USING WEIGHTED SCORING RUBRICS IN ENGINEERING ASSESSMENT

Juan J. Salinas Professor Emeritus Department of Civil and Environmental Engineering Carleton University juan.salinaspacheco@carleton.ca

Jeffrey Erochko Assistant Professor Department of Civil and Environmental Engineering Carleton University Jeffrey.Erochko@carleton.ca

Abstract

When evaluating student work by deducting marks for errors, it is possible to underestimate the importance of dominant concepts and assign grades at a level that might not be in agreement with academic policies. Rubrics facilitate and expedite the marking process but it is important to examine in detail the parameters and limitations of this structured approach to assessment. The structure of the scoring rubric considered in this study promotes the consistency and validation of the assessment process and discriminates between evaluation components by assigning different weights to dominant and secondary criteria. The shape of the weight distribution function plays an important role in this process. In the proposed rubric structure, an array of performance levels is multiplied by an array of task components to arrive at a mark and a grade. A uniform weight distribution is easy to develop and utilize, especially for large classes, but it fails to recognize the importance of dominant components. The proposed approach allows the incorporation of single and multiple dominant criteria modeled by using step or linear distribution functions and adjusting the relative value of dominant and secondary components.

Keywords: Conceptual Rubrics, Scoring Rubrics, Performance Levels, Task Components, Dominant Components, Secondary Components, Validation and Reliability.

1. INTRODUCTION

Mechanics is a common course for engineering programs in Canadian universities. At Carleton, and probably at other institutions, the first-year mechanics course has a large enrollment and is taught by several instructors, in multiple sections, to students enrolled in a variety of engineering programs. The final exam, in this class of nearly 1000 students, is marked by 5 instructors with varying backgrounds and degrees of experience. For some engineering programs at Carleton, this is the only Mechanics course that is required. For students in Civil, Mechanical, and Aerospace Engineering, this is the first in a series of related courses. For those students, the development of certain competencies within this course is critical to the understanding of important concepts and skills in followup courses. One such competency is the ability to draw and use Free Body Diagrams (FBD) to analyze the state of equilibrium of rigid bodies. In Mechanics, a FBD is a simplified sketch of an entire system, or a smaller part of a complex system. On this sketch, contact or remote forces and moments, external or internal, are drawn as vectors. This is done in preparation for analysis of the system using the equations of equilibrium. The approach of "cutting off" a small part of a complex system and analyzing conditions at its boundaries is also used other fields of study, such as: electronics, fluids, thermodynamics, transportation, economics, and structures. In mechanics, the ability to investigate individual parts of more complex systems, often in the form of a free body diagram, is one of the fundamental skills that students will need going forward.

Guided by a team of instructional design specialists from Carleton's Educational Development Centre (EDC), three experienced Mechanics instructors recently developed a new set of learning objectives (LO) for this course, which included the ability for students to:

- Calculate the effect of forces and moments on rigid 2D and 3D bodies
- Analyze the equilibrium conditions of particles and rigid bodies in 2D and 3D space

These objectives reflect the importance of the primary tool that is used to solve these types of problems, the FBD.

In a further study, these Learning Objectives were found to be aligned with the following Graduate Attributes used in our accreditation process:

• Problem analysis: *An ability to use appropriate knowledge and skills to identify, formulate, analyze, and solve complex engineering problems in order to reach substantiated conclusions*

- Investigation: An ability to conduct investigations of complex problems by methods that include appropriate experiments, analysis and interpretation of data, and synthesis of information in order to reach valid conclusions.
- Use of engineering tools: An ability to create, select, apply, adapt, and extend appropriate techniques, resources, and modern engineering tools to a range of engineering activities, from simple to complex, with an understanding of the associated limitations.
- Limits of Knowledge: An ability to appreciate important elements of other engineering disciplines and the limits of the graduate's own knowledge.

2. ASSESSMENT PRIORITY

Based on the topic's high importance, it is essential that assessment in the first year mechanics course gives the topic of FBDs high consideration. It may even be appropriate to develop departmental academic policies that reinforce the principle that certain key topics, such as the use of FBD, should be given priority in assessment. To be able to produce consistent, reliable, and fair marking schemes and to discriminate between the various evaluation criteria, the use of a scoring rubric is necessary for this course. This is especially true in a course that is taught and assessed by multiple different instructors.

