



Rationale

Rationale for the assessments

The rationale section of the *Inquiry Brief* establishes that the assessments the faculty uses in supporting each claim associated with *Quality Principle I* are reasonable and that there are plausible reasons for thinking the faculty members' interpretations of the assessment results are valid¹.

The program should describe its assessments in such a way that a reasonable person would conclude: *Yes, it makes sense that the measures selected are fitting, apt, and appropriate to test the claims. It makes sense, based on these measures, that the claims are (or could be expected to be) true.*

The rationale would show how the assessment of subject matter knowledge is connected to, for example, the grade point average in the completion of the major in the subject matter field, the individual course requirements, the grades given in the courses, the scores on standardized tests of the major field, pupil learning from the student teacher, an evaluation of a senior thesis, and the ratings of clinical supervisors.

The faculty members, in other words, explain why it is reasonable that they have chosen to support their claim that the teacher education candidates know their subject matter with such measures, for example, as grades in the major courses, scores on Praxis II, scores on the state curriculum tests, scores on the GRE subject matter test, grades on the senior thesis in the major, and cogency of the candidates' lesson plans in their subjects.

In sum, the rationale section explains why faculty members think it is reasonable to use the particular measures of candidate learning they have selected. Their reasons inevitably must indicate why they think the assessments will prove to be reliable and valid. Much of the rationale is implicitly revealed in the selection of the measures identified in Appendix E. The rationale section makes explicit the logic of the faculty's reasons for its choices.

Perhaps the most important feature of the rationale is that it gives the program's standard for its assessments and explains why the particular criterion the faculty believes indicates success is appropriate.

EXAMPLE: A rationale for the assessment of subject matter knowledge

The assessment (1) is tied to various program subject matter requirements, (2) has a basis and track record in the literature, (3) is empirically supported, (4) is practical and efficient, and (5) is otherwise a reasonable procedure for assessing subject matter knowledge.

In the rationale, the program faculty members give their reasons and their argument for using the measures they do. They address such questions as these:

1. *Why do we think this measure indicates subject matter knowledge?*
2. *How is this measure related to the teacher's competence to teach the subject matter?*
3. *How does the measure align with the program requirements?*
4. *Why would anyone think the measure has anything to do with subject matter knowledge?*
5. *What are the limitations of the measure and what are its strengths?*
6. *How did the faculty figure out what the criterion of success is for the assessment (the passing score)? How do they know those who show certain traits, skills, scores, and behaviors understand the subject matter while those who don't show these things do not meet the program's standard for subject matter understanding?*

¹ Empirical evidence of reliability and validity may be the major reason the faculty uses certain assessments and in that sense 1.5 and 2.1 may overlap considerably. However, there may be psychometrically sound assessments that could be dismissed in the rationale as weakly aligned with the claims.

Writing the rationale

The rationale is not simply a listing of the assessments (as presented in Appendix F) or an assertion that they measure the program's claims and goals, although it is partly that. It is an argument that gives the faculty's reasons for thinking its assessments are reliable and stable (usually because they have been used over time and the outcomes are consistent) and that they are valid (usually because the faculty sees that those who score highly on one assessment score highly on others and vice versa). The faculty members, if they are using an assessment for the first time and do not have a track record of experience with the assessment, may have some basis in the scholarly literature for thinking it will prove to be valid.

The rationale also provides the hypotheses that the faculty entertain in its inquiry into whether or not the assessments are valid: Why do faculty members think Praxis II scores and grades in the major should be related? Why do they think assessments of student teaching should be related to grades in the methods courses? Are the faculty supervisors or cooperating teachers more accurate in their assessment of the student's teaching? Can the pupils of student teachers assess the quality of the student teacher's teaching, etc?

The narrative of the rationale. The narrative might address such questions as these:

1. *Did the faculty measure what was covered in the program?*
2. *Did the faculty assess what the overall program was designed to produce?*
3. *Did the faculty's assessment procedures assure them and others that their graduates are competent, caring, and qualified?*

COMMENT

Why include a rationale? Many educators and other professionals have legitimate concerns about the reliability and validity of the evidence available in the field of education. To satisfy TEAC's *Quality Principle II*, the program faculty must have an ongoing investigation of the means by which it provides evidence for each component of *Quality Principle I*.

