Validity Evidence Needed for Rubric Use and Interpretation

(See page 2 for specific protocol.)

Validity includes gather evidence to demonstrate that the assessment content fairly and adequately represents a defined domain of knowledge or performance. The purpose of this paper is to provide guidance for collection of evidence to document adequate technical quality of rubrics that are being used to evaluate candidates in the College of Education at UNC Charlotte.

Definition of Validity

"Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests." (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014) Validation is the process of accumulating evidence that supports the inferences that are made of candidate responses for specified assessment uses. While there are several types of evidence commonly used to examine to support the validity of an assessment inferences and uses, this paper focuses on content-related validity.

Content-related evidence refers to the extent to which (a) a candidate's responses to a given assessment reflects that student's knowledge of the content area that is of interest and (b) concerned with the extent to which the assessment instrument adequately samples the content domain. Content-related evidence should also be considered when developing scoring rubrics. A well-designed scoring rubric cannot correct for a poorly designed assessment instrument, so selection of the tasks/activities should be examined to ensure it is aligned with the purpose and professional standards in the field of study.

The American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014) provided standards for considering content-oriented evidence.

> Standard 1.11 Content-Oriented Evidence
>
> "When the rationale for test score interpretation for a given use rest in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified." (p. 26)

Initial Plan for Collecting and Interpreting Evidence for Rubric Use

The following section provides the initial evidence needed to document the technical quality of rubric scores. It should be noted that this is the beginning of the evidence collection, and much more evidence is needed to document the quality of UNC Charlotte's candidates' assessments, including inter-rater reliability.

Establishing Content Validity for Internally-Developed Assessments/Rubrics

To establish content-validity for internally-developed assessments/rubrics, a panel of experts will be used. While there are some limitations of content validity studies using expert panels (e.g., bias), this approach is accepted by CAEP. As noted by Rubio, Berg-Weger, Tebb, Lee and Rauch (2003),

> Using a panel of experts provides constructive feedback about the quality of the measure and objective criteria with which to evaluate each item …. A content validity study can provide information on the representativeness and clarity of each item and a preliminary analysis of factorial validity. In addition, the expert panel offers concrete suggestions for improving the measure. (p. 95).

Protocol

1. **Complete the Initial Rubric Review (COED Assess FORM A)** for **each** rubric used to officially evaluate candidate performance in the program. Make sure that the "overarching constructs" measured in the assessment are identified (see #3-2 on FORM A). The form is completed via Google Drive link (a paper copy is available on the website for faculty to review).

2. **Identify a panel of experts and credentials for their selection.** The review panel should include a mixture of IHE Faculty content experts and B12 school/community practitioner experts. Minimal credentials for each expert should be established by consensus from program faculty; credentials should bear up to reasonable external scrutiny (Davis, 1992).

   The number of panel experts should include:
   a. At least 3 content experts from the program/department in the College of Education at UNC Charlotte;
   b. At least 1 external content expert from outside the program/department. This person could be from UNC Charlotte or from another IHE, as long as the requisite content expertise is established; and
   c. At least 3 practitioner experts from the field.
   TOTAL NUMBER OF EXPERTS: At least seven (7)

3. **Creating the response form.** For **each** internally-development assessment/rubric, there should be an accompanying response form that panel members are asked to use to rate items that appear on the rubric. Program faculty should work collaboratively to develop the response form needed for each rubric used in the program to officially evaluate candidate performance.
   a. For each item, the overarching construct that the item purports to measure should be identified and operationally defined.
   b. The item should be written as it appears on the assessment.
   c. Experts should rate the item's level of representativeness in measuring the aligned overarching construct on a scale of 1-4, with 4 being the most representative. Space should be provided for experts to comment on the item or suggest revisions.

d. Experts should rate the importance of the item in measure the aligned overarching construct, on a scale of 1-4, with 4 being the most essential. Space should be provided for experts to comment on the item or suggest revisions.
e. Experts should rate the item's level of clarity on a scale of 1-4, with 4 being the clearest. Space should be provided for experts to comment on the item or suggest revisions.

Helpful Notes:

- Faculty are welcome to make an electronic version of this tool that is customized to their specific rubrics to collect reviewer responses. The first row of information (see example) and the 4-level rating scale must be used for each item on each rubric. Survey Share or Survey Monkey or a Google Form could be adapted for this purpose.
- The COED Office of Assessment has scheduled rubric review sessions to discuss rubric development and/or content validity protocols. Faculty involved in this work are encouraged to attend. A list of the scheduled sessions can be found on the COED Assessment website.
- Copies of all this information may be found on the COED Assessment website.

(See example – faculty may cut and paste from the example to develop their response forms)

4. **Create an assessment packet for each member of the panel.** The packet should include:
   a. A letter explaining the purpose of the study, the reason the expert was selected, a description of the measure and its scoring, and an explanation of the response form. An example draft is included on the website (this is just a draft to get you started; faculty are welcome to develop their own letters). (see draft example)
   b. A copy of the assessment instructions provided to candidates.
   c. A copy of the rubric used to evaluate the assessment.
   d. The response form aligned with the assessment/rubric for the panel member to rate each item.

