**EPSY 8224 – Criteria for Class Project**

**Holding Assessments to High Standards**

Regardless of the purpose or format, quality assessments should meet certain common standards. The Center for Research on Evaluation, Standards, and Student Testing (CRESST), (Linn, Baker, and Dunbar 1991) has developed criteria that represent a touchstone throughout the assessment development process. The criteria include:

■ Consequences. Testing history is full of examples of good inten­tions gone awry. This criterion requires that we plan from the outset to assess the actual consequences of the assessment. Does it have positive consequences or are there unintended effects such as narrowing of curriculum, adverse effects on disadvan­taged students, and so on?

■ Fairness. Does the assessment consider fairly the cultural back­ground of those students taking the test? Have all students had equal opportunity to learn the complex thinking and problem solving skills that are being targeted?

■ Transfer and Generalizability. Will the assessment results sup­port accurate generalizations about student capability? Are the results reliable across raters, and consistent in meaning across locales?

■ Cognitive Complexity. We cannot tell from simply looking at an assessment whether or not it actually assesses complex thinking skills. Does an assessment in fact require students to use complex thinking and problem solving?

■ Content Quality. The tasks selected to measure a given content domain should themselves be worthy of students' and raters' time and efforts. Is the selected content consistent with the best current understanding of the field and does it reflect important aspects of a discipline that will stand the test of time?

■ Content Coverage. The content coverage criterion requires that assessment be aligned with the curriculum and, over a set of assessments, represent the full curriculum. Because time con­straints will probably limit the number of alternative assessments that can be given, adequate content coverage represents a signifi­cant challenge. Are the key elements of the curriculum covered by the set of assessments?

■ Meaningfulness. One of the rationales for more contextualized assessments is that they ensure that students engage in meaning­ful problems that result in worthwhile educational experiences and higher levels of motivation. Do students find the assessment tasks realistic and worthwhile?

■ Cost and Efficiency. To be effective tools, assessments must be cost effective. Labor-intensive performance-based assessments require efficient data collection and scoring procedures. Is the information about students worth the cost and time to obtain it?

Finally, it is important to note that alternative assessment is a developing field. New strategies are evolving as are new methodologies for ensuring their quality. As we learn more about alternative assessment, current approaches may be refined or even reformulated.

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: ASCD. pp. 10-11.

**Determining Priority Outcomes**

While designating goals may seem simple, it is challenging to set priori­ties from among the myriad possibilities. What major fields of knowl­edge, skills, and dispositions are worth teaching and worth assessing? What outcomes are you trying to achieve? Because performance assess­ments require considerable time and energy-both yours and your students-you will want to focus on a relatively small number of impor­tant outcomes, each perhaps representing a month or a quarter's worth of instruction. These assessments should aim at your major learning objectives for students. To help define these objectives, ask yourself this series of interrelated questions (to which we have supplied some sample responses):

1. **What Important Cognitive Skills Do I Want My Students To Develop?**

I want students to be able to:

• Communicate effectively in writing, or more specifically, to write persuasively, to write good descriptions, and to write stories.

• Communicate effectively orally.

• Analyze literature using plot, character, setting, and theme.

• Analyze issues using primary source and reference materials.

• Use algebra to solve everyday problems.

• Analyze current events from historical, political, geographic, and multicultural perspectives.

• Design and conduct studies to aid decision making about current or everyday problems.

• Use the scientific method.

• Use different media to express what they know.

2. **What Social and Affective Skills Do I Want My Students To Develop?**

I want them to be able to:

• Work independently.

• Develop a spirit of teamwork and skill in group work.

• Appreciate their individual strengths.

• Be persistent in the face of challenges.

• Have pride in their work.

• Enjoy and value learning.

• Have confidence in their abilities.

• Have a healthy skepticism about current arguments and claims.

• Understand that we all have strengths and that each person is able to excel in some way.

3. **What Metacognitive Skills Do I Want My Students To Develop?**

I want them to be able to:

• Reflect on the writing process they use, evaluate its effectiveness, and derive their own plans for how it can be improved.