Any general exam-type question may necessitate multiple different steps within the solution; however, not all of the steps are necessarily of equal importance relative to the learning objectives of the course or the graduate attributes. Some of the steps within a solution may rely on skills and methods that students have before undertaking the study of mechanics. For example, the solution of a typical equilibrium problem would involve the following steps:

- 1) FBD. Force Magnitude, Direction, Location
- 2) Coordinate Axes
- *3)* Equilibrium Equations
- 4) Solution. Units. Significant Figures
- 5) Organization. Clarity. Legibility

In this solution, a significant amount of time may be spent by students in the solution of the equilibrium equations, parts (3) and (4); however, the skills used to actually solve the equations is based on math that they learning in high school. The new conceptual understanding of the problem comes in at the formation of the free body diagram (FBD) in part (1) and in the construction of the equilibrium equations themselves (not in the solving of those equations). If we were assessing student performance on this question based on the amount of work, or based on the time that it takes to do each step, which may be a common approach, then we may risk undervaluing the steps that are most directly related to the course learning objectives or graduate attributes. Of course, this does not mean that the other parts of the question are not valued or are unimportant, just that the relative importance of each part should be considered in a more deliberate manner than by the relative amount of work. In fact, this approach also presents the opportunity to assess other secondary goals and learning objectives that may not have any time or work associated with them such as organization, clarity, or legibility as shown in step (5). The presentation of clear and legible engineering calculations is a skill that may often be overlooked from an assessment point-of-view, but that we may value in the training of a competent engineer.

To address the issue of importance, this paper will present an analysis of the consequences of the use of a *weighted scoring rubric* to balance the assessment of primary and secondary learning objectives in a course.

3. BACKGROUND ON RUBRICS

The development and use of rubrics have been amply discussed in the literature. The TLT Group (2002) presents "A Rubric for Rubrics" a very helpful tool to assess the quality and use of rubrics in education. This publication emphasizes the need for clarity, a distinction between levels of assessment criteria, and reliability of the scoring methodology. Carleton University's EDC (n.d.) provides instructors with very practical advice and examples on how to develop and use scoring rubrics. Moskal (2000) examined, in general terms, issues of validity regarding the authenticity of the assessment process itself; and the reliability or consistency of the markers. Moskal and Leydens (2000) investigated in more depth the extent to which performance competencies can be generalized to other relevant criteria; and how important it is to identify specific components of the assessment that might be related to professional engineering practice and relevant to other elements. There seems to be a need for a structured scoring rubric of direct application to engineering practice that can offer the ability to identify evaluation criteria especially relevant to the current course and to follow-up courses. It is important for this rubric to offer a structure that will promote both validity and reliability in the process and is, at the same time, capable of discriminating between dominant and secondary assessment criteria. We hope that the proposed rubric in this paper will be able to meet these requirements. Tierney and Simon (2004) caution educators of the pitfalls of ready-to-wear rubrics available in a variety of web sites and suggest that more attention should be paid to the development of independent evaluation criteria and well-defined performance levels in order to ensure consistency in the evaluation process.

3.1 Rubrics

A rubric is generally structured as an array of rows and columns. The top row contains a number of performance levels, for example ranging from "unacceptable" to "exemplary." When using letter grades in final assessment, the usual range of performance levels is: F, D, C, B, and A, with potential to include their additional variances, e.g. B- and A+. The left-most column of the rubric array consists of a list of criteria or components of the specific task being assessed. The cells within the array contain detailed descriptors explaining specific competencies to be demonstrated by the student's work, spanning the whole range of performance levels for each criterion. The formulation of this *conceptual* rubric requires considerable effort by the instructors, in consultation with instructional designers, academic and accreditation committees, and especially with students.

