

Using Rubrics for Course Assignments

Cy Leise¹, and Mohamed El-Sayed²

Abstract

Measurement of learning and growth is increasingly valued in higher education because of the focus on continuous improvement at all levels of the system. Well-constructed course rubrics can support the faculty goal of establishing clear standards for assessing and grading while also helping to make assignment and activity instructions more “transparent” for learners. This paper overviews different types of rubrics and analyzes a methodology for rapid design of course tools. Examples are used to illustrate how to connect course assignments and rubrics and also how to use measurement theory to improve course implementation of rubrics.

Introduction

Arguably, the most important task in designing any educational program, course, or learning activity is to select or prepare measurement tools that are of high enough quality to guide the processes of assessment and evaluation (Burke & Bargainnier, 2007). Any doubt by learners about what they are being instructed to produce creates ambiguity and reduces the value of data for improving the course. This causes the measurement process to lose its reliability and validity. Rubrics expand an educator’s measurement range beyond tests and exams, which work well for conceptual knowledge to the higher levels that are essential for learning and growth in any discipline or profession. To address the use of rubrics in course assignments this paper aims to provide:

- An overview of the types and purposes of rubrics.
- A process for designing rubrics for course use.
- Description of the measurement objectives in typical courses in coordination with learning/growth opportunities for all skill domains.
- An introduction to the use of measurement theory and criteria as a basis for insights about how to improve rubrics.
- Insights about rubric applications that will guide effective usage.

Rubrics and Quality Learning Environments

Every learning opportunity is a combination of one or more learning skills (Apple, Beyerlein, Leise, & Baehr, 2007) that are important for use or creation of knowledge relevant to a task, role, or purpose. A quality learning environment (Smith & Apple, 2007) is a “system” that must include sound curriculum design, effective facilitation, and valid measurement. Even in situations that require quick solutions, e.g., a new course

assigned to an educator at the last minute, it is possible to maintain quality if measurement tools match up with the main learning opportunities planned.

Rubrics are flexible tools that can be developed relatively easily in practice situations to capture the key performance criteria and standards expected. Assignments may need updating if they are inconsistent with the criteria and standards of “customized” rubrics, i.e., creating a rubric along with an assignments or activity provides a quality check. Whenever possible, of course, educators should use rubrics developed by a design team using an established methodology; this increases assurance that generalized criteria and standards of performance are articulated, which increases the quality of measurement data. Learners who can consistently meet generalized standards of performance are more likely to demonstrate accelerated growth across multiple contexts because of greater consistency and validity of assessment and evaluation from self, peers, and educators.

Rubrics as Measurement Solutions

A substantial literature in program assessment and measurement exists to support the use of rubrics and other performance-based measures. Linn, Baker, and Dunbar (1991) argued that discrepancies between indicators, e.g., tests, and goals, e.g., learning objectives, produce distortion in the overall results of educational assessment. “Authentic assessment” requires that efficiency, reliability, and comparability of measures be carefully planned and delivered. Miller and Linn (2000) analyzed the need for a focus on measuring what teachers actually teach and observe, and found, among other factors, that task variability is an important source of error. Without multiple representative tasks, it is difficult to achieve reliability. Many educational standards include processes for which scoring rubrics are the appropriate type of measure; Miller and Linn review established methods for researching the reliability and validity of such measures.

¹ Department of Psychology and Human Services, Bellevue University

² Department of Mechanical Engineering, Kettering University

Messick (1995) distinguished six validity criteria for educational measures: content, substantive, structure, generalizability, external, and consequential. Measures that meet these validity standards meet both scientific and social-consequences expectations. Clearly, validity research provides a “big picture” perspective on what educators need to work on in order to advance the quality of educational processes and outcomes. The purpose of noting these larger issues is to create a sense of context regarding scholarship and research. Educators can be part of this larger conversation to the extent that they gather data on the quality of their course measures.

Types and Purposes of Rubrics

Rubrics vary in type and purpose (Bargainnier, 2007) and must be selected to fit one’s measurement purpose. Task-specific rubrics with an analytic focus are perhaps the most common type because they are customized for specific assignments or tasks, e.g., a course or activity assignment. Analytical rubrics help with establishing the basic learning skills needed for a type of performance, e.g., writing a news item for a school newspaper. The content for a news article must be assessed, but learning and growth occur mainly in terms of the basic information processing, communication, and analysis (critical thinking) skills used by journalists at any level of professionalism. An analytical rubric can represent an “expert” perspective in the sense of providing criteria and standards that are ideal for the type of assignment while also providing specifics to use in assessing and grading (evaluating) selected products submitted by students during a course.

