Whether in the science or language laboratory, carrying out health care procedures or demonstrating performance arts, faculty can improve skill evaluation through transparency and authenticity in exam construction, format, and grading.

Performance-Based Assessment: Improving the Value of Laboratory and Skills Examinations

Judy M. Silvestrone

Assess it, and it becomes important. Assess it explicitly... with transparent criteria, and perhaps they will see why.

—James, 2000, p. 355

Enhancing the quality of examinations is always a creditable goal. Improving skill assessment presents unique challenges because of the nature of the evaluation and the traditions of various disciplines. In this chapter, I address how choices in the construction and grading of practical examinations can improve transparency (explicitness) and authenticity (relation to real tasks) and thereby increase the capacity to examine the most essential skill sets.

Transparency in examination refers to the explicit definition and communication of guidelines for evaluation (James, 2000; Moskal, 2003; Simon and Forgette-Giroux, 2001). Whereas this is consciously and consistently prepared for cognitive assessment in most courses, there are subtleties to skill examination that warrant special attention. In particular, skill assessment benefits from the development of specialized learning objectives, the elucidation of performance expectations, and the separation of cognitive from skills evaluation.

Authenticity refers to the use of acquired knowledge in relationship to other courses, ideas, and real-world tasks. Validity and reliability in any examination are improved by presenting real-world challenges (Wiggins, 1990). Skill examination benefits by the selection of the appropriate form
of evidence, delineation of the weight of various elements, and the identification of variables that can decrease validity. In the ensuing discussion, I examine means for maximizing transparency and authenticity in learning objectives, assessment construction, and grading.

**Learning Objectives**

One of the simplest methods faculty can use to ensure the quality of skill evaluation is to construct and communicate explicit learning objectives for each skill set. Although cognitive objectives are typically included in course syllabi, psychomotor and affective objectives may not be as well developed or expressly conveyed to students. Psychomotor objectives include specific performance criteria such as speed, dexterity, strength, and functional expectations. Affective learning objectives can include attitude and communication. For instance, in a nursing program, it would be appropriate to include both proper venipuncture technique (psychomotor) and appropriate patient concern (affective) as valued learning objectives. These objectives are vital to the future application of the skill and deserve appropriate consideration and assessment.

**Examination Construction**

There are several issues to consider with regard to the construction of examinations in courses using performance-based assessment.

**Cognition Versus Performance.** When constructing an examination, faculty should consider separating cognitive tasks from performance unless they are inextricably linked. Different understanding is demonstrated through action than through written examination. For assessment in the Spanish language laboratory, James (2000) indicates that disaggregation of performance helps students see the importance of individual skills and increases their transferability to other venues.

An instructor might consider offering a separate cognitive exam or provide a reading period before the examination for students to prepare responses necessary during the skill evaluation. In medicine, Wilkerson and Lee state that “knowing how to perform the physical examination and knowing when to perform it may be very different skills” (Wilkerson and Lee, 2003, p. S31, emphasis added).

**Context of the Test.** Context, however, remains important in the relevance of the skills to be learned, the manner in which they will be applied in future endeavors, and the realism of the conditions in which students are examined. Scrutinizing the expected application of the skill allows the suitability of both the objective and method of assessment to be determined. For example, does sight-reading have a convincing rationale as part of a performance evaluation when music students would otherwise realistically have rehearsal time before recital? The answer may be that sight-reading is
a crucial requirement, but the decision to use it must be made consciously and with the targeted application in mind. The reasons for conditions imposed on performance should be made explicit to students so that the relevance of the limitations are evident, rather than appearing arbitrary.

**Appropriate Evidence.** Moskal and Leydens (2000) indicate that reflection on the learning objectives and purpose of the assessment will make apparent the best forms of evidence for identifying these skills. Examination of the method should prompt the question: “Does this method truly reveal competency in the skill sets I want to examine?” For example, is there a strong rationale for examining students in an anatomy lab by having them rotate through thirty one-minute stations in which they identify tagged structures? What would be more typical of the real-life demands of a paleontologist or a physician? Will such a rapid pace ever really be essential? In real life, what is more critical, accuracy or speed?