5. **Initiate the study.** Set a deadline for the panel to return the response forms to you / complete the response form online.

6. **Collecting the data**. Once response data for each internally-developed rubric have been collected from the panel participants, that information should be submitted to the COED Assessment Office. Copies of all forms and/or an excel file of submitted scores (if collected electronically) should be submitted in the designated file on the S: drive.

   Save expert responses in the following format: Rubric name (or shortened version)_Expert Last Name_Degree_Program
   (example: "STAR Rubric_Smith_BA_CHFD" "Present at State Read Conf_Smith_MEd_READ")

   To access the S: drive file to submit Content Validity Results, go to Computer → Shared Drive (S:) → coed → Shared → Assessment → Content Validity Results → select your department → select the program where the assessment is used. Multiple files may be added.

7. **Once Content Validity Results have been submitted**, the COED Assessment Office will generate a Content Validity Index (CVI). This index will be calculated based on recommendations by Rubio et. al. (2003), Davis (1992), and Lynn (1986):

$$\frac{\text{The number of experts who rated the item as 3 or 4}}{\text{The number of total experts}}$$

A CVI score of .80 or higher will be considered acceptable.

References:

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research, 5*, 194-197.

Lawshe, C. H. (1975). A qualitative approach to content validity. *Personnel Psychology, 28*, 563-575.

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 382-385.

Rubio, D.M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27(*2), 94-104.

DRAFT EXAMPLE - COPY AND PASTE WHATEVER PIECES ARE HELPFUL IN CREATING THIS FORM

Establishing Content Validity – Rubric/Assessment Response Form (note: creating an electronic version of this form via Google Drive or an online Survey tool is acceptable).

Name of Reviewer: _____ Position: _____

INSTRUCTIONS: This measure is designed to evaluate the content validity of _____(insert title of assessment)_____. Please rate each item as follows:

- Please rate the level of representativeness of item in measuring the aligned overarching construct on a scale of 1-4, with 4 being the most representative. Space is provided for you to comment on the item or suggest revisions.
- Please rate the importance of the item in measuring the aligned overarching construct on a scale of 1-4, with 4 being the most essential. Space is provided for you to comment on the item or suggest revisions.
- Please rate the level of clarity for each item on a scale of 1-4, with 4 being the most clear. Space is provided for you to comment on the item or suggest revisions.

This row **doesn't change** – same for all rubrics →

Overarching construct (i.e, "big idea to measure")	**Operational Definition**	**Item measuring overarching construct** (use the exact wording as appears on the assessment rubric).	**Representativeness of item in measuring the overarching construct** • 1 = item is not representative • 2 = item needs major revisions to be representative • 3 = item needs minor revisions to be representative • 4 = item is representative	**Importance of item in measuring the overarching construct** • 1 = item is not necessary to measure the construct • 2 = item is provides some information but is not essential to measure the construct • 3 = item is useful not but essential to measure the construct • 4 = item is essential to measure the construct	**Clarity of item** • 1 = item is not clear • 2 = item needs major revisions to be clear • 3 = item needs minor revisions to be clear • 4 = item is clear	**Comments:**
Construct 1: ___(fill in the blank) – *the construct "Content Knowledge" is used for this example.*						
Content Knowledge (Example)	Knowledge about actual subject matter that is to be learned or taught (Example)	K2a: Demonstrates knowledge of content (Example)	1 2 3 4	1 2 3 4	1 2 3 4	
Content	Knowledge about	K2b:Implements	1 2 3 4	1 2 3 4	1 2 3 4	

These rows will change depending on your program rubric … these rows will be different for each rubric used

Knowledge	actual subject matter that is to be learned or taught	interdisciplinary approaches and multiple perspectives for teaching content				
Content Knowledge	Knowledge about actual subject matter that is to be learned or taught	K2c: Demonstrates awareness of literacy instruction across all content areas	1 2 3 4	1 2 3 4	1 2 3 4	
Content Knowledge	Knowledge about actual subject matter that is to be learned or taught	K2d: Makes content relevant for all learners	1 2 3 4	1 2 3 4	1 2 3 4	

These three (3) open-ended response rows are inserted after each group of items aligned with an identified overarching competency

1 **To the reviewer**: What additional items would you recommend including to measure the construct? If you have no suggestions, please enter "none."

(this row would be inserted after each group of items aligned with an identified overarching construct).

2 **To the reviewer**: What additional items would you recommend deleting? If you have no suggestions, please enter "none."

(this row would be inserted after each group of items aligned with an identified overarching construct).

3 **To the reviewer**: Please provide any additional information you believe may be useful in assessing the identified construct with this instrument. If you have no suggestions, please enter "none."

(this row would be inserted after each group of items aligned with an identified overarching construct).

Construct 2: ___(fill in the blank) – *the construct "Learning Environments" is used for this example*.

Start with a new competency for next group of items

Learning Environments (Example)	The diverse physical locations, contexts, and	Etc – form would go on to list all items,	Etc	Etc	etc	etc

	cultures in which students learn. (Example)	etc.				