In many courses and programs, learners are required to demonstrate the quality of a general type of performance, e.g., watercolor painting or software development, by preparing portfolios; a holistic rubric may be more practical, and valid, for capturing the overall quality of a portfolio than repeated application of several analytic rubrics. Table 1 includes examples of uses of rubrics from the analytic and holistic perspectives for both specific tasks and generic skills.

Use and Adaptation of Rubrics

Although hundreds of rubrics can be found by searching the internet and publications, most will not match up well with what an educator has in mind. If there are no copyright restrictions, or permission can be obtained, it may be possible to revise a published rubric. Often, however, it may be less work to create an original rubric by identifying the performance criteria and standards (which are set for evaluation but useful also for assessment) that match the assignment. Any assignment designed with one or two key learning skills in mind will have a predictable

pattern (Apple, Beyerlein, Leise, & Baehr, 2007). For example, to return to the journalism example, review of the information processing “skill cluster” (from the cognitive domain) and the communicating skill cluster (from the social domain) will result in identification of specific learning skills that will need to be demonstrated and assessed if a student is performing “to standard” in writing news articles for the school newspaper. The standards obviously must be set realistically to match beginner performance capabilities as well as higher competency levels that may not be achieved until much later. A holistic rubric for journalism could add to the overall validity of learners’ writing performance by representing quality levels based on important qualities and dimensions of this field of writing. Examples of all types of rubrics are available in the *Faculty Guidebook* (Beyerlein, Apple, and Holmes 2007).

Growth in Measurement Practices

Professional motives change throughout educators’ careers from an earlier focus on discipline knowledge and specific achievements required for advancement, to increasing awareness of the performance potential that could result if educational systems and practices could be improved.

Table 2 is based on lead author impressions from personal experience and from many discussions with peers, of some typical professional growth challenges that educators must meet in regard to measurement of learning, growth, and program development. An assumption is that, over time, professional development of educators should be correlated with increasingly stronger forms of measurement (Myrvaagnes, 2007, Racine, 2007, and Smith & Apple, 2007). Comments below provide reflective insights about the table content.

1. While one is learning the content knowledge of a discipline or practice it is difficult to maintain a clear focus on higher-order learning outcomes, e.g., as described in Bloom’s taxonomy for cognitive learning objectives (Bobrowski, 2007). During this professional development phase, the evaluation of knowledge and theory by means of written or multiple-choice tests is well-supported by textbook publishers and is highly acceptable to students and administrators.
2. Educators who have integrated the content knowledge and the skills associated with a discipline, even if they do not yet consider themselves experts, are likely to become interested in facilitation of student learning at an application level. They have concluded that no matter how well learners know facts, principles,

Table 1 Examples of Uses of Rubrics by Type

| | Analytic | Holistic |
|----------------------|---|---|
| Task-specific | Accuracy of validation of math solutions in a specific course | Typical level of math problem validation skills of a math major |
| Generic | Writing quality of research reports | Typical writing quality regardless of task or context |

Table 2 Growth in Measurement Practices

| Professional Focus | Practice Description | Recommendation |
|---|--|---|
| First year; discipline focus | Tests and exams are well-supported by publishers and accepted by most students | Explore use of one rubric per course |
| Second year; learning focus | Revise selected assignments to learning above the Bloom comprehension level | Assess own activity facilitation with a self-assessment rubric |
| Improvement of learning standards | Revise selected courses by using a course design methodology | Add learning activities with clear objectives, criteria, and task-specific rubrics |
| Increasing quality of performance assessment and evaluation | Revise the program or major with emphasis on key integrated growth outcomes | Include analytic rubrics that represent all major dimensions of integrated performances |
| Implementing quality learning environment assessment | Assess current system and practices against the criteria and standards of a quality learning environment | Use diverse measures including quality assurance monitoring of the system |
| Improving program assessment | Assess major outcomes of a program or major | Develop a quality measurement table for the program |
| Implementing continuous quality improvement of the learning environment | Use quality assurance methods to identify areas to improve | Continuously improve quality based on data from measures in the quality measurement table |

and theories, there is no substitute for using these in applied situations such as labs, projects, research, or service. This creates a need for customized measures such as rubrics because tests are ill-suited for application and other higher-order learning.

