Rubrics in assessment

John Gough — Deakin University
jugh@deakin.edu.au

Introduction

I use to think that a “rubric” appeared in a Prayer Book as a special margin-note, sometimes printed in red ink, advising the reader about something that needs to be done at that stage in the reading or the church service. For example, such a margin-note might advise the unsure congregation when it was appropriate to stand, sit, kneel, sing, and so on. (Note that, as with most specialist educational vocabulary, consulting an ordinary dictionary will NOT reveal the intended technical meaning.)

A few years ago someone surprised me by using the term in the context of educational assessment. This (new?) sort of “rubric” was a way of formalising observations and judgments about what a student could do with a particular topic.

Here is a simple example. Students were given an open-ended task, and reported the following:

Yesterday we surveyed how some children traveled home after school.
- Five children went home in cars.
- Two children rode home on bikes.
- Eight children walked.
- Twelve children went home in a school bus.

The next stage of the task asks students to:
• explain how many wheels took the children home?
• draw or write to explain your answer.

The students’ responses to this task are then assessed by locating each response on a numbered scale, marked with the following descriptive-analytic categories. The student:
• draws some wheels and children, but either does not address the mathematics or does not solve any part of the problem correctly;
• shows the correct number of children, but not a correct (or plausible) number of wheels; or the child shows a correct (or plausible) number of wheels, but not the correct number of children;
• shows correctly (or plausibly) both the number of children and number of wheels, reaching a possible solution to the problem;
• shows more than one possible solution, or is at least aware of the uncertainty in the initial problem (all the children who went home in a car, for example, could have car-pooled in one car);
• uses pictures, numbers and equations (number statements) in displaying detailed correct (or plausible) alternative solutions to the problem.

(Adapted from Stenmark & Bush, 2001, p 123; this was itself based on an example originally given in Christina Myren Posing Open-Ended Questions in the Primary Classroom: 1995.)

Typically in this new sense a rubric is also a table, or matrix, with rows and columns.

- The columns identify the LEVEL or QUANTITY of achievement, or understanding, or skill - ranging across:
  • Not Begun, and Beginning; to
  • Developing; and beyond to
  • Established.
  OR
  • None or Not Very Much, or Rarely;
  • A Little, or Seldom;
  • Quite a Lot, or Often; and
  • Usually or Large Amount.

- The rows identify different subject-related aspects of what was being learned.

That is, a rubric is essentially a slightly elaborated and tabulated checklist of expected learning outcomes, usually augmented by exemplars of observable behaviours that enable a rubric-user to identify that Student X has learned Objective P to Definable/Observable level D.

Another way of describing a rubric is to say it is a weighted, or quantified checklist. In other words, although the word “rubric” is being used in a new way, what it means is NOT new as far as actual teaching and assessment practice goes.
Establish clearly what is to be learned, and what different levels of achieved learning would look like in students’ work.

Consider the assessment criteria for VCE CATs and SACs that VCE teachers have learned to use: these are essentially rubrics for grading project performance.

Moreover, as with any might-be educational innovation, rubrics are not a fool-proof method of assessment. (There are no educational cure-alls!) Viewed realistically, rubrics have good points, and not so good points.

One of the best features of a rubric is that, like the VCE criteria, students are usually given the rubric, before any instruction on a topic. In some cases, teachers and their students actively negotiate both the topic that they students will engage in, as well as the rubric-based assessment criteria that will be used to judge the level or quantity of the students’ eventual learning of this topic.

The rubric’s examples of aspects of a topic and their different levels of performance serve as nutshells outlines of the content knowledge being considered. This means that a rubric, if it is explained at the outset, is an example of an Ausubelian “advance organizer”. (For example Lefrancois describes David Ausubel’s theory of cognitive learning, and the technique of using an advance organizer to increase the learner’s ability to make sense of, and remember, new instruction as it proceeds: 1985: Chapter 6: Cognitive Learning). An advance organizer gives some preparatory glimpse of the topic ahead, that helps to set the scene, establishing a suitable mind-set for the learning experiences that are about to develop.

This is a formal version of the classic advice for effective public speaking:

• tell them about what you’re going to tell them;
• tell them;
• tell them about what you just told them.

Hence a rubric also objectifies, to some extent, the content that the student is going to learn, and the way the student will be assessed, increasing the open-ness and fairness of assessment.

Problems With Using Rubrics:
Student Manipulation; Criteria Ambiguity
As we know with the VCE criteria, students can take some proposed assessment criteria and manipulate their performance deliberately, and artificially, to ensure that every assessment box gets an appropriate tick, regardless of what the individual student might really want to do with the project topic.

Also, how effective, clear, or objective can a rubric be? In my experience the assessment categories “Not Evident through to Beginning, Developing, Established to Advanced” are sometimes difficult to interpret, and rely on subjective judgments, especially in the more open-ended subject areas, or skills areas, such as learning to read, or learning to write, or learning to multiply.

For example, what does “established” mean with learning to read? What is “advanced” writing? What standards are implicit in determining “established” or “advanced”? Similarly, with mathematics, what does “established” mean with a topic such as “addition” or “place-value”?

Usually this means only that the student has reached whatever the expected target-level is, for now. Hence, at Year 7, a student might have established “addition” of whole numbers, but still be developing “addition” of decimals, and be beginning “addition” of fractions. As for “addition” of vectors, matrices, or complex numbers - Not Begun - wait until Year 9 or later.