The weighted scoring rubric proposed in this study operates in parallel with the conceptual rubric discussed above, except as follows: (a) beneath the top row (levels of *performance*) of the scoring rubric there is another row containing numerical values of marks associated with each level of performance, and (b) to the right of the leftmost column (criteria) there is another column with numerical values or weights corresponding to the relative importance of each criterion. In this weighted scoring rubric, the cells inside the array are scored with a value of either 0 (Zero) or 1, as explained below. Assigning a relative weight to each criterion of the assessment allows us to identify dominant criteria or primary criteria which we feel have significant relevance to the course learning objectives and graduate attributes. Using the score for each cell, numerical performance levels, and the weights assigned to each criterion, the overall score for an assessment task would be evaluated as follows:

$$\begin{split} P_{j} &= \text{Value of performance level } j \\ 0 &\leq P_{j} \leq 100 \text{ for } j = 1 \cdots k \\ W_{i} &= \text{Weight of criterion } i \\ 0 &\leq W_{i} \leq 100 \text{ for } 1 = 1 \cdots n \\ h_{ij} &= \text{Evaluation cell} \\ h_{ij} &= \begin{cases} 1 \text{ If criterion } i \text{ is assessed at level } j \\ 0 \text{ Otherwise} \end{cases} \\ \mu &= \frac{1}{100} \sum_{i=1}^{k} \sum_{j=1}^{n} W_{i} h_{ij} P_{j} \\ \mu &= \text{Student mark (%)} \end{split}$$

For example, as previously discussed, the concept of Free Body Diagrams (FBD) plays a critical role in the analysis process in Engineering Mechanics as well as in several other related courses and fields of study. It should, therefore, receive special consideration in the assessment process and be assigned a larger weight. The other (*secondary*) elements of the assessment criteria would then receive lower weights.

When using a conventional *conceptual* rubric, instructors examine a particular component of the task (criterion) and assign a number of "marks" depending on the degree of competency demonstrated by the student's performance. For example, if using a "point" system, and the specific performance level has a value of, say 2 pts, the instructor might assign 1.5 pts to a particular student for one criterion and 1.0 points for another criterion. To attain an overall mark, the instructor then may average the scores or perhaps give a grade depending on the instructor's conception of the relative importance of each criterion. There is a degree of subjectivity here, which may result in uneven assessments between instructors. The proposed scoring approach assigns explicit weights to the components of the evaluation, and offers a degree of consistency when multiple instructors are involved. These weights, and their relative values must be determined after considerable discussion between instructors. The proposed approach using an analytical scoring rubric is capable to discriminate dominant components of the assessment from secondary criteria using the general formulation explained below.

Consider a generalized weight distribution given by the function $W_i = f(a,b,i)$ shown in Figure 1.



Figure 1 Component weight distribution

There are certain *criteria* (*i*) which play an especially important role in the assessment process, to the extent that failure to meet competencies associated with these criteria would result in unacceptable marks or grades. We call these the *dominant* criteria. The rest of the criteria in the set are also important but to a lesser extent. We call these the *secondary* criteria. The relative weight of dominant and secondary criteria can be adjusted to achieve the level of discrimination desired for a particular situation. For example: Component W_1 (FBD) of the scoring rubric may be so important that a student that does not demonstrate a level of competency equal or greater than, say 48%,

should not receive a passing grade for the task, even if the level of competency in the other, *secondary*, criteria is good.

The shape of the weight distribution function and the condition that the sum of all weights must add up to 100 will allow us to determine either "a" (dominant) or "b" (secondary), extreme values.

4. DISTRIBUTION MODELS

4.1 Uniform Distribution

One commonly used marking approach consists of using equal weight for all components,

 $W_1 = W_2 = W_3 = W_4 = W_5 = 100/5 = 20 = a = b$ as shown in Table 1 below.

| Table 1 Sample Rubric | – Uniform Weights |
|-----------------------|-------------------|
|-----------------------|-------------------|

| | | | Performance | | | | | | |
|------|-------|----|-----------------------|-----------------------|----------------|-----------------------|----------------|--|--|
| | | | P ₁ | P ₂ | P ₃ | P ₄ | P ₅ | | |
| | | | F | D | С | В | Α | | |
| | | | 0 | 55 | 65 | 75 | 85 | | |
| | W_1 | 20 | 1 | 0 | 0 | 0 | 0 | | |
| ia | W_2 | 20 | 0 | 0 | 0 | 1 | 0 | | |
| iteı | W_3 | 20 | 0 | 0 | 0 | 1 | 0 | | |
| Ċ | W_4 | 20 | 0 | 0 | 0 | 1 | 0 | | |
| | W_5 | 20 | 0 | 0 | 0 | 1 | 0 | | |