- Initial experiences with rubrics may involve using, with only slight adaptations, the tools offered by others. Use of new measures tends to focus attention back to course learning objectives and to the design of activities that provide appropriately-structured learning opportunities relevant to specific objectives. Extensive experience with using rubrics to score a performance results in assessment of the tool itself. However, lack of criteria and standards for what a rubric should do may restrict efforts to make a rubric better until an educator learns more about quality educational processes in general and about measurement design in particular.
- Recognition that quality learning environments must be developed with guidance from methodologies for program, course, and activity design initiates a new professional development phase for an educator; careful design always requires measurement of performance. Programs, courses, and activities must be integrated; measurement choices are among the most obvious indicators of how such integration has been conceptualized.

Methodology for Designing Rubrics

The principles from Burke & Bargainnier (2007) overview of measurement provide a basis for assessing how well a rubric is likely to work for a given purpose. They summarized the foundation principles of measurement for educators, whether the measurement target is knowledge, performance, a product, or an organizational entity such as an academic department. The principles are paraphrased here.

- Measure what is important.
- Use observable data as the frame of reference.
- Keep the focus well-defined—not too large or too small.
- Select an appropriate measurement tool.
- If possible, obtain data from multiple sources.
- Analyze outliers (very high or very low scores or ratings).
- Test reliability before using a measure.
- Compare and contrast validity under varying conditions.
- Assess the cost/benefit of using a measure.

The abbreviated methodology for designing rubric measures presented here is intended to provide guidance for initial understanding; in the next section a sample rubric for assessing and evaluating student research proposals provides an illustration of the steps in this methodology.

1. Create an assignment or activity that is planned within a course to achieve certain learning objectives and to “grow” one, or at most two, learning skills.
2. After piloting the assignment oneself and involving a few colleagues and students, revise for internal consistency, clarity, and focus.
3. Identify the key criteria that students must attend to in order to demonstrate performance. For initial rubric design, it often works reasonably well to identify the main sections or features of the assignment or activity as a way to establish criterion areas.
4. Establish “anchors” for rating scales that will allow quick assessment or evaluation of each feature. If there are multiple elements to a criterion, it helps to separate these by using short phrases versus a lengthy paragraph.
5. Although the goal is to keep the focus on assessment, students will also need score information to help understand where they stand in terms of evaluation (course grading). Giving a minimum total score for satisfactory performance or approximate scores for the traditional letter grades may be useful if accompanied by extensive work on self-assessment.
6. Use the SII Assessment Technique (Wasserman & Beyerlein, 2007) as an open-ended way to provide feedback meant to be for assessment. Use a score only when the intent is to provide evaluation.

Example of a Course Rubric

Table 3 provides an example of a rubric produced by the methodology described above that is used in a research course taught by the lead author in a Master’s of Human Services program designed to prepare mental health practitioners.

1. The integrated outcome for the research course is a proposal that should demonstrate knowledge of research design and methodology plus ability to write an appropriate literature review. Minimal emphasis is placed on data collection or analysis and interpretation of data. Two main learning skills include “filtering” information to match specific proposal requirements (from the “organizing data” cluster under Processing Information in the Cognitive Domain in the *Faculty Guidebook* (Beyerlein, Apple, and Holmes 2007)) and designing a research method for testing a hypothesis (from the “obtaining evidence” cluster under Conducting Research, also in the Cognitive Domain).
2. The assignment was improved over several years by the lead author and peers; earlier versions of the rubric also helped to add clarity to the assignment and how students are prepared with preliminary assignments and activities earlier in the course.
3. The left-most column in the rubric identifies the criteria considered of most importance for assessing or evaluating the quality of the drafts and final versions of proposals. Some writing criteria, e.g., orienting readers, have been found to be helpful to guide the perspective used to write a proposal that will communicate an “empirical” approach to knowledge.
4. Four “anchors” or scale values were selected in the example but this is relatively arbitrary depending on the importance of the quality distinctions and the total points for the final performance. One significant innovation is division of the descriptors within a row (criterion area) into two or more elements separated by font changes. This avoids an overly complex information load while also allowing for full representation of what is intended for each criterion quality level.
5. A rubric can serve both assessment and evaluation needs but careful facilitation of “buy-in” of students for the assessment/evaluation distinction is essential. Learner experiences with assessment earlier in a course or program must have convinced them that measures exist to increase objectivity and transparency about quality of current performance and that the goal is personal performance improvement. Without preparation learners tend to assume that evaluation

