

2006-723: DEVELOPING PERFORMANCE CRITERIA AND RUBRICS FOR BIOMEDICAL ENGINEERING OUTCOME ASSESSMENT

Kay C Dee, Rose-Hulman Institute of Technology

Kay C Dee is an Associate Professor of Applied Biology and Biomedical Engineering at Rose-Hulman Institute of Technology. Her educational research interests include learning styles, teaching faculty about teaching, student evaluations of teaching, and assessment. Her teaching portfolio includes courses on: biology; biomaterials; cell-biomaterial interactions; cell and tissue mechanics; bioethics, science fiction, and tissue engineering; interdisciplinary engineering problem-solving; and teaching engineering.

Developing Performance Criteria and Rubrics for Biomedical Engineering Outcome Assessment

Abstract

This paper describes the use of criteria and rubrics to evaluate student achievement of biomedical engineering program outcomes. As a case-study example, a process for creating and using performance criteria and rubrics for program assessment is presented, and the evolution of a sample biomedical engineering-related rubric is described. This paper also includes a “how-to” section, to help faculty develop and critique their own performance criteria and rubrics.

Introduction – Accreditation as Motivation for Performance-Centered Assessment

All engineering programs seeking accreditation from ABET (the major accrediting body of university/college engineering programs within the United States) are required to define the types of skills that students will possess by the time they graduate, and to provide evidence that program graduates possess a set of skills/knowledge designated by ABET. In other words, and using ABET terminology, all engineering programs seeking ABET accreditation must define and measure student achievement of *Program Outcomes*, and through this process, must demonstrate that their students attain an ABET-designated set of abilities, criteria lettered “a” through “k”^[1]. Biomedical Engineering programs must additionally demonstrate that their graduates have: “an understanding of biology and physiology, and the capability to apply advanced mathematics (including differential equations and statistics), science, and engineering to solve the problems at the interface of engineering and biology; the ability to make measurements on and interpret data from living systems, addressing the problems associated with the interaction between living and non-living materials and systems.”^[1]

Curriculum maps or topic lists that show how Program Outcomes or ABET criteria are addressed across a curriculum or within given courses can be used to demonstrate that certain types of material *are presented to students*, but these lists *do not provide evidence of students mastering the designated skills*. Moreover, “Student self-assessment, opinion surveys, and course grades are not, by themselves or collectively, acceptable methods for documenting achievement of outcomes,”^[2] since these assessments provide evidence of either student opinions, or of generalized student achievement across a potentially broad area of study. Programs seeking ABET accreditation must use an assessment strategy which demonstrates the level of student achievement of clearly-defined, designated criteria. Ideally, the assessment strategy will also have the ability to be logically coordinated across a program as a whole; provide feedback that is informative as well as easily organized and interpreted; and facilitate reflection and improvements on multiple levels – from specific, focused areas of the program to a broad, holistic overview of the program. To meet these needs, we have chosen to assess our students’ achievement of Program Outcomes with performance criteria and associated evaluative rubrics.

Background – Performance Criteria and Rubrics

Performance criteria and rubrics are used extensively for educational assessment in many fields and at many levels. **Performance criteria are statements that set standards for the demonstration of an outcome – in this case, of an educational outcome.** Performance criteria are more specific than goals, since they set some standard for determining whether or not a goal has been achieved, but are generally broad enough in scope to encompass multiple individual tasks that might be associated with achieving a goal. For example, in Table 1, the performance criterion is more specific than the related educational goal, since it clarifies that students who have learned to present experimental results should be able to apply statistical analyses to data, and to use data to create graphical displays. A student could demonstrate achievement of this performance criterion through the completion of many different tasks or assignments, only one of which is given in Table 1 as an example.

Table 1. Levels of Specificity: Goals, Criteria, and Tasks.

Educational Goal	Performance Criterion	Individual Task/Assignment
Learn to present experimental results.	Demonstrate the ability to present experimental data, using graphical displays and statistical analyses.	Create a dot plot using the data in the attached file. Include a linear regression trendline for the data and display the regression equation and the R-squared value in the lower right corner of the plot.

Rubrics are descriptions of how the standards set in performance criteria can be met, at varying levels of quality. Rubrics can be used to discriminate between many levels of performance (for example, ‘excellent,’ ‘good,’ ‘fair,’ ‘poor,’ ‘unacceptable’) or few levels of performance (e.g., ‘pass’ or ‘fail’). Rubrics can include a scale of points to be awarded across a continuum of performance levels, or can be intended to bin samples into descriptive categories.