• Discuss and evaluate their problem-solving strategies.

• Formulate efficient plans for completing their independent pro­jects and for monitoring their progress.

• Evaluate the effectiveness of their research strategies.

4. **What Types of Problems Do I want Them To Be Able To Solve?**

I want them to:

• Know how to do research.

• Solve problems that require geometric proofs.

• Understand the types of problems that trigonometry will help them solve.

• Apply the scientific method.

• Predict consequences.

• Solve problems that have no right answer.

• Make healthy choices.

• Create their own unique expressions.

5. **What Concepts and Principles Do I Want My Students to Be Able to Apply?**

I want them to be able to:

• Understand what a democracy is.

• Understand cause-and-effect relationships in history and in everyday life.

• Understand the meaning of various logical propositions.

• Criticize literary works based on plot, setting, motive, and so on.

• Understand and recognize the consequences of substance abuse.

• Apply basic principles of ecology and conservation in their everyday lives.

Be as specific as possible in formulating your answers to these questions. While you shouldn't produce the excruciating detail found in behavioral objectives of the past, you should describe your primary outcomes with enough detail that others can agree on what the outcomes mean and whether or not students have attained them.

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: ASCD. pp. 25-26.

**Table 3**

***Criteria for Performance Assessment***

1. Content/skill coverage and correct method

The assessment:

• Clearly states skills and content to be covered

• Correctly uses performance assessment to measure these skills and content

• Avoids irrelevant and/or unimportant content

• Deals with enduring themes or significant knowledge

• Matches statements of coverage to task content and performance criteria

2. Performance criteria

• Include everything of importance and omit irrelevant features of performance

• State criteria clearly and provide samples of student work to illustrate them

• Are stated generally, especially if the intent is use as an instructional tool

• Are analytical trait, especially if the intent is use as an instructional tool

3. Performance tasks

General

• Elicit the desired performances or work

• Recreate an "authentic'' context for performance

• Exemplify good instruction

• Are reviewed by others (students, peers, experts)

Sampling/representativeness/generalizability

• Cover the content or skill area well; results can be generalized

• Sample performance in a way that is representative of what a student can do

Bias and distortion

• Avoid factors that might get in the way of students' ability to demonstrate what they know and can do

4. Fairness and rater bias

Performance tasks

• Have content and context that are equally familiar, acceptable, and appropriate for students in all groups

• Tap knowledge and skills all students have had adequate time to acquire in class

• Are as free as possible of cultural, ethnic, or gender stereotypes

• Are as free as possible of language barriers

Performance criteria and rater training

• Ensure that irrelevant features of performance do not influence how other, supposedly independent features are judged

• Ensure that knowledge of the type of student does not influence judgments about performance quality

• Ensure that knowledge of individual students does not affect judgments about performance quality

5. Consequences

The assessment:

• Communicates appropriate messages

• Results in acceptable effects on students, teachers, and others

• Is worth the instructional time devoted to it; students learn something from doing the assessment and/or using the performance criteria

• Provides information relevant to the decisions being made

• Is perceived by students and teachers as valid

6. Cost and efficiency

The assessment:

• Is cost efficient-the results are worth the investment

• Is practical

Arter, J. (1999). Teaching about performance assessment. *Educational Measurement: Issues and Practice, 18*(2), 30-44.

**Describing your Assessment Task**

While the nature of the assessment task will dictate what needs to be specified, the following aspects usually need specification:

■ What outcome(s) are intended for the assessment?

■ What are the eligible content/topics?

■ What is the nature and format of questions to be posed to students? What is the audience for the response?

■ Is it group or individual work? If group work, what roles are to be filled?

■ What options/choices are allowed? What are the choices in response mode? What will they include, for example portfolios? Who makes the choices-the teacher or students or both?

■ What materials/equipment/resources will be available to students? Are there any specifications?

■ What directions will be given to students?

■ What administrative constraints are there? How much time is allowed? What is the order of tasks? How will student questions be answered? What help will be allowed?

■ What scoring scheme and procedures will be used?