Quality Principle II, in fact, is partly about the need for this investigation. The investigation must accomplish two goals related to the assessment of candidate learning:

1. Support the choice of the assessments, particularly their links with the program's design, the program's goal, and the faculty claims made in support of the program goal.
2. Reduce the credibility of confounding factors associated with the evidence from which the faculty draws its inferences.

Finally, when faculty use the same assessment to support several claims, the rationale has to make clear which components of the assessment instrument support the several claims and that the faculty's interpretations of parts of the instrument are valid. The cooperating teaching rating form, which may be cited in support of each component of *Quality Principle I* may be a weaker indicator of subject matter knowledge than teaching skill and vice versa for course grades or license test results. The rationale would acknowledge these differences in the validity of the interpretations based on various components of the assessment instruments.

EXERCISES from *TEAC Exercise Workbook, 2010*
pages 59-63

Exercise 39: Draft your rationale. Using the space below, draft a short argument for one of your assessments for one of your claims that shows

1. Why the assessment was selected? (The answer to this question often entails showing how the assessment procedures reflect the features of the program, e.g., graduation requirements, admission criteria and procedures, coursework, field assignments, and experiences.)
2. What the passing score is for the assessment, how it was determined, and why does the faculty think it is appropriate?
3. What is the faculty's basis for thinking the assessment is reliable and that they can interpret the results validly?

You might produce the draft together with colleagues from your campus or write individual drafts and then compare and combine drafts.

Exercise 40: The program's standards: An expectation for the rationale section is to argue for the cut or passing scores that are in use in the program. When a measure is given, what level does the faculty take to be acceptable? Whatever the answer to that question, TEAC asks that the rationale address the question: What makes this faculty judgment credible?

Some examples of cut scores include: How high does the GPA have to be to graduate from the program? How high does the score on the state licensure test have to be to be judged by the faculty as "competent"? What levels of ratings from cooperating teachers are expected before a candidate is dropped from the program? What are the stated expectations for advancing to student teaching in the program's screening process?

a. List the current cut scores that apply to your program:

Assessment	Cut score
SAT	
GRE	
High school index	
License scores	
GPA	
Selected course standards	
Satisfaction survey ratings	
Student course evaluations	
Course	
Instructor	
Entry to student teaching	
Education GPA	
Subject Matter GPA	

b. Below are some proposed arguments for particular cut scores. Which arguments do you find compelling? Which ones do you find less than convincing? Mark those that are compelling with a C in front of its number. Mark those that are not convincing with an N. Place a question mark in front of those for which you are unsure.

C, N, or ?	Argument for cut scores
	Prior to entering the TEAC process, there were a number of cut scores in place. We intend to examine each of them as we develop habits of “institutional learning.” As of now, we have chosen to leave most of them in place until our inquiry can challenge them and suggest changes in them.
	We wrote to the programs ² that were honored by Carnegie in the Teachers for a New Era project to find out what cut scores they used in three particular areas – GPA to exit the program; GPA to enter student teaching; and satisfactory levels of ratings received from cooperating teachers on a five point scale. We adopted their cut scores.
	We hired a statistician who has worked with the National Board for Professional Teaching Standards (NBPTS) to help us empirically set cut scores. She worked for over a year with our files that included a complete data set for 500 of our graduates over a five year period. Her procedures yielded cut scores for GPA to exit, GPA to enter student teaching, and the lowest level satisfactory rating score received from cooperating teachers. We adopted our consultant’s recommendations.
	We couldn’t find in the research literature or in our own findings guidance for setting our cut scores. For this reason, we adopted the TEAC suggested 75% guideline. All of our cut scores represent the application of the 75% rule to our empirical maximum score.
	In a faculty meeting, we discussed the issue of cut scores and based the discussion on files of students who were near the current cut scores. After lengthy give and take among faculty members, we voted to approve the cut scores reported in the <i>Brief</i> .

