The 5^{th} Iran International Conference on Engineering Education
(with Focus on Internationalization of Engineering Education)

Nov. 21-23, 2017, K. N. Toosi University of Technology, Tehran, Iran.

Paper Code: IICEE2017-12

Utilizing Outcomes Based Analytic Rubric in Engineering Classroom

Bahar Memarian

Abstract – This article aims to promote the use of analytic rubrics and provide guidelines for benchmarking from the rubric bank developed by the University of Torontos Faculty of Applied Science and Engineering as part of the Higher Education Quality Council of Ontario (HEQCO) Learning Outcomes Assessment Consortium. The Development of Analytic Rubrics for Competency Assessment project (herein called DARCA) sought to develop valid universal rubrics that detail the expected learning outcomes for five Canadian Engineering Accreditation Board (CEAB) graduate attributes namely Investigation, Problem Analysis (used interchangeably with Problem Solving), Teamwork, Communication, and Design. The analytic rubric bank content has been carefully refined through testing with graduate students and instructors. We encourage instructors and faculty to choose performance criteria from this work that are best suited to the learning outcomes, learning activity, and learning environment.

keywords

Rubrics,Criterion Based Assessment, Learning Outcomes Assessment

Bahar Memarian Department of Mechanical and Industrial Engineering, University of Toronto, Canada

1 What are analytic rubrics?

Rubrics are used as tools in the instructional design process to better assess student performance in order to provide timely feedback and effective direction on student learning. As outlined in literature [1] rubrics: (1) provide a task description; (2) develop a scale for evaluation; (3) develop dimensions of the task necessary for success; and (4) provide a description of the dimensions relative to the scale being used. Rubrics could be used for diagnostic, summative and formative purposes. Rubrics that have two or more separate scales and criteria for evaluation are called analytic where as single scale rubrics are called holistic. The multiple scale feature of analytic rubrics in turn allows more customizability and specificity in identifying students learning outcomes as compared to holistic rubrics which give a broader and overall perspective on student performance.

Prior to the design and benchmarking of analytic rubrics, instructional designers or assessors first need to identify and address the desired learning outcomes of their course. Learning outcomes seek to find out: what outcomes will this learning have for students? As learning outcomes are built upon Blooms revised taxonomy of learning or similar taxonomies, the objectives are outlined as a statement

containing a verb and a noun with the verb identifying the intended cognitive process and the noun specifying the knowledge students are expected to acquire or construct [2]. It would be best to aim for clear, specific, achievable, observable, and measurable learning outcomes when crafting our assessment tools [3].

Analytic rubrics are commonly developed as two-dimensional tables as shown in Figure 1. The table constitutes a framework to specify what we want students to learn on the left most column of the table (indicators) and an evaluation scale on the top most row of the table (achievement or performance levels). The performance of the student based on the achievement levels and indicators is then described in each cell within the table (descriptors). Descriptors could be further broken down into sub-descriptors specifying alternative scenarios for meeting a certain achievement level of student learning outcome. When choosing achievement levels, two factors need