Table 1 shows the case where student work on the *dominant* component (W_1) is at the F-level while the work on the *secondary* components is at the B-level. Therefore, the cell value h_{11} for criteria W_1 under the heading 'F' is assigned a value of '1' and the rest of the criteria are assigned cell values h_{ij} of '1' in the cells under the heading 'B'. All other cells are assigned a value of zero. This assumes that a separate rubric has been developed to define what constitutes each performance level for each criterion. The student mark in this case is calculated as follows:

$$\mu = [1/100] \{ 20(1)0 + 20(1)75 + 20(1)75 + 20(1)75 + 20(1)75 + 20(1)75 \} = 60$$

which corresponds to a somewhat ambiguous grade between D+ and C-. This probably constitutes a "passing" grade, even though the performance level for the dominant criterion W_1 is unacceptable. This would violate a "minimum grade" policy.

4.2 Step Distribution

To attain more discriminating result, the *dominant* component may be assigned a higher weight, as shown in Table 2. In this table,

$$W_1 > W_2 = W_3 = W_4 = W_5$$
 and $\Sigma W_i = 100$.

| 1 abie 2 Sample Rubie – Dominant Weight | Table 2 | Sample | Rubric - | Dominant | Weight |
|---|---------|--------|----------|----------|--------|
|---|---------|--------|----------|----------|--------|

| | | | Performance | | | | | | | |
|---------|----------------|----|-----------------------|----------------|----------------|----------------|----------------|--|--|--|
| | | | P ₁ | P ₂ | P ₃ | P ₄ | P ₅ | | | |
| | | | F | D | С | В | Α | | | |
| | | | 0 | 55 | 65 | 75 | 85 | | | |
| | W_1 | 40 | 1 | 0 | 0 | 0 | 0 | | | |
| iteria | W_2 | 15 | 0 | 0 | 0 | 1 | 0 | | | |
| | W ₃ | 15 | 0 | 0 | 0 | 1 | 0 | | | |
| C_{r} | W_4 | 15 | 0 | 0 | 0 | 1 | 0 | | | |
| | W_5 | 15 | 0 | 0 | 0 | 1 | 0 | | | |

In this case, the resulting student mark is calculated as:

$$\mu = [1/100] \{ 40(1)0 + 15(1)75 + 15(1)75 + 15(1)75 + 15(1)75 + 15(1)75 \} = 45$$

which corresponds to a grade of F. This weighting is clearly more able to identify poor performance in a *dominant* criterion. The degree of discrimination might be adjusted to be more tolerant by reducing the weight of W_1 and adjusting the secondary weights. For example, if

 $W_1 = 32$ and $W_2 = W_3 = W_4 = W_5 = 17$

the resulting grade would be 51, a passing low grade. The important point is that it may be possible to incorporate important academic policies, for example regarding graduate attributes, within the structure of a *weighted scoring rubric*.

4.3 Multiple dominant criteria

One other scenario might be considered by assuming there are two dominant components, eg:

$$[W_1 = W_2] > W_3 = W_4 = W_5$$
 and $\Sigma W_i = 100$

| | I | ab | le 3 | 38 | Samp | le R | lubri | ic – ' | Two | D | ominant | W | /eig | hts |
|--|---|----|------|----|------|------|-------|--------|-----|---|---------|---|------|-----|
|--|---|----|------|----|------|------|-------|--------|-----|---|---------|---|------|-----|

| | | | Performance | | | | | | | |
|---------|----------------|----|-----------------------|----------------|----------------|----------------|----------------|--|--|--|
| | | | P ₁ | P ₂ | P ₃ | P ₄ | P ₅ | | | |
| | | | F | D | С | В | Α | | | |
| | | | 0 | 55 | 65 | 75 | 85 | | | |
| | W_1 | 38 | 1 | 0 | 0 | 0 | 0 | | | |
| ria | W_2 | 38 | 0 | 0 | 0 | 1 | 0 | | | |
| iteı | W ₃ | 8 | 0 | 0 | 0 | 1 | 0 | | | |
| C_{r} | W_4 | 8 | 0 | 0 | 0 | 1 | 0 | | | |
| | W_5 | 8 | 0 | 0 | 0 | 1 | 0 | | | |

