COE Assessment and Program-Reporting Help-Sheet: Possible Questions and Issues to Consider

Suggestion: These questions may be of use in guiding faculty as they review and evaluate their program key assessment data and summaries. Addressing these three categories of reliability, validity, and fairness will help document the credibility of our assessments as tools are developed and used. Documentation within program reports and committee minutes is advised.

Reliability: consistent over repeated applications; dependable

- Are the instructors/raters consistent in how they score performances using the rubric or other instrument? (Which data/summary tells us this?)

- Are performance scores similar if/when a second faculty member scores a sample student's work product? (How is this ensured?)

- Do other available data support the reliability/consistency of program assessments? (Which data?)

Program report writers could also describe how training and reliability scoring is conducted or how carefully constructed proficiency level descriptions were developed or revised, including how these activities might affect assessment reliability.

Validity: degree to which accumulated evidence and theory support proposed interpretations

- Alignment: Do the performance tasks and scoring instruments adequately reflect the standards defined for the program? (How do we know this?)

- Content: Are the key assessment performance tasks authentic representations of necessary job knowledge or skills? (How do we know this?)

- Instrument Use: Are the three performance levels meaningfully defined and used so as to yield accurate score differences among target, acceptable, and unacceptable performances for each skill, or element? (How do we know this?)
  If everyone receives scores of 3, for example, is everyone truly at target level? Does the instrument need and receive some revision to better specify levels and discriminate between target and acceptable performance? Or, is additional rater training needed and implemented?

Program report writers could also describe how internal and external reviews were used with assessments, instrument development, revision, and scoring.

Fairness: students are assessed in an equitable way

- Do all candidates have an equal opportunity to demonstrate the assigned tasks? (How do you know?)

- Do procedures contribute to equity such that all candidates can know the requirements, have access to resources, have the opportunity to learn content, and can remediate when necessary? (How?)

Program report writers could also describe the clarity and availability of instructions and rubrics, timely feedback, multiple assessments, and formative assessment opportunities provided.
Definitions

Reliability
1. The degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence inferred to be dependable, and repeatable for an individual test taker; the degree to which scores are free of errors of measurement for a given group (AERA, APA, NCME, 1999).

2. NCATE does not define reliability but instead uses the terminology “Consistency in Assessment,” defined as follows: The assurance that key assessments produce dependable results or results that would remain consistent on repeated trials. Institutions can document consistency through providing training for raters that promote similar scoring patterns, using multiple raters, conducting simple studies of inter-rater reliability, and/or comparing results to other internal or external assessments that measure comparable knowledge, skills, and/or professional dispositions (NCATE, 2008).

Validity
1. The degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of the test (AERA, APA, NCME, 1999).

2. NCATE does not define validity but instead uses the terminology “Accuracy in Assessment,” defined as follows: The assurance that key assessments are of the appropriate type and content such that they measure what they purport to measure. To this end, the assessments should be aligned with the standards and/or learning proficiencies that they are designed to measure (NCATE, 2008).

Fairness
1. In testing, the principle that every test taker should be assessed in an equitable way (AERA, APA, NCME, 1999).

2. The assurance that candidates have been exposed to the knowledge, skills and dispositions that are being evaluated in key assessments and understand what is expected of them to complete the assessments. To this end, instructions and timing of assessments should be clearly stated and shared with candidates. In addition, candidates should be given information on how the assessments are scored and how they count toward completion of programs (NCATE, 2008).

References