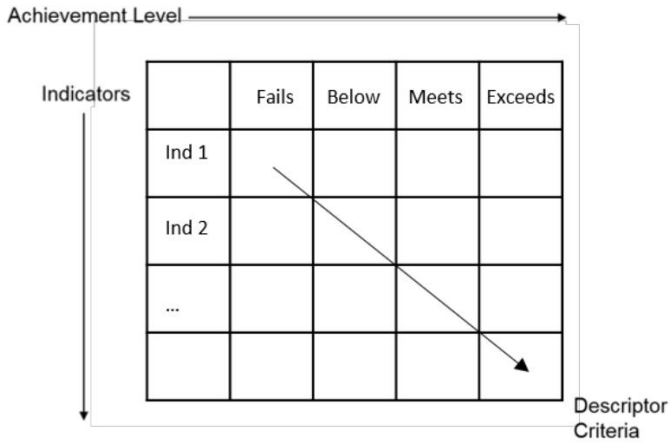

Figure 1: Organization of analytic rubrics

to be taken into account: the nature and number of levels. Achievement levels could be presented in a formative manner using words such Fails-below-meets-exceeds or in a summative manner having some numerical attribution tied to it such as D-C-B-A or 0-25, 25-50, 50-75, and 75-100 percent. The nature (summative vs. formative) and scale (3 vs. 4 vs. 5 level scales) of achievement level are often selected based on preference of the instructor and learning activity. Learning activities include solving problems, carrying out experiments, participating in discussions, engaging in direct observation, and conducting research [3]. In the DARCA project for example rubrics for competencies such as Communication, Teamwork, and Design have been assigned four level achievement criteria, breaking down fail into two parts. What these rubrics aim to achieve by partitioning failure into two categories is to distinguish between work that exhibits no effort at all (not demonstrated) and work that demonstrates a fundamental misunderstanding of the relevant concepts.

2 Why use analytic rubrics?

Literature suggests the use of rubrics brings quality and sheds light on teaching, learning, and assessment in a number of different ways. To better picture the importance and need for the assessment tools such as analytic rubrics one could consider the desired learning outcomes, lessons, learning activities, and evaluation or scoring each as puzzle pieces. Without a sound assessment tool the puzzle pieces may be piled in one place or be put together incorrectly and provide an inaccurate view of teaching and learning. The use of analytic rubrics therefore enhances the educational experience through [4]:

- Promotion of learning and student judgement: Promoting continuous learning by providing a roadmap of what steps students need to take now to be successful in the future through means of formative, self-, and peer assessment.
- Higher credibility: Increasing the consistency of judgement when assessing performance and authentic tasks across students, assignments, as well as between different raters.
- Teaching clarity and outcomes based evaluation: Quantifying desired student learning outcomes into measurable means as well as classification, ranking,

and weighing competencies based on relevancy to what we aim to assess.

With exposing students to the instructor and institutions view of teaching and learning, analytic rubrics enable students to better self-assess and adjust their performance over the learning period.

3 Where to use analytic rubrics?

As we see the rising shift from traditional testing of knowledge towards assessment for learning [5], analytic rubrics are becoming more viable tools for appropriate assessment of student learning as they outline a framework and concept map to learning success rather than the traditional right or wrong approaches. Analytic rubrics are particularly effective in evaluating student performance when situated in authentic activities in which students are required to come up with original solutions to real world problems. While analytic rubrics have been mainly utilized for providing feedback after a learning activity, current technologies make it be plausible to design analytic rubrics that meet the following objectives:

1. Understanding each student learning style
   (a) Learning patterns, where the student looks for information
   (b) Learning habits, and examining times the student dedicates attention to learning - student learning data analytics would be beneficial
2. Examining student perception and understanding of topics, determining if students have appropriate mental model of the different concepts covered (for timely and effective guidance and feedback)
3. Student learning compared to peers (for creating better teams and diagnosing team issues)
4. Evaluating student compared to the class (to observe class standing)

4 When to use analytic rubrics?

Determining the effective times to use rubrics as a means of directing student learning in a formative manner may be a more challenging task than using rubrics as a summative grading tool after a learning activity. While we may be interested in examining specific discipline focused skills in different learning activities of our course, we may find there are some major themes and competency patterns that are part of and iterated over our different learning activities. For this reason, we suggest making use of rubrics such as the DARCA rubric bank that focus on graduate student attributes valued in accreditation and have a top-down approach on student learning outcomes. Overall, iterative feedback and reinforcing students to adjust their understanding through the use of analytic rubrics can result in enhancement of student learning.