In this case, the resulting student mark is calculated as:

$$\mu = [1/100] \{38(1)0 + 38(1)75 + 8(1)75 + 8(1)75 + 8(1)75 + 8(1)75 + 8(1)75 \} = 46.5$$

which corresponds to a grade of F. Thus, the method is clearly able to identify a poor performance in one of the *dominant* criteria, even though performance was acceptable in the other dominant criterion. It is possible to incorporate a more tolerant assessment policy by using $W_1 = W_2 = 32$ and $W_3 = W_4 = W_5 = 12$ with a resulting grade of 51, also a passing low grade.

4.4 Linear distribution

It is also possible to consider varying degrees of dominance amongst the assessment criteria. Consider for example a linear variation of weights as follows:

W1 > W2 > W3 > W4 > W5 and $\Sigma Wi = 100$

Table 4 Linear variation of weights

| | | | Performance | | | | | | |
|-------|-------|----|-----------------------|-------|----------------|-----------------------|----------------|--|--|
| | | | P ₁ | P_2 | P ₃ | P ₄ | P ₅ | | |
| | | | F | D | С | В | Α | | |
| | | | 0 | 55 | 65 | 75 | 85 | | |
| | W_1 | 36 | 1 | 0 | 0 | 0 | 0 | | |
| ia. | W_2 | 28 | 0 | 0 | 0 | 1 | 0 | | |
| iteı | W_3 | 20 | 0 | 0 | 0 | 1 | 0 | | |
| C_r | W_4 | 12 | 0 | 0 | 0 | 1 | 0 | | |
| | W_5 | 4 | 0 | 0 | 0 | 1 | 0 | | |

In this case, the resulting student mark is calculated as: $\mu = [1/100] \{ 36(1)0 + 28(1)75 + 20(1)75 + 12(1)75 \\ + 4(1)75 \} = 48$

which still corresponds to a grade of F, but is able to distinguish between the relative weights of multiple secondary criteria. Alternatively, we might wish to use $W_1 = 30$; $W_2 = 25$; $W_3 = 20$; $W_4 = 15$; $W_5 = 10$ to obtain a grade of 52.5, a more tolerant policy. Risk-tolerant and riskaverse non-linear weight distribution functions were also evaluated but the results do not seem to deviate much from the linear case.

5.0 SUMMARY

A structured assessment rubric, the *weighted scoring rubric*, is proposed to evaluate tasks with single and multiple dominant components following uniform and nonuniform distribution functions. The shape of the weight distribution function allows the discrimination between dominant and secondary assessment criteria. The rubric has the ability to adopt specific academic grading policies, allowing a match between assessment and learning objectives and graduate attributes. For a given weight distribution function it is also possible to adjust the degree to tolerance of the rubric and ensure consistency of the marking process for large classes with multiple instructors.

References

Moskal, Barbara M. (2000). *Scoring rubrics: what, when and how*?. Practical Assessment, Research & Evaluation, 7(3). Available: http://PAREonline.net/getvn.asp?v=7&n=3

Moskal, Barbara M. & Jon A. Leydens (2000). *Scoring rubric development: validity and reliability*. Practical Assessment, Research & Evaluation, 7(10). Available: http://PAREonline.net/getvn.asp?v=7&n=10 Tierney, Robin & Marielle Simon (2004). *What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels*. Practical Assessment, Research & Evaluation, 9(2). Available: <u>http://PAREonline.net/getvn.asp?v=9&n=2</u>

TLT Group. (2002) *A Rubric for Rubrics*. Available: http://www.tltgroup.org/resources/rubrics/A_Rubric_for_ Rubrics.htm

Educational Development Centre, Carleton University (EDC). (n.d.) *Scoring Rubrics*. Available: http://carleton.ca/edc/wp-content/uploads/Rubrics.pdf