What is Satisfactory Performance? Measuring Students and Measuring Programs with Rubrics

Some assessment experts strongly recommend that a desired level of achievement be stated when measuring student performance on stated student learning outcomes. According to Nichols, the criteria should be stated in quantitative terms, as this example illustrates: "Eighty percent of those taking the CPA exam each year…will pass three of four parts of the exam" (Nichols, 1989, p. 178). In the era of rubrics, this can easily be translated to "Eighty percent of students…will score at least 'satisfactory' on three of the four rubric rows."

But why eighty percent? The world of measurement is full of numbers used as rules of thumb, such as the familiar 95% confidence interval. If the performance desired is "passing", 70% should be adequate, or even 60% if a D is considered a passing grade. But such thinking conflates grading, with its traditions of placing students into a distribution of some sort, with quality assurance, which implies that all who complete some educational experience should meet some minimum standard of quality.

The purpose of rubric usage may affect the choice of an acceptable level of performance. Two common purposes are the diagnostic study of student performance to find areas of strength and weakness and the evaluation of program effectiveness. Although a database of rubric scores can be used for both purposes, the latter is what many stakeholders are looking for today. The latter would require setting a program-wide acceptable level, such as the 80% referred to above. The concept of acceptable performance is often included in rubric scores or levels (which often appear as columns in a matrix), but it would not be necessary to set a program-wide acceptable level simply to compare student performance on one criterion or trait (often represented by the rows) with another.

For comparison, we might look at a standard of quality from an established professional field, public accounting, which employs the CPA exam Nichols' hypothetical example used. According to its website, the American Institute of CPAs' Uniform CPA Examination is not constructed to place test-takers into a normal distribution, but to discriminate among them. The acceptable level of performance is determined by experts in the field who have supervised beginning CPAs. They start by rating samples of examinee performance on the four subtests, each with both multiple-choice and constructed responses. "For each item type (multiple-choice, simulation, essay) panelists will rate candidate performance profiles as either failing, passing, or borderline (just passing)" (American Institute of CPAs, 2010). After the initial rating, the expert panelists discuss one another's ratings and submit final ratings, which are averaged for the final recommendation to the Board of Examiners. That board also considers the resulting pass rates if the recommended passing score is implemented.

This description suggests that the procedures for creating the CPA exam pursue three goals: discrimination among

test-takers, assurance of uniformity of versions, and a convenient method of substituting a test for expert judgment. Even though rubrics directly record expert judgment using raters for every candidate, they need the other two qualities as well.

How can rubrics discriminate among subjects? Linda Suskie (2009) suggests that faculty construct rubrics using samples of student work and conferring with colleagues as to what is exceptional, acceptable, and inadequate. Even though these three levels give each rubric row only minimal discrimination, multiple rows can augment this discrimination. Another approach is David Dirlam's use of a theory of the development of professional expertise or skill, asking faculty to describe what each of the developmental stages (beginning, easy, practical, inspiring) would look like on each rubric row (Dirlam, 2011). The AAC&U VALUE rubrics were created with a similar developmental philosophy: "learning develops over time and should become more complex and sophisticated as students move through their curricular and cocurricular educational pathways toward a degree." The "over time" concept is implicit in the labeling of the wording of the column heads (Milestones, Capstone) and the wording of the descriptions in the cells.

Once criteria are agreed upon, the focus shifts to how rubrics are utilized to score student work. Uniformity in the rubric as an instrument, analogous to the CPA exam's uniformity of versions, is dependent not only on the rubric itself but also on the assignment given to the students whose work will be rated. This consideration may be neglected when faculty adopt a rubric from another source as though it were a standardized test. The rubric may be like the answer key to a test, but the assignment is analogous to the questions in a test. The rubric and the assignment to which it will be applied are yoked together to create the actual measurement, and there has to be a solid relationship between the two. Secolsky and Wentland (2010) allude to this when discussing the application of rubrics to collections of varied student work within a portfolio.

Once the rubric is sufficient to discriminate among student products and uniformity is established, two additional questions become pertinent: (1) what constitutes satisfactory performance for the individual student, and (2) what constitutes satisfactory performance for the program? Satisfactory individual performance is implicit in both kinds of rubric design mentioned above: with the developmental approach, satisfactory performance would be achievement that matches the student's current stage. Obviously, setting an acceptable level of performance for the program could be as simple as stating that "all students will score 'satisfactory' or better on all rows." More complex formulas allow for the percentages of students who may be expected to score over or under the "satisfactory" level, depending on faculty judgment and goals (Allen, 2006).

Even though the unit of assessment is usually considered to be the program, not the student, the satisfactory preparation of individual students is the end result of the program. Given that philosophical stance, there is little rationale for accepting 80% of students scoring at satisfactory. A 20% failure rate would be unacceptable in any other context. Faculty should determine what combination of scores will characterize a well-prepared student, much like the CPAs do with their professional entry examination. This may be very complex, and the data will be useful for diagnosis, but the overall aim should be for all students (100%) to reach the expected level of performance.

References

Allen, M. J. (2006). *Assessing general education programs*. Boston, MA: Anker Publishing Company.

American Institute of CPAs (2010). *Setting passing scores on the uniform CPA examination*. Retrieved on 3/15/2012 from http://www.aicpa.org/BECOMEACPA/CPAEXAM/PSYCHOMETRICSANDSCORING/Pages/PsychometricsandScoring.aspx

Association of American Colleges and Universities (2012). *VALUE: Valid assessment of learning in undergraduate education*. Retrieved on 3/15/2012 from http://www.aacu.org/value/index.cfm.
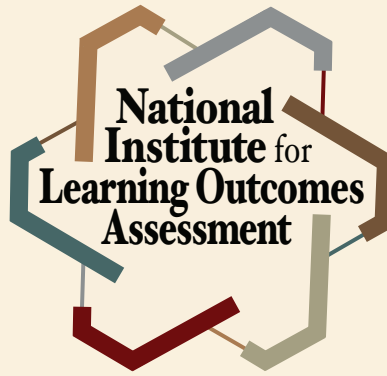
Dirlam, D. (2011, June). *Using developmental interviews to create learning outcomes networks*. workshop presented at the annual conference of the Association for Assessment of Learning in Higher Education, Inc., Lexington, Kentucky.

Nichols, J. O. (1989). *Institutional effectiveness and outcomes assessment implementation on campus: A practitioner's handbook*. New York, NY: Agathon Press.

Secolsky, C., & Wentland, E. (2010). Differential effect of topic: Implications for portfolio assessment. *Assessment Update: Progress, Trends, and Practices in Higher Education, 22*(1), 1-2, 15.

Suskie, L. (2009). *Assessing student learning: A common sense guide*. (2nd Ed.). San Francisco: Jossey-Bass.

Walvoord, B. E. (2004). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. San Francisco, CA: Jossey-Bass.

Please Cite As:

Follow us on social media:

@NILOA_web

@LearningOutcomesAssessment

Sign up to receive our monthly NILOA Newsletter and stay up to date with our research and publications.