5 How to use analytic rubrics?

To better design an analytic rubric that is valid and relevant to what we aim to assess and results in higher reliability of use among one or more scorers, the following questions first need to be answered (Figure 2):

1. What are the outcomes and applications of designed learning experiences? (Table 1) To make analytic rubrics, it is suggested to design indicators around methodologies, concepts and learning processes rather than formulae and numeric end results.
2. What groups of skills and learning outcomes are we expecting students to acquire through a learning experience?
   - Oral and soft skills such as communication and teamwork
   - Hands on and training skills such as procedural lab work or use of a software
   - Higher order thinking skills such as analysis,

evaluation, application, and designing original ideas as a solution to a real-world problem

3. What is the learning environment? Are students provided with appropriate and meaningful resources and deadlines?

4. Are we making sure the desired learning outcomes, learning activity, and learning environment are well aligned?

5. What is our approach to the delivery and supervision of evaluation processes and merging them into the instructional design?

Table 1: Traditional versus authentic activities

Learning experience	Pros	Cons
Authentic activities: Assignment, Lab Work, Portfolio, or Project, Open ended tests	Provoking higher thinking and acquisition of authentic engineering skills	Open ended with higher levels of uncertainty as exposed to real life problems, More difficult to grade
Traditional Timed & Text based activities: True or False, Solve a Closed-ended Problem, Test or Exam	Standardized, Easier to grade for large population classes	Mainly focused on remembering, restating, computing, and deducing factual knowledge

Like any other tool, analytic rubrics are best utilized when designed and used appropriately. An analytic rubric design that critically concerns and focuses on the content validity, structure, and a fair and consistent methodology for weighing and assessing student performance is considered an appropriate tool to guide and assess student learning outcomes. Whether the analytic rubric is used for high stakes or classroom assessment, best practices are ones that provide evidence on the alignment, validity and reliability of analytic rubrics. Sample indicators for the problem-solving analytic rubric are shown in Table 2.

5.1 Feedback

With institutional and accreditation expectations on one hand and meeting diverse set of students learning needs, instructors and assessors are under pressure when it comes to the assessment of students performance. As humans are prone to making errors in making decisions under deadlines, making use of outcome based and aligned assessment tools such as the DARCA rubric bank helps in reducing erroneous and inflated judgements while bridging the program/accreditation expectations and students needs.
Literature suggests meaningful assessment requires continuous reflection and revision. The provision and use of explicit and specific feedback enables students to review their cognitive process model and adjust their learning accordingly [3],[6]. Feedback is the key factor in revision and formative assessment as it identifies how students can engage in adjusting the discrepancies, misconceptions, and gaps in their learning. Even though instructors and students social, psychological, and cultural background and perceptions inevitably play a role in both delivering and receiving feedback and consequently its effective use, objective, topic specific and neutral feedback has been shown to improve student performance.

5.2 Alignment

Alignment, as the name suggests refers to the association and alliance of two or more ideas or objects. In the context of learning outcome assessment, alignment concerns the relevant context and homogeneous linkage between learning outcomes, learning experiences, and assessment tools. On a high level, alignment seeks consistency and connection between course level, program level, and institutional level learning outcomes to ensure students have acquired desired graduate attributes to meet institutional level expectation and accreditation goals. In engineering