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: ASCD. p. 41.

**Task Specifications**

**PA Class Project Section (d):**

1. Consider relevant task development guidance (e.g., Brookhart 1993, Table 1;

Solano-Flores, 1997, Table 4) to provide task specifications – how tasks are to be

developed

1. Briefly describe example task (given task characteristics and targets for consideration)
2. Provide sample examples
3. Describe context of assessment (conditions under which assessment is administered)

**Task Characteristics**

* Complete instructions
* Reading level
* Source of reference materials
* Familiarity
* Authenticity
* Level of support allowed
* Amount of structure, scaffolding
* Restricted or extended task
* Task length – response length expectations
* Minimize construct irrelevant features

**Targets for Consideration**

* Uncover thinking processes
* Problem solving
* Focus on higher level demands
* Integrate knowledge and skills
* Employ real-world context
* Allow for innovation, creativity

From the Handout on *Performance Assessment Task Development*

**Table 11.1**

***Guidelines for Writing CR Items***

**CONTENT CONCERNS**

1. Clarify the domain of knowledge and skills to be tested.

2. Ensure that the format is appropriate for the intended cognitive demand.

3. Ensure construct comparability across tasks.

**FORMATTING & STYLE CONCERNS**

4. Edit and proof instructions, items, and item formatting.

5. Pilot items and test procedures.

**WRITING THE DIRECTIONS/STIMULUS**

6. Clearly define directions, expectations for response format, and task demands.

7. Provide information about scoring criteria.

8. Avoid requiring implicit assumptions; avoid construct-irrelevant task features.

**CONTEXT CONCERNS**

9. Consider cultural and regional diversity and accessibility.

10. Ensure that the linguistic complexity is suitable for intended population of test takers.

Haladyna, T.M., & Rodriguez, M.C. (2013). *Developing and validating test items*. New York, NY: Routledge.

**Table 12.1**

***CR Scoring Guidelines***

**CONTENT CONCERNS**

1. Clarify the intended content and cognitive demand of the task as targets for scoring.

2. Specify factors in scoring that are irrelevant to the task demands.

**SCORING GUIDE DEVELOPMENT**

3. Select an appropriate scoring method.

4. Begin scoring guide development during task construction (Item writing).

a. Clarify distinctions across score points.

b. Define clear justifications within score points.

c. Do not over specify expected responses.

d. Expectations for the same cognitive demand should be the same across similar tasks and scoring rules.

5. Review actual responses to refine scoring guide.

**SCORING PROCESS**

6. Qualify raters.

7. Train raters.

8. Rate consistently.

9. Minimize bias.

10. Obtain multiple ratings.

11. Monitor Ratings.

Haladyna, T.M., & Rodriguez, M.C. (2013). *Developing and validating test items*. New York, NY: Routledge.

**Criteria for Rubrics**

**Trait 1: Content coverage**

1. Ready to roll

1. justification for the dimensions of performance
2. face validity
3. if counting, the counts are real indicators of quality
4. emphasis is right
5. definitions are correct – reflect current thinking in the field
6. the number of points makes sense

**Trait 2: Clarity**

1. Ready to roll
   1. Different teachers will give the same rating to the same performance/product
   2. Will yield consistent ratings across observations, participants
   3. Words are specific and accurate
   4. It is clear why samples were scored the way they were
   5. Terms are defined
   6. Enough descriptive detail
   7. Each score point is defined with indicators and descriptions

**Trait 3: Practicality**

1. Ready to roll
   1. Manageable, not too many things to attend to
   2. May need to translate results into training/instruction
   3. It is analytical for complex skills/ products
   4. If task-specific or holistic rubrics are used, their justification is clear and appropriate
   5. The rubric can be used by participants themselves to adjust performance

**Trait 4: Technical Quality (not evaluated)**

1. Ready to roll
   1. Technical information is available regarding rater agreement (at least 65% exact, 95% within one-point)
   2. Language used is appropriate given diversity of participants
   3. Formal bias reviews of rubric content

From *Metarubric* Handout

****

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: ASCD. p. 77.