Well-written rubrics clearly *define components necessary to demonstrate levels of achievement* of performance criteria, and are *supplemented with specific indicators or examples* of what would constitute the various levels of performance to be determined. Well-written rubrics *focus on identifying the presence or absence of demonstrable items, avoiding overly broad or vague terms* that are difficult to directly demonstrate (e.g., ‘appreciate,’ ‘understand’) or that are very likely to be interpreted in different ways by different reviewers (e.g., ‘appropriate,’ ‘adequate’).

One possible rubric and examples for the performance criterion given in Table 1 is:

“The submission should:

- graphically display data, including the mean values of and an indication of the range of variation within the data,
- include a statistical analysis of any data trend, with the level of statistical significance clearly presented.”

Accompanying this rubric might be these examples:

Fail

The submission does not include a plot of the data. The submission includes a statistical analysis but does not define or present the level of statistical significance. The submission plots mean data values but gives no graphical representation of the variation within the data.

Pass

Example passing submissions might be a plot showing mean values and standard deviations of data, with a written paragraph describing the results and statistical significance of a t-test used to analyze the data.

Performance criteria and rubrics can be created to assess student achievement of Program Outcomes. Having faculty use the criteria and rubrics to evaluate samples of student work may increase “buy-in” to an overall program agenda of continual improvement, since developing and using criteria/rubrics for outcomes assessment is, in essence, setting performance standards and establishing directions for change. Additionally, this assessment strategy:

- Provides a framework for program faculty to discuss openly and come to consensus on the educational priorities of the program.
- Generates clear guidelines for instructors and students about the levels of achievement expected.
- Directly assesses demonstrated evidence of student abilities.
- Reinforces educational continuity when different instructors teach the same course or different sections of the same course, since the same criteria are assessed in each course/section.
- Gives instructors the opportunity, through reviewing and rating submissions of student work, to become more familiar with what their colleagues are teaching and what the program’s graduates are learning in other areas.
- Yields clear feedback on potential improvements for individual Program Outcomes or courses as well as for the program as a whole.

The main disadvantage of this strategy is essentially the time commitment required to determine rubrics/criteria, review submitted evidence, and determine appropriate individual and program-level responses to the results. Note, however, that all engineering programs seeking ABET accreditation will have to devote time to setting up assessments, acquiring results, and determining how to use the information generated by the assessments. We believe this strategy yields quality, useful information without requiring the expenditure of inordinate faculty or administrative time.

Case-Study Example – Our Experiences

Introduction

Our Biomedical Engineering program began working on a program assessment strategy by discussing direct assessment techniques, in which students demonstrate learning by doing

something (e.g., passing a test), and indirect assessment techniques, in which students report about learning (e.g., surveys, interviews, focus groups)^[3]. We used material from chapters 5 and 6 of *Assessing Academic Programs in Higher Education*^[3] as a basis for this discussion, since these chapters are brief, informative, and accessible to faculty who are not involved in assessment or educational research. We quickly decided that direct assessments at the core of our program-level assessment strategy, with supplemental indirect assessments where needed or desired, would meet our accreditation-related needs and were likely to yield useful information.

We next needed to choose a direct assessment strategy. We could not think of an external national, regional, or discipline-specific exam that we believed would appropriately assess our Program Outcomes. Furthermore, we did not want to create, score, test, validate, and convince students to take seriously our own version of a Biomedical Engineering mastery exam. We wanted an assessment strategy that could be smoothly incorporated into and coordinated across our existing courses and curriculum.

Our institution has used performance criteria and associated rubrics to annually evaluate student electronic portfolios, for institutional-level assessment purposes, for a number of years^[4-6]. It made sense to use data from the institutional reviews of Biomedical Engineering student electronic portfolios as direct evidence of student learning where appropriate. However, the majority of our technical and discipline-specific Program Outcomes were not assessed by these institutional reviews.

Choosing to use a performance criteria/rubric system at the program level allowed us to utilize the experience of the faculty and staff who had previously participated in the institutional rubric/criteria revisions and electronic portfolio reviews. ***Would we have chosen this strategy if we had no pre-existing institutional assessment system?*** Likely, yes, due to the advantages of performance criteria and rubrics as listed in the ‘Background’ section of this paper.

