

GaPSC Guidelines for Key Assessments for Initial Programs

The Georgia Professional Standards Commission (GaPSC) uses the term “key assessment” to include content tests, observations, projects or assignments, and surveys. Each EPP must provide five key assessments for each program (with an optional sixth key assessment), and in PRS-II there are seven prompts to which EPPs must respond for each of those key assessments:

- A copy of the assessment.
- The instructions to candidates about the assigned task.
- The purpose of the assessment.
- Data tables showing results of this assessment for the last three cycles, an analysis of the findings, the interpretation of those findings, and changes made in (or planned for) the program as a result. (This item does not apply to developmental programs.)
- Statement of how the program assures that validity and reliability of the assessment have been met. (4 of 5 key assessments at the program level for teaching fields)
- A brief description of how the assessment is used in this program, including how often and when it is administered, who administers it, and how this assessment is used to make decisions about candidates' progress through the program.
- Scoring guides or criteria used to score candidates' responses to the task, including the required passing score. Also include what is expected when a candidate does not meet the required passing score.

This document provides additional context for these prompts and will assist EPPs and Site Visit Teams in considering key assessment evidence. **EPPs should use the guidelines in this document as they design, pilot, and judge the adequacy of the assessments they create and utilize. During approval reviews, Site Teams will also use these guidelines as they examine the key assessments and resulting data.**

GENERAL GUIDING PRINCIPLES FOR EXAMINING KEY ASSESSMENTS

- Key assessments and scoring guides are used by faculty to evaluate candidates and provide them with feedback on their performance. (Key assessments are not only assessments OF learning – they are also assessment FOR learning.)
- Key assessments and scoring guides should address relevant and meaningful attributes of candidate knowledge, performance, and dispositions, and they must be aligned with InTASC standards and/or program standards.
- Key assessments are those that all candidates are expected to complete as they transition from admission through completion, or that are used to monitor their progress in developing proficiencies during one or more stages of preparation.
- Surveys may be used to gather evidence on aspects of candidate preparation and candidate perceptions about the quality of their preparation and their own readiness to teach. Surveys can be strong key assessments when they are clearly aligned to InTASC standards and/or program standards and are executed with best practice protocols.

Note that the GaPSC administers the Inductee Survey and the Employer Survey one year after completion. Although resulting data might be used as evidence for meeting Standard 4 and as supporting evidence of proficiency on the InTASC Standards, as Inductees are completers and not candidates, data from these surveys should not be used as key assessment evidence in Standard 1.

ADMINISTRATION AND PURPOSE (*informs relevancy*)

- The point(s) at which the assessment is administered during the preparation program are explicit.
- The **purpose** of the assessment and its use in candidate monitoring or consequential decisions regarding progression are specified and appropriate.
- Instructions provided to candidates about what they are expected to do are informative and unambiguous.
- The basis for judgment (criterion for success, or what is “good enough”) is made explicit for candidates.
- Evaluation categories or assessment tasks are aligned with Georgia Standards for the Approval of Educator Preparation Providers and Educator Preparation Programs, InTASC standards, and/or program standards.
- Candidate progression is monitored and information is used for mentoring and/or remediation.
- Candidates are informed about how the instrument results are used in reaching conclusions about their status and/or progression.

CONTENT OF ASSESSMENT (*informs relevancy*)

- Indicators assess explicitly identified aspects of the Georgia Standards for the Approval of Educator Preparation Providers and Educator Preparation Programs, InTASC standards, and/or program standards.
- Indicators reflect the degree of difficulty or level of effort described in the standards.
- Indicators clearly describe the proficiencies being evaluated.
- When the standards being assessed set an expectation for higher level functioning, the indicators require higher levels of intellectual behavior (e.g., create, evaluate, analyze, and apply).
- Most indicators (at least those comprising 80% of the total score) require observers to judge consequential attributes of candidate proficiencies described in the standards. In other words, only 20% of the indicators include expectations such as, “Timely Submission,” “Grammar/Writing,” “Document Formatting.”
 - The word *indicators* is used as a generic term for assessment items.
 - For content tests, *indicators* refers to a question.
 - For projects or assignments, *indicators* refers to a prompt or task that the candidate is to perform.
 - For an observation, an *indicator* might be a category of performance to observe or a specific aspect of candidate performance that a reviewer will record.
 - For a survey, an *indicator* is represented by a question or statement for which a response is to be selected from multiple options.

SCORING (*informs reliability and actionability*)

- The basis for judging candidate performance is well defined.
- Each proficiency level descriptor is qualitatively defined by specific criteria aligned with indicators.