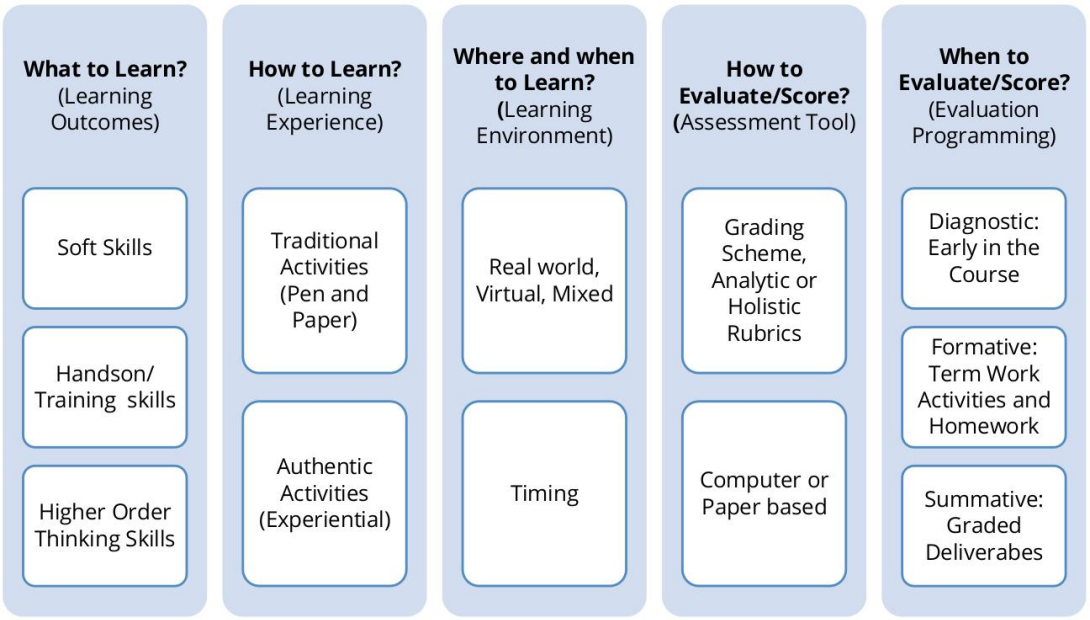

Figure 2: Considerations for design and employment of analytic rubrics

disciplines for example, by following CEAB requirements through design of outcomes based assessment strategies such as the DARCA rubric bank, instructional designers should expect to see a higher level of alignment compared to not using such strategies.

5.3 Reliability

Experimental evidence in rubric studies show that using rubrics can make assessments more reliable [4],[7],[8]. Since raters may each have their own method of grading students off a rubric and are prone to making grading errors, it is important to have a common strategy for making use of the analytic rubric to ensure reliability and uniformity in evaluating students work. The reliability of assessment increases as scores over different raters and occasions become more consistent [9]. Ensuring inter-rater reliability for large enrolment courses handled by several raters as well as intra-rater reliability for consistency in grading by one rater are some of the important factors when it comes to the topic of reliability. Inconsistency in evaluations could be due to the rater or raters attitudes regarding students ethnicity, as well as the content [10]. Other than social factors and perceptions, lack of systematic approaches in using analytic rubrics and lack of training could result in variability in assessment.

A number of statistical methodologies are used when measuring reliability among several raters for high stakes assessments. Using such statistical strategies could give instructors insight on their grading methods and ways in which they could revise the achievement scales for a more fair depiction of student performance. Table 3 lists such statistical methodologies and associated computational approaches. They could be grouped as:

- Consensus estimates: Measuring the degree to which markers give the same score to the same performance.
- Consistency estimates: Measuring the correlation of scores among raters.
- Measurement estimates: Measuring, for instance, the degree to which scores can be attributed to common scoring rather than to error components.

Table 2: University of Toronto Sample Indicators for Investigation Skills

02B Collect existing information	No information collected	Information collected covers few/no important concepts or relevant prior work in research area.	Information collected covers some important concepts and relevant prior work in research area.	Information collected covers most important concepts and relevant prior work in research area.	Information collected covers all important concepts and relevant prior work in research area.
		Information collected is not at all credible	Information collected is somewhat credible.	Information collected is mostly credible.	Information collected is entirely credible.
02C Perform a pedagogical review of a research paper or article	Review does not contextualize the article.	Review contextualizes the article incorrectly or superficially.	Review contextualizes the article ineffectively.	Review contextualizes the article.	Review contextualizes the article very effectively.
	No identification of: trends and patterns gaps in article major findings related to research questions	Incorrect identification of: trends and patterns gaps in article major findings related research questions	Minimal identification of: trends and patterns gaps in article learnings to research questions	Sufficient identification of: trends and patterns gaps in article learnings to research questions	Comprehensive identification of: trends and patterns gaps in article learnings to research questions
02D Perform a critical review of a research paper or article	No summary provided No key ideas captured	Summary is incorrect Incorrect ideas captured	Summary is marginally sufficient Some key ideas missing	Summary is complete Captures key ideas of the article	Summary demonstrates substantial depth of understanding All key ideas emphasized
	Criticism of key components not covered	Criticism of key components: incorrectly covered complete lack of insight	Criticism of key components: missing some key components are superficial	Criticism of key components: mostly covered shows sufficient insight	Criticism of key components: exceptionally well covered are insightful