Developing The Criteria and Rubrics

We decided on general parameters of our assessment strategy before working through details. First, we decided to collect samples of student work through the institutional student electronic portfolio system when convenient, and through instructor-generated ‘collective portfolios’^[3] – samples of student work, saved by the instructor in the format of their choosing – where convenient. We planned to use or modify existing course assignments, projects, *etc.* as the basis for our review whenever possible.

We next identified courses which most clearly required students to demonstrate achievement of our Program Outcomes. Faculty teaching these courses volunteered to save paper or electronic copies of all the student work submitted (*i.e.*, work from all students in the class) for something (an assignment, a test question, a project, a report) that the instructor believed – if completed correctly – would demonstrate achievement of a designated, relevant Program Outcome.

Prior to beginning to write our criteria and rubrics, we decided to keep the rubrics as simple as possible. At this initial stage we were primarily interested in whether a given sample of student work did or did not demonstrate achievement of a Program Outcome, so we created categories of

“Pass” and “Fail.” We decided that our next priority would be to encourage and recognize excellence, so we created the category of “Pass Commended” for samples of work that were extremely well-done.

Our writing began with a team of faculty writing one performance criterion as well as an associated scoring rubric and examples. This team writing exercise took approximately 45 minutes and involved a good deal of discussion; it proved a useful calibration exercise since it let us reconfirm definitions and ideas, and work through some of the guiding questions listed in the last section of this paper. We next asked pairs of faculty members to choose the Program Outcomes most closely related to the areas of interest/courses taught by those faculty, and asked each pair to draft criteria, rubrics, and examples for those Program Outcomes. Drafts were distributed electronically prior to another meeting, at which the whole team discussed each set of criteria/rubrics and examples. The paired faculty members revised their drafts, the assembled team met again to discuss them, and after one final round of revisions our initial set of rubrics and criteria were established.

We then used our criteria and rubrics to evaluate the samples of student work that had been saved from the previous academic year. *Each sample of student work was examined by a pair of faculty raters*; these faculty were chosen such that *no instructor reviewed material that had been saved from one of their own courses*. Each *pair of raters came to agreement* on which samples passed the criteria; for submissions that did not pass the criteria, the pair of raters came to agreement on the reasons why it did not pass. We saved the samples of student work and the raters’ comments for the record, and calculated the percent of submissions that passed the criteria.

Our goal is to have 90% or more of submitted student work pass the criteria we have established for each Program Outcome. If between 70 and 89% of the submissions pass the criteria for a Program Outcome, we pay more attention to this Outcome over the next academic year. This attention may entail increased coverage in a course or courses, additional assessments, and/or minor revision of the assignment(s) used to demonstrate the competency in question. If less than 70% of the submissions pass the criteria for a Program Outcome, we are committed to discussing and determining a program-level response. In the past this type of response has included increased coverage in multiple courses, assessment in additional classes, major revision of the assignment(s) used to demonstrate the competency in question, and/or reconsideration of the rubric used for assessment.

Following our initial program-level review, we revisited our performance criteria and rubrics to address any inconsistencies or redundancies that had been noticed by the faculty raters. We then began the next cycle of review. As such, our current program assessment cycle starts in the Fall with a meeting to discuss the results from the last review, to confirm which courses and instructors will be saving samples of student work throughout the academic year, and to confirm any assessment changes or changes in criteria and rubrics. Indirect assessments that occur during the year are discussed in departmental meetings as they occur, and are tabulated for subsequent review in the context of our direct assessments. Shortly after classes have finished in the Spring, our faculty meet to conduct a Program Review, in which pairs of faculty use our performance criteria and rubrics to evaluate student achievement of our Program Outcomes. We

tabulate and discuss the results, re-revise any criteria and rubrics if necessary, and we are then prepared for our first Fall meeting and another assessment cycle.

The Evolution of a Sample Rubric

One of our Biomedical Engineering Program Outcomes is that students “will have an advanced and current body of knowledge within biomaterials, biomechanics, or biomedical instrumentation.” As we reviewed and revised our overall assessment strategy, the rubric associated with this Program Outcome has changed four times, as described below and shown in Table 2.