Table 3 Research Proposal Rubric (Human Services)

| Criterion Area "Quality of..." | Improvement Needed (0 pts) | Minimum Performance (1 pt) | Medium Performance (2 pts) | Excellent Performance (3 pts) |
|--|---|---|--|--|
| 1. Reader Orientation | Title focus unclear | Title focus needs work | Title focus fairly clear | Title focus very clear |
| | Variables not noted | Variables unclear in title | Variables noted indirectly but accurately in title | Variables directly noted in title |
| | First paragraph(s) on page 3 did not orient readers to topic and Hypothesis | First paragraph(s) unclear about topic and hypothesis | First paragraph(s) on page 3 fairly clear for readers about topic and hypothesis | First paragraph(s) on page 3 introduce(s) topic and hypothesis |
| 2. Abstract | Abstract not present | Abstract too general or lacks operational detail | Abstract fairly balanced and operationally oriented | Abstract balanced and operationally oriented |
| | | Hypothesis not central enough | Hypothesis kept fairly central | Hypothesis kept central |
| | | Substantial APA format errors | Minor APA format errors | In APA format |
| 3. Literature Review | Citations all low quality | Citations not of high quality | Citations of fairly high quality | Citations of high quality |
| | Paraphrasing not used | Paraphrasing not used well | Paraphrasing fairly accurate | Paraphrasing accurate |
| | Citation selection incoherent | Citations selected for topic, not purpose | Citations sometimes redundant | Citations met logical needs |
| 4. Hypothesis | Hypothesis not stated | Hypothesis topical | Hypothesis partially operational | Hypothesis operational |
| | | Hypothesis unrealistic | Hypothesis global | Hypothesis realistic |
| | | Hypothesis ambiguous | Hypothesis not a future tense statement | Hypothesis one future tense statement |
| 5. Definition of Variables | Ind. & dep. vars not described | Ind. & dep. vars defined qualitatively | Ind. & dep. vars partially operationally defined | Ind. & dep. vars operationally defined |
| | Ind./dep. vars incorrectly labeled | Ind/dep. vars given unclear labels | Ind/dep. vars given fairly clear labels | Ind/dep. vars given clear labels |
| 6. Design | No design specified | Design ambiguously labeled | Design fairly accurately labeled | Design accurately labeled |
| | | Design possible but weak for hypothesis | Design fairly good fit for hypothesis | Design best fit for hypothesis |
| | | Design addresses few validity issues | Design addresses some validity issues | Design addresses main validity issues |
| 7. Method Section: Participants | Participant pool incompletely described or missing | Mentioned sample but left details unclear | Participant pool fairly well described | Participant pool clearly identified |
| | | Assignment (to group or level) procedure unclear | Assignment (to group or level) procedure left slightly unclear | Assignment (to group or level) procedure specified |

(Table 3 continued)

| Criterion Area "Quality of..." | Improvement Needed (0 pts) | Minimum Performance (1 pt) | Medium Performance (2 pts) | Excellent Performance (3 pts) |
|--|---|--|--|--|
| 8. Method Section: Measure(s) (Dep. Vars) | Measure(s) not identified | Measure(s) described but not named | Measure(s) described too informally | Measure(s) identified by title |
| | | Measures incomplete or poor fit for hypothesis | Measures fairly sound fit for operational hypothesis | Measures fit operational hypothesis |
| | | Reliability & validity not noted | Reliability & validity incompletely noted | Reliability & validity specified |
| 9. Method Section: Procedures | Procedures discussed only briefly or abstractly | Unclear data collection procedures | Fairly clear data collection procedures | Clear data collection procedures |
| | | Incomplete materials, forms, & resources | Missing some materials, forms, & resources | Materials, forms, & resources specified |
| 10. Results & Conclusion Sections | Results presented in past tense or as actual data | Results not well-matched to procedures | Results overly elaborated for proposal | Results presented as one-sentence empirical estimate |
| | | Conclusion not clearly related to hypothesis | Conclusion fairly well related to hypothesis | Conclusion related to hypothesis |
| 11. Appendix | No appendix | Did not include specified documents | Included fairly well-designed supporting documents | Included well-designed supporting documents |
| | | Did not follow format instructions | Appendix format somewhat inconsistent | Appendix format follows APA style or instructions |
| 12. Use of APA style | Well below expectations in use of APA standards | Many grammar errors, etc. | Some grammar errors, etc. | Correct grammar, etc. |
| | | Two format errors in title page | One format error in title page | Correct format of title page |
| | | Used some headers but not logical | Used headers but one or more not appropriate | Appropriate headers |
| | | Substantial errors in citations & references | Some format errors in citations & references | Correct format of citations and references |
| | | Substantial organizational issues | Some organizational inconsistencies | Properly organized and balanced throughout |
| Total Points/Grade: A+ (36); A (33-35); A- (30-32); B+ (27-29); B (24-26); B- (21-23); C+ (18-20); C (15-17) | | | | |

