluation of EPP-Created Assessments & Surveys

### CAEP Criteria for Evaluation of EPP-Created Assessments

#### 1. Administration and Purpose

##### Sufficiency Criteria

- The time/point at which the assessment is administered during the preparation program is explicit.
- The purpose of the assessment and its use in candidate monitoring or decisions on progression are specified and appropriate.
- Instructions provided to candidates about what they are expected to do are informative and unambiguous.
- The basis for judgment is made explicit for candidates.
- Evaluation categories or assessment tasks are aligned with CAEP, InTASC, national/professional, and state standards

#### 2. Content of Assessment

##### Sufficiency Criteria

- Indicators assess explicitly identified aspects of CAEP and InTASC Standards, in addition to national, professional, or state standards.
- Indicators reflect the degree of difficulty or level of effort described in the standards.
- Indicators unambiguously describe the proficiencies to be evaluated.
- When the standards being informed address higher level functioning, the indicators require higher levels of intellectual behavior (e.g., create, evaluate, analyze, and apply). For example, when a standard specifies that candidates' students "demonstrate" problem solving, then the indicator is specific to candidates' application of knowledge to solve problems.
- Most indicators require observers to judge consequential attributes of candidate proficiencies in the standards.

### 3. Scoring

#### Sufficiency Criteria

- The basis for judging candidate performance is well defined.
- Each proficiency level descriptor (PLD) is qualitatively defined by specific criteria aligned with indicators.
- PLDs represent a developmental sequence from level to level (providing raters with explicit guidelines to evaluate candidate performance and giving candidates explicit feedback on their performance).
- Feedback provided to candidates is actionable—it is directly related to the preparation program and can be used for program improvement as well as for feedback to the candidate.
- Proficiency level attributes are defined in actionable, performance-based, or observable behavior terms. [NOTE: If a less actionable term is used such as “engaged,” criteria are provided to define the use of the term in the context of the category or indicator.]

### 4. Data Reliability

#### Sufficiency Criteria

- A description or plan is provided that details the type of reliability that is being investigated or has been established (e.g., inter-rater, internal consistency, consensus building activities with documentation) and the steps the EPP took to ensure the reliability of the data from the assessment.
- Training of scorers and checking on inter-rater agreement and reliability are documented.
- The described steps meet accepted research standards for establishing reliability.

## 5. Data Validity

### Sufficiency Criteria

- A description or plan is provided that details steps the EPP has taken or is taking to ensure the validity of the assessment and its use.
- The plan details the types of validity that are under investigation or have been established (e.g., construct, content, concurrent, predictive) and how they were established.
- If the assessment is new or revised, a pilot was conducted.
- The EPP details its current process or plans for analyzing and interpreting results from the assessment.
- The described steps meet accepted research standards for establishing the validity of data from an assessment.

### Notes