5.4 Validity

In simple terms, validity questions and examines whether the assessment measures what it was intended to measure. It could be argued that validity of analytic rubrics is more important than reliability for fair evaluation of student work. This is because if we are measuring competencies different from what was taught, even if we have high reliability and consistency in grading among one or more

Table 3: University of Toronto Sample Indicators for Investigation Skills

Methodology:	**Approaches**:	**Assumptions**:	**Cons**:
Consensus estimates	1. Simple percent agreement form. Can be done with small sample, e.g. two judges 1. # of same rates/ Total rates 2. Cohens Kappa	Reasonable observers should be able to come to exact agreement about how to apply various levels of a scoring rubric.The percent agreement at the extreme ends of the rating scale is almost always lower than in the middle.	Including the adjacent scoring categories on the rating scale can lead to inflated estimates of inter-rater reliability if there are only a limited number of categories to choose from.
Consistency estimates	1. Pearson correlation coefficient 2. Spearmans rank coefficient 3. Cronbachs alpha coefficient	It is not really necessary for two instructors to share a common understanding of the rating scale.	Pearson and Spearmans rank coefficient can be calculated only for one pair of judges at a time and for one item at a time.
Measurement estimates	1. Factor analytic technique of principle components analysis 2. Generalization theory 3. Many-facets Rasch model	One should use all of the information available from all judges including discrepant ratings.	Would identify a rater who had responded randomly to the instrument (therefore scoring near the mean) or idiosyncratically to a few items

raters, we have still failed to appropriately capture student learning outcomes and thus our assessment is poorly constructed. Some important aspects of validity proposed in literature include but are not limited to [11]:

- Content: Does the content represent the knowledge and skills demonstrated on the assessment and it is not only limited to the sample of assessed tasks?
- Generalizability: Could the score interpretations be generalizable across groups, occasions, tasks, etc?
- External: Does the assessment score relate to other measures relevant to the construct being assessed? (i.e. correlations with other measures or instruments, relevance and utility of the rubric for its intended purpose)
- Structural: Does both the learning activity and the analytic rubric follow rationally from the domain structure? (i.e. factor analysis or raters evaluating the alignment of rubrics, tasks, and learning outcomes)
- Substantive: Does the level of cognitive processes match and is there consistency in responses that reflect the thinking processes used by experts in the field?
- Consequential: What are the implications of score

interpretation, both intended and unintended as well as short- and long-term consequences?

6 Discussion

As explained previously analytic rubrics have variations in structure (scale level, # of indicators, presence of sub-descriptors), subject (evaluating an executive summary versus a team presentation), and specialization field (mechanical engineering versus chemical engineering). Though thesevariations would in turn result in variations in the design of analytic rubrics, the essence of the assessment tools will remain the same. Some best practices when using the analytic rubrics such as the DARCA rubric bank as a benchmark are:

- Analytic rubrics should be utilized as 1) generic: detailing general criteria in a given performance, 2) task-specific: demonstrating use for particular assessment activity, 3) longitudinal: representing the progression in understanding of complex skills and interrelatedness of concepts. An effective rubric connects specific task requirements to overall performance goals, objectively discriminates among different degrees of proficiency, fits all potential performances in the rubric, sets out mutually exclusive indicators, and uses student anchors to set standards based on student artifacts [12].

- Instructors can select the number of learning outcomes from each category based on the weight allocated for each category and order indicators based on relevancy or chronology of steps. For example, for a course deliverable such as a team oral presentation, you would expect to see more weight and indicators in the analytic rubric to be allocated to communication and fewer for team work learning outcomes.