Table 4 Selected Measurement Qualities Related to Rubrics

| Quality Dimension | Criteria/Standards | Recommendations |
|---|---|---|
| Ease of Creation | <ul style="list-style-type: none"> Time and effort requirements | Adapt an available rubric to match a well-written assignment or activity |
| Ease of Use | <ul style="list-style-type: none"> Clarity of presentation Judgment expertise requirements | Restrict the rubric criteria to specific elements provided in instructions; use simple phrases for each element within a criterion. |
| Reliability | <ul style="list-style-type: none"> Inter-rater agreement | Involve other educators in using the rubric for similar tasks |
| Validity | <ul style="list-style-type: none"> Keyed to learning process Authenticity | Focus on a key learning skill while also reflecting organization/format; assess realistic need in course for performances assessed |
| Generalizability | <ul style="list-style-type: none"> Usefulness across contexts | Focus on learning skills useful across courses, disciplines, and situations |
| Support of quality learning environment | <ul style="list-style-type: none"> Relation to learning objectives Usefulness for self-assessment | Design some rubrics for educator self-assessment of quality performance |
| Program assessment value | <ul style="list-style-type: none"> Connection to course and program outcomes | Design tasks, and related rubrics, to support learning and growth related to program outcomes |

is the only process supported by a measure and will view any but the highest scores as failure. Letter grade equivalents for total scores are included in the example but this can and does create learner concerns that must be addressed with clear communication.

- Any well-designed rubric includes many indicators of performance that can be useful as a resource for selecting assessment targets when using the SII technique (Wasserman & Beyerlein, 2007). Learners must assess actual performances in order to grow in competency.

Improving Rubrics

Table 4 includes selected measurement quality dimensions of rubrics with suggested criteria, standards, and recommendations for educators. Some of these are practical matters while others are based on scientific validity criteria, e.g., those recommended above from Messick (1995). Linn et al. (2000) provide specific guidance on these and more features that can be strengthened. Although the steps in the section above provide some guidance for getting started with designing rubrics, it is essential to keep in mind that the scientific improvement of all measures is an ongoing challenge.

Conclusions

Quality learning environments must be developed with guidance from methodologies for program, course, and activity design. Rubrics can easily be developed to capture the key performance criteria and standards

expected. Simpler rubrics are highly valuable because even a lower-quality measure will provide much better evidence about the nature of learning than estimation on the basis of “experience.” Like any tool, however, one will become more expert only by direct work with creating and applying of rubrics in real contexts.

Rubrics must be selected to fit one’s measurement purpose so the distinctions between the types of rubrics are quite important. Holistic rubrics should be used only for general assessment of an individual’s performance quality across many tasks within a skill domain, e.g., see the holistic as well as the analytic rubrics for writing and mechanical design published in the *Faculty Guidebook* (Beyerlein, Apple, and Holmes 2007). Analytical rubrics are closer to the working conditions that educators deal with on a day-to-day basis. The relationship between these important types is that holistic “levels” are quite useful for established the rating “anchors” used on analytical rubrics. The relationship between these is that holistic “levels” are quite useful for establishing the rating “anchors” used in analytical rubrics. Holistic rubrics keep the most important challenge in front of the educator, which is how to facilitate growth in learners.

Educators must continuously improve their uses of measurement in order to facilitate quality learning and growth. The information and recommendations presented are intended to encourage educators with practical methods and examples that will lead to creative applications of rubrics as flexible solutions to measurement of higher-order learning across all learning domains.

References

- Apple, D., Beyerlein, S., Leise, C., & Baehr, M., (2007). Classification of Learning Skills, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest. 201-204.
- Bargainnier, S. (2007). Fundamental of Rubrics, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest, 75-78.
- Beyerlein, S., Apple, D., & Holmes, C., (Eds). (2007). *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest.
- Bobrowski, P., (2007). Bloom's Taxonomy-Expanding its Meaning, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest.161-164.
- Burke, K., & Bargainnier, S. (2007). Overview of Measurement, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest, 71-74.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20, 15-21.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning, *American Psychologist*, 50, 741-749.
- Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments, *Applied Psychological Measurement*, 24, 367-378.
- Myrvaagnes, E., (2007). Performance Levels for Learners and Self-Growers, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest. 87-90.
- Racine, M., (2007). Constructing a Table of Measures, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest. 125-128.
- Smith, P., & Apple, D., (2007). Overview of Quality Learning Environments, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest. 311-314.
- Wasserman, J., & Beyerlein, S., (2007). SII Method for Assessment Reporting, *Faculty Guidebook: A Comprehensive Tool for Improving Faculty Performance* (4th ed.). Lisle, IL: Pacific Crest. 465-470.