- Proficiency level descriptors represent a developmental sequence from level to level (to provide raters with explicit guidelines for evaluating candidate performance and for providing candidates with explicit feedback on their performance).
- Feedback provided to candidates is actionable—it is directly related to the preparation program and can be ***used for both feedback to the candidate and program improvement.***
- Proficiency level attributes are defined in actionable, performance-based, or observable behavior terms.
 - If a less actionable term (e.g., “engaged”) is used, criteria are provided to define the use of the term in the context of the category or indicator.

DATA RELIABILITY

- A detailed description of the type of reliability that has been established (e.g., test-retest, parallel forms, inter-rater, internal consistency), and the process the EPP took to ensure the reliability of the resulting data.
- Process used to establish reliability should be a widely accepted practice.
- Based on the type of reliability and process used, a reliability coefficient (statistic) is reported with a description of what the EPP defines as an acceptable level of reliability.
- Procedures used for training scorers and checking on inter-rater agreement and reliability are documented.

SURVEYS

- Questions or topics are explicitly aligned with aspects of the EPP’s mission and Georgia Standards for the Approval of Educator Preparation Providers and Educator Preparation Programs, InTASC standards, and/or program standards.
- Individual items have a single subject; language is unambiguous.
- Leading questions are not used.
- Items are stated in terms of behaviors or practices, rather than opinions.
- Scoring is anchored in performance/behavior relative to practice.
- Surveys of dispositions make clear to candidates how the survey is related to their area of preparation (teaching field, service field, leadership).

SURVEY DATA QUALITY

- Scaled choices are qualitatively defined using specific criteria aligned with key attributes.
- Feedback provided to the EPP is actionable, with a minimum response rate of 20%.
- EPP provides evidence that questions were piloted to determine that candidates interpret them as intended, and necessary modifications were made.
- EPP provides evidence of survey construct validity.

NOTE: During an Approval Review

- EPP provides evidence that assessment data are compiled and tabulated accurately.
- Data are disaggregated by program, where appropriate (e.g., Standard 1, Component 3.2, Component 3.3, Standard 4, and Standard 5).
- The analysis and interpretation of the data are sufficient to inform EPP/program improvements.
- Interpretations of assessment results are appropriate for the items and resulting data.

- Any interruptions in the collection of data are thoroughly described.
- Results from successive administrations are compared (for evidence of reliability).

Examples of Key Assessment Attributes *Below Sufficient Level*

<p>Administration and Purpose</p> <ul style="list-style-type: none"> ○ Use or purpose is ambiguous or vague. ○ There is limited or no evidence regarding information given to candidates. ○ Instructions given to candidates are incomplete or misleading. ○ The criterion for success is not provided or is not clear. <p>Content of Assessment</p> <ul style="list-style-type: none"> ○ Indicator alignment with Georgia Standards, InTASC, and/or program standards is incomplete, absent, or only vaguely related to the content of standards being evaluated. ○ Indicators fail to reflect the degree of difficulty described in the standard. ○ Indicators are not described, are ambiguous, or include only headings. ○ Higher-level performances, as represented in the standards, are not apparent in the indicators. ○ Description of established validity does not inform reviewers about the method used. ○ The type of validity established is misidentified or not described. <p>Scoring</p> <ul style="list-style-type: none"> ○ Many indicators (more than 20% of the total score) require judgment of candidate proficiencies that are of limited importance in Georgia Standards, InTASC, and/or program standards. ○ Rating scales are used instead of rubrics (e.g., “level 1= significantly below expectation,” “level 4 = significantly above expectation”). ○ Proficiency level descriptors do not align with indicators. ○ Proficiency level descriptors do not represent developmental progressions. ○ Proficiency level descriptors provide limited or no feedback to candidates specific to their performance. ○ Proficiency level descriptors are vague or not defined and may simply repeat the language from the standards. 	<p>Reliability</p> <ul style="list-style-type: none"> ○ Description of established reliability does not inform reviewers about the method used. ○ Described steps of establishing reliability do not meet accepted standards. ○ Limited evidence is provided that scorers are trained, and their inter-rater agreement was documented. <p>Survey Data Quality</p> <ul style="list-style-type: none"> ○ There is no evidence that the items were piloted before administration. ○ Described steps do not meet accepted research standards for establishing validity. For example, validity is determined through an internal review by only one or two stakeholders. ○ Questions or topics are not aligned with EPP mission or standards. ○ Individual items are ambiguous or include more than one subject. ○ There are numerous leading questions. ○ Items are stated as opinions rather than as behaviors or practices. ○ Dispositions surveys used as key assessments do not demonstrative a relationship to effective teaching. ○ Scaled choices are numbers only, without qualitative descriptions linked with the item under investigation. <p>FOR ALL ASSESSMENTS</p> <ul style="list-style-type: none"> ○ Data analysis process and the interpretation of data are not presented or are superficial. ○ Limited or no feedback provided to the EPP for improvement purposes.
--	---

Adapted from the [CAEP Evaluation Framework for EPP-Created Assessments](#)