The performance criterion we initially established for this Program Outcome has remained the same: “A student should demonstrate possession of an advanced and current body of biomedical engineering knowledge within the subfield of biomaterials, or biomechanics, or biomedical instrumentation.” In this case, our Program Outcome was more specific than a general educational goal of, say, “being competent in biomaterials, biomechanics, or biomedical instrumentation,” and we wanted to keep the performance criterion broad enough to be applicable to all three sub-areas of biomedical engineering. A fairly broad performance criterion therefore suited our needs.

The rubrics and examples first evolved to provide more specific guidance, and more specific examples, for reviewers. Instead of stating that a passing submission should “involve both living and non-living materials or systems,” a next draft stated that a passing submission should “involve challenges associated with interfaces between biologically-derived/living and man-made/non-living materials or systems.” Upon further reflection we realized that a student could demonstrate advanced and current biomedical engineering knowledge without necessarily directly addressing challenges associated with interfaces between living/biologic and non-living/artificial items. Since we already had, and were assessing, another Program Outcome which clearly involved these interfaces and challenges, we ultimately revised this rubric to eliminate the non-essential requirement. Instead of using comparative examples which were likely to be interpreted by individual faculty in different ways (*e.g.*, “at a level comparable to that expected in the top ten percent of a class of graduate students,”), we tried to use comparators that would be familiar to the vast majority of our program faculty (“on a level we might teach in BE200” is familiar to nearly all of our faculty since we take turns team-teaching this course) and to be as specific as possible about defining different levels of achievement.

Developing and Critiquing Your Own Rubrics

Because performance criteria and rubrics are used widely for educational assessment at levels ranging from pre-kindergarten through advanced university work, it can be difficult to sort through the many web sites and published articles. The intent of this section of the paper is therefore to provide a few key resources and guidelines that can be used as a broad starting point to help faculty consider whether a criteria/rubric-based direct assessment strategy would be useful for their own courses and/or programs, and to develop and critique their own performance criteria and rubrics. There are a number of very good online resources^[7] that can help faculty

Table 2. The Evolution of a Sample Rubric.

Performance Criterion: Demonstrate possession of an advanced and current body of biomedical engineering knowledge within the subfield of biomaterials, or biomechanics, or biomedical instrumentation.		
<p>Rubric, Version 1 The submission should:</p> <ul style="list-style-type: none"> • show knowledge of a current and/or advanced topic in the relevant subfield, • demonstrate understanding of both a) biology/physiology and b) engineering, mathematics, or science, • involve both living and non-living materials or systems. 	<p>Rubric, Version 2 The submission should:</p> <ul style="list-style-type: none"> • show knowledge of a current and advanced topic in the relevant subfield, • demonstrate understanding of both a) biology/physiology and b) engineering, mathematics, or science, • involve challenges associated with interfaces between biologically-derived/living and man-made/non-living materials or systems. 	<p>Rubric, Version 2 The submission should:</p> <ul style="list-style-type: none"> • show knowledge of a current and advanced topic in the relevant subfield, • demonstrate the ability to solve problems at the interface of engineering and biology/physiology.
<p>Examples: Fail</p> <ul style="list-style-type: none"> • The submission demonstrates understanding of only biology/physiology or engineering, mathematics, or science, rather than both sides of the interface between biology and engineering. • The submission deals with a trivial or fundamental topic, on a level we might teach in BE200. Examples: a broad general description of a tensile test of a tissue; describing the gross anatomy of the heart and acknowledging that heartbeat can be measured; recalling that metal biomaterials are often used for orthopaedic implants. 	<p>Examples: Fail</p> <ul style="list-style-type: none"> • The submission demonstrates understanding of only biology/physiology or engineering, mathematics, or science, rather than both sides of the interface between biology and engineering. • The submission deals with a trivial or fundamental topic, on a level we might teach in BE200. Examples: a broad general description of a tensile test of a tissue; describing the gross anatomy of the heart and acknowledging that heartbeat can be measured; recalling that metal biomaterials are often used for orthopaedic implants. 	<p>Example: Fail</p> <p>The submission deals with a trivial or fundamental topic, on a level we might teach in BE200. Examples: a broad general description of a tensile test of a tissue; describing the gross anatomy of the heart and acknowledging that heartbeat can be measured; recalling that metal biomaterials are often used for orthopaedic implants.</p>
<p>Example: Pass</p> <p>The submission meets all three requirements of the rubric.</p>	<p>Example: Pass</p> <p>An example submission would be a derivation of viscoelasticity equations relevant to confined compression testing of cartilage, with equation parameters explained in terms of the molecular-level physiology of the tissue, and some evaluation of how well the equation describes behavior directly under the indenter as opposed to at the confined edges.</p>	<p>Example: Pass</p> <p>An example submission would be a derivation of viscoelasticity equations relevant to confined compression testing of cartilage, with equation parameters explained in terms of the molecular-level physiology of the tissue, and some evaluation of how well the equation describes behavior directly under the indenter as opposed to at the confined edges.</p>
<p>Example: Pass Commended</p> <p>The submission meets all three requirements of the rubric while demonstrating assimilated knowledge or analytical capabilities at a level comparable to that expected in the top ten percent of a class of graduate students.</p>	<p>Example: Pass Commended</p> <p>The advanced and current nature of the passable submission is clearly supported and documented within the submission (for example, with a literature review, references, etc.), the submission is well-grounded in both biology/physiology and engineering, mathematics or science, and the submission integrates multiple aspects of biomedical engineering rather than remaining entirely focused within one subfield.</p>	<p>Example: Pass Commended</p> <p>The advanced and current nature of the passable submission is clearly supported and documented within the submission (for example, with a literature review, references, etc.), the submission is well-grounded in both biology/physiology and engineering, mathematics or science, and the submission integrates multiple aspects of biomedical engineering rather than remaining entirely focused within one subfield.</p>