Exercise 41: Figuring out the passing score (The Angoff method)

The faculty has used the following form of eight items to rate a student teacher’s proficiency with technology. The rating form allows the rater to score each item as “satisfactory” or “unsatisfactory.” Using this form, the highest possible score is 8. The faculty asks: *What should represent a cut score on this rating form to distinguish students who meet the department expectations from those who don’t meet the department expectations?* To determine the cut score, the faculty used the Angoff method (Livingston & Zieky, 2004). Here is the procedure:

1. Identify judges who know the students in the program and who are familiar with their practices with technology in the classroom. (For this exercise, you and your colleagues at the workshop will be judges).
2. Think about a borderline student in your program – one who falls near the cut between competent in technology and incompetent in technology. Discuss this student in some detail with your colleagues or in a “conversation with self” if you are the only person from the program in the workshop. Describe his practices in the classroom and his uses of technology in some depth.
3. Work to understand the following steps of the procedure: For each of the 8 items, decide how likely it would be for borderline students, such as the one you selected and described in step 2 above, to be rated satisfactory on this item. Score each item as 1 if the borderline student would know it and 0 if not. Determine the total score correct for each rater and take the mean of those scores as the cut-score (rounded up or down).
4. Practice the Angoff method on the scale we have provided. What cut score do you recommend?

Directions: Rate the borderline student 1 if he or she could do the item or 0 if he or she could not do the items for each of the 8 items of this scale.

1 or 0	Rating scale for the use of technology in the classroom
	1. The student teacher (intern) uses Power Point (or another presentation application) when presenting information to the class.
	2. The student teacher (intern) uses the Internet to connect the class with others doing similar work in schools across the nation and/or the world.

² www.teachersforanewera.com

	3. The student teacher (intern) uses the Internet to locate sources appropriate for the students to use.
	4. The student teacher (intern) finds software items that are useful for teaching understanding of difficult concepts.
	5. The student teacher (intern) coaches students to use computers for many purposes.
	6. The student teacher (intern) is careful to preview software for appropriateness and efficacy before it is introduced to the class.
	7. The student teacher (intern) shows ingenuity in selecting software for use in the classroom.
	8. The student teacher (intern) uses well software to help manage the class, record grades, register feedback, and/or to communicate to students and their parents.

Reference: Livingston, S.A., & Zieky, M. J. (2004). *Excerpts from passing scores*. Princeton, NJ: Educational Testing Service.

Another way to think about it, if there were 100 students who were judged borderline in technology, how many of the 100 would likely be rated satisfactory on each of the 8 items. When you are finished with this task, each item should have a number assigned to it from 0 to 100. Sum the numbers you have assigned to the 8 items, and divide by 100. The quotient should estimate the cut score on the rating scale. For example, if the estimates for the 8 items were as follows: 40, 20, 30, 20, 40, 10, 5, 10 – the sum is 175 and the quotient is 1.75. Rounding to the nearest integer, we have determined that the cut score should be 2. Any student receiving a 2 or lower on the scale should be deemed not to have met faculty expectations.

Exercise 42: The rationale requirement in both the *Inquiry Brief* and the *Inquiry Brief Proposal* asks that an argument be advanced to justify the selection of assessments. Which of the following arguments might be used in the rationale section? Circle the number(s) of the arguments that might be used in a rationale.

1. A description of the process by which the instrument (survey, rubric, test, interview schedule, etc.) was developed.
2. A description of a critique the instrument received from outside reviewers.
3. A content analysis showing how the instrument addresses elements of the construct that is being assessed – caring teaching, subject matter knowledge, etc.
4. If the assessment is a standardized measure published by a testing company, simply say so in the argument.
5. Let TEAC know that if the auditors are interested in the validity of the instrument, they should write to the publisher.
6. Simply advance the assumption that since the local authors of the instrument are all well prepared for their roles in the program, they must have developed a valid instrument.
7. Carry out a pilot study with candidates in the program, and assess the reliability of the instrument empirically and report the estimated reliability coefficient.
8. Report the program's historical experiences with the assessment instrument.
9. Explain how evaluators were trained in the use of the assessment.