For evidence of student development over time, teachers in many disciplines use portfolios of accumulated work accompanied by a self-reflection on progress and process. The use of videotaped skill performance over time with student self-assessment can provide a profound learning experience for the student and good evidence of development of the skill.

**Overcoming Measurement Obstacles in Performance Testing**

Obstacles that limit accuracy of skill assessment include situations that require transfer before the student is ready, time constraints, and performance anxiety.

**Too Far a Transfer Required.** Although transferability of skill is a goal of instruction, an examination should not be the first time that a student is asked to use a skill in a new manner or for a new situation. Students should get the opportunity to rehearse using skills in various circumstances; explore connections between cognitive and skill sets; and practice using the critical thinking, synthesis, or problem solving expected of them before being asked to do so for evaluative purposes.

**Not Enough Time Allowed.** Another pertinent question is whether the use of time constraints in performance assessment is necessary. Such constraints are more frequently found in performance tests, which must be administered to a series of students in a short time or which involve several students needing access to equipment sequentially, as in a lab stations format. Temporal limits may be required because of class time but can just as often be based on faculty convenience or disciplinary tradition. There may be alternative forms of evaluation that allow for a more accurate assessment of the skill set. Can the same behaviors be assessed by a less complex procedure or with a checklist rather than a full-blown demonstration? For example, is it really necessary for a student to work all problems completely if demonstrating how to set up the problem correctly
would 90 percent of the time result in a correct computation? Can time on assessment be saved by targeting the key points rather than requiring a full demonstration each time?

The use of other class sessions (such as recitation sessions) or open lab hours to allow more examination time could be considered, as well as the accumulation of quiz opportunities rather than one-shot tests to assess student progress and retention over time. Discussion with colleagues, teaching assistants, and student focus groups (such as small-group instructional diagnosis) may result in some innovative methods for performance assessment that will not be hampered by time constraints.

For some settings—food preparation, emergency procedures, or simultaneous translation—performance under time constraints may be crucial. A common recommendation for those instances is to allow one-and-a-half greater time than required by a graduate student for adequate performance of the same assessment (Smee, 2003). However, it has been my experience that when an exam is constructed well (and externally reviewed), most students do not need the full time that was initially allotted. When the urgency of time is eliminated, students perform better and often faster. When a student’s progress through a performance test is affected by the progress of others (such as in moving from one lab station to the next), credit can be given for timely completion or for partial completion to encourage students not to linger over a single station.

If time is essential, removing disruptive prompts to indicate time passing (such as a bell, buzzer, or timer) may improve accurateness of evaluation. Having a human voice prompt to “move to the next station or question” provokes less anxiety. When multistation examinations are necessary, use a “revisit” period at the end of the examination when students can check back on a few stations. Unnecessary transit time and time spent orienting to a specimen or problem can be reduced by decreasing the number of stations or problems and asking multiple questions at each. Wilkerson and Lee (2003) recommend that if time constraints are an essential component of the skill, time management should also be part of instruction and practice for that skill. The bottom line has never been more accurately summarized than by a question raised by one professor: “Do we want our students to know it better or faster?”

**Coping with Performance Anxiety.** Whatever assessment type is chosen, sharing the details of the format, conditions, and process of the examination with students improves exam quality. Skill assessment inherently carries with it the increased risk for test anxiety (a form of performance anxiety) due to the requirements of demonstration and observation.

For an assessment of what students know to be fair and accurate, with minimal influence by performance anxiety or lack of familiarity with exam format, students need to be apprised of precise expectations. Sharing objectives, test system, and grading criteria with students early in the course is vital to decrease their anxiety with novel exam formats (Moskal, 2003;
McGregor and Elliott, 2002). Simulated examinations for practice can provide grading opportunities, frequent feedback on performance, and familiarity with exam schema. There are special concerns when videotaping or audiorecording students’ performance. If they are used only in the testing environment, students may view their purpose as potentially punitive or threatening. The purpose of recording should be made explicit, previously modeled, and best used for self-assessment.

**Grading the Performance**

In the grading process, an explanation of the weight of various elements, attention to level of detail, and distribution of sample grading forms and graded work can support a quality assessment and reduce grading challenges.