learn about creating rubrics and view rubrics created for different subjects and educational levels (for example, Rubistar^[8] or Rubrician^[9]). For other examples specific to biomedical engineering, past ASEE proceedings papers have presented information on rubrics for deriving program assessment data and grading undergraduate projects in a physiology course^[10] or evaluating undergraduate design presentations^[11].

Regardless of the educational level or subject, creating a system of performance criteria and rubrics is similar to the “big picture design approach” to education advocated by Wiggins and McTighe in *Understanding By Design*^[12]. As you write performance criteria, and then move on to related rubrics, you will essentially be asking three questions, here quoted directly from Wiggins and McTighe:

1. What is worthy and requiring of understanding?
 2. What is evidence of understanding?
 3. What learning experiences and teaching promote understanding, interest, and excellence?”^[13]
- In other words, what should your criteria encompass? What would constitute evidence of achieving those criteria? What does or could occur in the context of a course/curriculum that would demonstrate and promote achievement of those criteria?

When writing or critiquing performance criteria, it may be helpful to consider different types of criteria. In *Educative Assessment: designing assessments to inform and improve student performance*^[14], an excellent resource, Wiggins describes different types of criteria including “Impact of performance,” “Work quality and craftsmanship,” “Adequacy of methods and behaviors,” “Validity of content,” and “Degree of expertise.”^[15] These categories may not be equally important – and some may not even be of interest – for all assessments, but these categories provide a useful way to start thinking about what types of things can/should be included in performance criteria. Bloom’s taxonomy, or other taxonomies of ways of knowing^[16], might also be useful in developing performance criteria.

Taxonomies of cognitive processes can also be useful in writing rubrics. How might a student show “understanding” of something: might they interpret, classify, compare, explain, summarize, map, construct a model, paraphrase something?^[16] In *Understanding By Design*, Wiggins and McTighe provide a rubric describing six facets of understanding^[17]. This thought-provoking rubric can be helpful in moving assessments past simply the mastery of facts, and into areas such as: how masterfully can students apply knowledge/skills? How thoughtful are their stated perspectives? How empathetic and open are they? What level of maturity do they exhibit in their self-knowledge? Again, these categories are likely not equally important – and may not even be of interest – for all types of assessments, but it is important to acknowledge that rubrics are one way that affective and/or behavioral criteria can be assessed.

Once you have written a set of performance criteria and rubrics, evaluation and revision is a continuing process. One strong resource for evaluating criteria and rubrics, as well as other performance-related tasks and assessments, is the “Assessment Design Rating Checklist” given by Wiggins in *Educative Assessment*.^[18] Presented as two pages of checklist questions, this tool can help you self-evaluate (or have peers rate) how credible, instructionally worthy, user-friendly, and well-designed your criteria and rubrics are.