- Instructors can customize the rubric bank by adding or merging examples and concepts to indicators and descriptors that are important and need to be learned by the students in the course. This is to ensure instructors are designing and evaluating learning activities through a uniform analytic rubric structure meeting program level and institutional outcome goals while making sure what is being assessed is geared towards and centered around the topics in their course.

- Analytic rubrics are best utilized by students when followed by tangible and discipline specific learning activities and example scenarios.

- Analytic rubrics need to be followed by comments along with an overall grade to describe the areas students performance needs improvement for future evaluations.

- Analytic rubrics are more effective when they not only incorporate student learning outcomes but the trajectory of student learning and students improvements in learning over the course of the term.

- While the aim of achievement level is to classify and evaluate students on an individual level, it is important to keep in mind the students performance relative to each other and the overall class population performance when using rubrics. It is plausible to go through student work prior to scoring with an analytic rubric to have an understanding of the overall performance of class population. This is to better map and draw connection between different achievement levels in the class.

- Often time instructors have a mental picture of a students grade and what constitutes a pass and would adjust the rubric scoring accordingly [13]. It would be therefore more reasonable to use the analytic rubrics as a formative assessment framework focusing on directing student learning.

7 Conclusion

To address the concern that adding transparency to learning expectations may stifle creativity, we suggest that rubrics should not restrict the format or method but rather provide examples or anchors to show there are many ways to approach the same task. Therefore, rubrics such as the rubric bank focusing on meta-cognitive processes bring transparency and credibility to assessment without the need to sacrifice student creativity. While there is debate in the literature on the effectiveness and perceptions of students on receiving graded feedback on their consequent deliverables, there seems to be an agreement regarding the positive effect of descriptive, task specific, neutral and praise-free feedback on learning criteria. It would be therefore safe to say rubric designs that diagnose students knowledge about a concept, provide guidance on students performance through formative feedback or track students progress over the course of the learning period, and provide a consistent approach for their use better capture students performances and allow enhancement of student learning outcomes.

8 Acknowledgements

We would like to thank Professor Susan McCahan at the University of Toronto for her review and feedback on the article.

References

[1] Stevens, A., & Levi, A. (2005). Introduction to Rubrics: an assessment tool to save grading time, convey effective feedback, and promote student learning. Sterling, VA: Stylus Publishing.

[2] Anderson, L., & Krathwohl, D. A. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

[3] Whitlock, B., & Nanavati, J. (2013). A systematic approach to performative and authentic assessment. Reference Services Review, pp. 32-48.

[4] Jonsson, A. &. (2007). The use of scoring rubric: Reliability, validity and educational consequences. . Educational Research Review, 130144.

[5] Dochy, F., Gijbels, D., & Segers, M. (2006). Learning and the emerging new assessment culture. In F. D. L. Verschaffel, Instructional psychology: Past, present and future trends. Amsterdam: Oxford.

[6] Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. College Composition and Communication, 365387.

[7] Silvestri, L. &. (2006). Using rubrics to increase the reliability of assessment in health classes. International Electronic Journal of Health Education, 2530.

[8] Wolf, K. &. (2007). The role of rubrics in advancing and assessing student learning. The Journal of Effective Teaching, 314.

[9] Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability . Practical Assessment, Research & Evaluation, 71-81.

[10] Davidson, M., Howell, K., & Hoekema, P. (2000). Effects of ethnicity and violent content on rubric scores in writing samples. Journal of Educational Research, 367373.

[11] Messick. (1996). Validity of performance assessments. Washington, DC: National Center for Education Statistics.

[12] Mayes, R. (2004). Think like an Assessor. University of Wyoming, Science and Mathematics Teaching Center.

[13] Gantt, L. (2010). Using the Clark Simulation Evaluation Rubric with associate degree and baccalaureate nursing students. Nursing education perspectives, 101-105.