**What Is Included?** Performance can be graded on the use of process, the final product, or a combination of both. The relative value and effect of these elements should be delineated (Moskal and Leydens, 2000; Woods and others, 2001). Just as in a cognitive problem in mathematics, students should be aware what weight is assigned to the reasoning or construct behind the performance and the resultant outcome. For example, in a musical recital, what is the expected weight of technical performance versus emotive factors? Transparency of partial credit policies also serves to decrease perceived variance and subjectivity in grading.

**What Level of Detail Is Expected?** Explicit grading criteria, weight, and level of detail help students understand expectations (Silvestrone, 2002). Students should be aware of the value assigned to different aspects of their performance. Some errors or omissions may have relatively little weight, but others may prompt inevitable failure or retake. A radiological technologist might fail an examination by taking the wrong x-ray image on a patient (or simulation) but not for using the incorrect film size. For a comparative anatomy examination, if a student has identified a groove on a bone, does he or she also need to name that bone? Is the identification required in English or Latin nomenclature? Examples of expected work (such as vocabulary lists) may help students identify the level of detail necessary. Including these lists in the syllabus or before the examination gives students a necessary frame of reference for the depth and orientation of their response or performance.

**Value of a Good Model.** Furthermore, there is a strong foundation for providing graded examples to students. Seeing examples of student work and the reasons for the grade assignments communicates standards to new students and lecturers (James, 2000; Simon and Forgette-Giroux, 2001). Even more benefit can be derived if students see multiple examples or are allowed to participate in the evaluation of an example. When students see how the criteria are being applied, they get a clearer idea of how to prepare their own responses to an exam.
**Inter-Rater Reliability.** Idiosyncratic grading can be an issue with performances being evaluated by multiple raters or multiple sections (Baughin, Brod, and Page, 2002; Smee, 2003). To increase inter-rater and intrarater reliability, the use of primary trait analysis (grading rubric or matrix) is optimal (Simon and Forgets-Giroux, 2001; Chapter Twelve of this volume). With a substantial history in primary and secondary education, rubrics have special value for performance-based examination in higher education courses and programs. Rubrics can decrease grade variance with multiple raters, multiple assessments, and student progress over time. Teaching assistants and lead faculty can work together to construct rubrics with up to five levels of accomplishment over a broad number of evaluation areas. James (2000) recommends the use of rubrics in the language laboratory to allow evaluation of technical skills as well as language. Rubrics allow for a single rater to evaluate key skill components with the stroke of a pen or the touch of a palm-held computer screen.

An additional benefit of rubrics is that grade challenges are decreased when students have the opportunity to see the evaluative criteria and attendant grade levels in advance (Moskal and Leydens, 2000). Students use these criteria for exam preparation, monitoring their content coverage, and self-feedback.

Many authors espouse the use of a three-tiered rubric, with expected characteristics, behaviors, or skill proficiency delineated for each level. Such ratings as “good, pass, not ready” or “meets expectations, exceeds expectations, does not meet expectations” with descriptors provide sound feedback for students and support consistency in grading. Instructions for their construction and sample rubrics may be found in Walvoord and Johnson’s (1998) *Effective Grading*, or on Web sites such as http://www.rubistar.4teachers.org or http://www.teach-nology.com/web_tools/rubrics. Many disciplines have rubrics already constructed in light of the outcomes-based assessment movement affecting certification by professional associations.

**Conclusion**

Increasing authenticity and transparency of learning objectives and grading criteria does not entail decreasing intellectual challenge or integrity of skill assessment. The quality of an examination increases when learning objectives are constructed in depth, clearly communicated, and applied throughout examination administration and grading. Both students and faculty benefit when this communication is coupled with strategies for reducing variables that decrease validity and relevance of the skills and assessment to the larger world.

Moskal (2003) states the goals of skill assessment nicely when discussing the use of rubrics in engineering: “Performance assessments require more time to administer than do other forms of assessment. The investment of this classroom time should result in a higher payoff. This payoff should
include both an increase in the teacher’s understanding of what students know and can do and an increase in the students’ knowledge of the intended content and constructs” (p. 2).

References


Wilkerson, L., and Lee, M. “Assessing Physical Examination Skills of Senior Medical Students: Knowing How Versus Knowing When.” Academic Medicine, 2003, 78(10), S30−S32.


JUDY M. SILVESTRONE is professor of clinical sciences at New York Chiropractic College and director of the Center for Academic Support.