Practical Tips

The most important, no-nonsense, practical tips we can offer for people who would like to minimize the amount of criteria/rubric revisions are:

1. Avoid using the word “and” in your criteria and rubrics unless you really mean it. If a submission must do something, and something else, and something else, and something else, to meet a criterion, then strictly speaking the lack of only one of these things makes the submission fail overall. Use “and/or”, or “or” – or even better, separate compounded items into multiple categories or multiple criteria/rubrics.
2. Wherever possible, use specific language when writing rubrics. If a rubric states that to meet a criterion, a page of text must contain “minimal errors” – does “minimal” mean one error? Five errors? Does ‘error’ refer solely to grammatical errors, or also to formatting errors, or errors in the content/thesis of the text? The more specific the rubric, the more consistently it can be applied by multiple people.
3. Write relatively general performance criteria. This will give you the option of changing the way the criteria are assessed (*i.e.*, changing the specific demonstration/assignment examined, or changing the course from which samples of student work are kept) without having to change everything.
4. Build consensus for the criteria and rubrics within the faculty; seek feedback on them from other constituencies (students, advisory boards, *etc.*); and use them for program improvement in clearly-communicated ways. Having participated in setting performance goals, and seeing evidence that assessment results are used in beneficial ways, faculty and students will be far more willing to dedicate time and effort to participating fully.
5. Focus on judging the impact or effect of a student performance rather than on the processes, formatting, or effort expended by the students^[19], if your goal is to directly assess achievement of outcomes.

Troubleshooting: Inter-Rater Reliability

Every assessment strategy has inherent potential pitfalls, and low inter-rater reliability is one potential problem associated with having different people review and rate samples of student work. ***The first step to improve inter-rater reliability is to create rubrics that identify specific and demonstrable aspects of performance, supplemented with descriptive examples of levels of performance. Calibration, discussion, and consensus are the next three steps toward improving inter-rater reliability.*** Before beginning a full set of reviews, ask raters to discuss, as a group, a small set of samples (perhaps saved from past years, with the past ratings revealed after some discussion). Ask the raters to come to consensus on the samples as a “calibration.” Then, have each submission reviewed by two or three raters. This can be accomplished either by having them work together in real-time resolution, or by having them work individually and then meet to compare ratings and discuss/resolve any discrepancies. ***The next step toward improving inter-rater reliability is to treat issues brought up during calibration, discussion and consensus interactions as feedback on the rubrics and criteria – and to revise the criteria and rubrics accordingly for the next cycle of assessment.*** Revision is a critical aspect of rubric efficacy, and revision to improve consistency is an important part of rubric development^[20]. Finally, if raters absolutely cannot come to consensus, have a contingency plan developed in advance through which other raters will join the discussion. If further mediation is needed, the discussion

should probably shift from determining individual scores to how to improve the criteria, rubrics, and/or overall rating process. Ensuring that raters have opportunities to calibrate, discuss, and come to consensus may increase the total time spent on the assessment – but will yield information important to the revision of the criteria and rubrics, and will dramatically improve the prospect of obtaining quality, meaningful data that can be used with confidence.

Troubleshooting: Faculty Resistance

When considering a performance criteria/rubric assessment strategy, the complaint or resistive comment most likely to be heard from faculty is something like “*Aren’t we just ‘teaching to the test’? Aren’t we just ‘gaming the system’ here?*” These questions could probably be asked regarding any system of assessment in which the results are used as a basis for changes, especially when these changes might include altered methods of assessment. Regarding the use of criteria/rubrics, this type of comment is most likely to come from faculty who are unfamiliar or uncomfortable with articulating and openly communicating academic performance standards. These may be faculty who generally use normative-based rather than criterion-based grading systems (*i.e.*, who curve grades rather than grading on achievement of learning objectives), who grade on general criteria but who have not worked on clearly articulating demonstrable learning objectives for their courses, or who have chosen not to share articulated course learning objectives with their students. Instead of ignoring or dismissing such comments, consider some of the possible responses presented below. The resulting discussion may prove informative and useful.

- What part of this review system are you concerned about? The fact that we are setting the performance standards for our own students, and therefore for our own program? The fact that we are telling our students what they need to do in order to meet the criteria and/or earn a good grade on an assignment?
- Who, if not us, should be setting standards for our program? Are there other resources you’re aware of that we ought to be using?
- Why is telling students what they should be learning ‘gaming the system’? How will they know what we expect of them if we don’t communicate our expectations?
- Do you believe that the criteria/rubrics we are discussing are things that students are not or should not be doing in your class? Why? Do you believe these skills are important enough to assess on a program level? Do they correlate with our Program Outcomes? Is there another place in the curriculum that these skills should be demonstrated instead? Should we revisit our Outcomes?

Conclusions

We believe the advantages (clear standards and feedback, facilitating communication, community-building) of using performance criteria and rubrics as the core of a program-level assessment approach outweigh the disadvantages of this approach, which are essentially the time commitments required to determine the rubrics/criteria, review the submitted evidence, and determine appropriate responses to the results. Our performance-centered direct assessment strategy is coordinated across a program as a whole; provides feedback that is informative as

well as easily organized and interpreted; and facilitates reflection and improvements on multiple levels – from specific, focused areas of the program to a broad, holistic overview of the program. We present this paper with the hope that faculty from other programs may find the examples, tips, and suggestions useful as they work through their own program assessment strategies.

Acknowledgements

The case-study performance criteria and rubrics described in this paper were developed by the Rose-Hulman Institute of Technology Biomedical Engineering ABET Working Group, which at the time consisted of Jameel Ahmed, Christine Buckley, Kay C Dee, Glen Livesay, Renee Rogge, and Bill Weiner. I deeply appreciate the teamwork, resilience, and support exhibited by this group of faculty.

Bibliography

- [1] Engineering Accreditation Criteria. Published by ABET. Accessible online as of 01/07/06 at <http://www.abet.org/forms.shtml>
- [2] “Guidelines to Institutions, Team Chairs and Program Evaluators on Interpreting and Meeting the Standards Set Forth in Criterion 3 of the Engineering Accreditation Criteria,” 2004, page 2. Published by ABET. Accessible as of 10/03/05 as “Whitepaper – Guidelines for Criterion 3,” at <http://www.abet.org/Linked Documents-UPDATE/Program Docs/EAC Guidelines for Criterion3.pdf>
- [3] Allen, MJ. *Assessing Academic Programs in Higher Education*. Anker Publishing Company, Inc., Bolton, MS, 2004. See especially pages 99-102 and 127-129 for summary tables.
- [4] Rogers, G. and Williams J. “Asynchronous Assessment: Using Electronic Portfolios to Assess Student Outcomes,” *Proceedings of the 1999 American Society for Engineering Education Annual Conference & Exposition*.
- [5] Rogers, G. and Williams, J. “Building a Better Portfolio,” *ASEE Prism*, 30-32, Jan 1999.
- [6] Williams, J. “The Engineering Portfolio: Communication, Reflection, and Student Learning Outcomes Assessment,” *International Journal of Engineering Education* 18(2):199-207, 2002.
- [7] Holzberg, C.S. “Designing Rubrics – taking advantage of online resources,” *Technology & Learning*, 26(3):36-37, 2005.
- [8] Accessible online as of 02/23/06 at <http://rubistar.4teachers.org/index.php>
- [9] Accessible online as of 02/23/06 at <http://www.rubrician.com/>
- [10] Blanchard, S.M., McCord, M.G., Mente, P.L., Lalush, D.S., Abrams, C.F., Lobo E.G. and Nagle, H.T. “Rubrics Cubed: Tying Grades to Assessment to Reduce Faculty Workloads.” *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition*.
- [11] Gassert, J.D. and Milkowski, L. “Using Rubrics to Evaluate Engineering Design and to Assess Program Outcomes.” *Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition*.
- [12] Wiggins, G. and McTighe, J. *Understanding By Design*. Association for Supervision and Curriculum Development, Alexandria, VA, 1998.
- [13] Wiggins and McTighe, page 18.
- [14] Wiggins, G. *Educative Assessment: designing assessments to inform and improve student performance*. Jossey-Bass Inc., San Francisco, CA, 1998.
- [15] Wiggins, page 130.
- [16] Anderson, L.W., Krathwohl D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. and Wittrock, M.C., eds. *A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom’s taxonomy of educational objectives*. Addison Wesley Longman, Inc., 2001.
- [17] Wiggins and McTighe, pages 76 and 77.
- [18] Wiggins, pages 151 and 152.

[19] Wiggins, page 184.

[20] Hanson, J. "Rubrics: Helping you and students perform better," Assessment and Learning Forum presented 01/24/06 at Rose-Hulman Institute of Technology, Terre Haute, IN. Slides, video and resources accessible online as of 02/23/06 at <http://www.rose-hulman.edu/irpa/TLForum/index.html>