Rubrics can help state what is meant, in particular situations. But remember that rubrics are just a tabulated, weighted way of collecting and using learning outcomes, or educational objectives, or performance indicators (we are burdened with a pluralist variety of seemingly distinct terms, most of which seem to me to be broadly equivalent, that is, synonymous).

Equally a rubric, so called, often seems essentially equivalent to a checklist of observable behaviours. We have, implicitly, lived with rubrics
7/10 means what? Seven (similar) correct answers? Seven (different) achieved learning outcomes &/or performance indicators?

for a long time, without necessarily using the term, or knowing we were doing anything remarkable. Maybe this is one of the best features of a rubric. Aside from the novel name, it is no more than common-sense good practice - we have often practised this way.

Pursuing this a little further, if a traditional test is constructed so that specific questions (or test items) can be directly linked with a specific learning outcome (or objective, rubric category, etc.), then the test score becomes an indication - a measure - of achievement of all the learning outcomes (rubric categories, educational objectives, observable behaviours, etc.).

Ordinarily a numerical test or assignment mark of, say, 7 out of 10, means nothing more than the fact that 7 questions were correctly answered, and 3 weren't (with a possibility that some of the three zeroes were only partially faulty, and others might have been non-attempts, or omissions) - assuming that there is 1 mark for each test question, and the decision to award a single mark is made on an either/or right/wrong basis without splitting a mark. (We can discriminate more finely if we allow for half-marks, and partial answers, and so on. Equally we can discriminate more closely if we allow, for example, 3 marks for a question that contains, we expect, three distinct steps or parts, and allot 1 mark to each step. We are the controllers of our own marking schemes. All we require is common-sense and consistency - and efficiency.)

But if we know more about what the overall mark means we can see that a numerical score, carefully interpreted, is not JUST a NUMBER. For example, a mark of 7 out of 10, can refer to the ten items that represent the ten learning outcomes for Level Whatever of the CSF or VELS. Similarly the mark can be based on the ten educational objectives assigned for this specific topic in the school's integrated problem-based learning curriculum. Or the mark can be an overall measure of achievement of ten rubric categories. Hence 7 out of 10 is a measure of the AMOUNT of learning, not just a score for CORRECTNESS.

However this can be taken a step further.

Scoring a Rubric, Overall

Using a rubric, we can generate an overall rubric-mark by weighting the successive table-boxes or cells: e.g. any box in the Not Begun column scores 0, a box in the Beginning column scores 1, and so on up to, perhaps, 4 or 5 for a cell in an Advanced - one column higher in achievement than Established. Tick the boxes, according to observed evidence. Add up the corresponding boxes, and there is that student’s current rubric-mark. You can turn that total rubric-score into a percentage if you like, to facilitate direct comparisons between a score on one rubric and a score on another.

However, as is often the case with a bare numerical score, a number on its own tells little, if anything about ONE student, and does not necessarily enable us to make comparisons between TWO or more students. For example, Student A might have six Beginning scores, and zero everywhere else in the rubric; and Student B might have one Beginning and one Developing and one Established score, and zero for everything else; and Student C might, rather oddly, have two Established scores. Using the simple but crude numerical scoring system I outlined above, all three of the students have an overall rubric-total of 6.

As an alternative, we might adopt a rubric scoring system that maintains a clear separation between the different levels of achievement, with a student scoring 1 for any particular box in the rubric. The three students I have given as hypothetical examples already would have the following scores:

• Student A = B(6) D(0) E(0);
• Student B = B(1) D(1) E(1); and
• Student C = B(0) D(0) E(2).

This at least shows the important differences between the three that had otherwise been concealed in single-number scores.

However this leaves open how a rubric approach might then lead to a final "grade" or overall "mark" for a term, or half-year, or year. Moreover, it remains unclear how a rubric approach can provide some kind of ranked comparisons between students. While everyone knows that comparisons are odious, it would be naïve to imagine that schools can totally abandon the use of comparisons between students. End-of-Year-12 assessment, naturally linked with competitive vocational and tertiary selection, is necessarily based on scrupulous and cautious comparisons,
Can we use analytic rubric categories to distinguish development of work-habits, personal and social values? Schooling = CSF + VELS?

because of the need to allocate scarce and expensive resources as parsimoniously as possible, with greatest practical effect, and least cost and damage to the larger community.

Hop on the Rubric Bandwagon?
The rubric approach has been taken up with amazing enthusiasm by the National Council of Teachers of Mathematics (NCTM) in the USA, particularly in their recent publications, such as *Mathematics Assessment: A Practical Handbook* (2003). Virtually the ONLY assessment method used in this so-called “practical handbook” is that of rubrics, as though no other assessment method is practical. Perhaps this is more a reflection on the textbook and test-dominated assessment methods currently used in the USA, and the NCTM’s recent agenda for shifting to both curriculum and assessment based on investigative projects, problem solving and thinking mathematically.

Rubrics have an obvious role, potentially, with cognitive and performance learning. What about other aspects of schooling and learning? What about the woolier aspects of VELS?

If attitudes and study-habits and other affective aspects of learning matter to us as instructors, then we can try to build judgments about such things into part of the whole approach to assessment. This may include a checklist or rubric for such non-cognitive, affective things as:

- attendance;
- punctuality;
- class-contribution;
- neatness of personal presentation;
- clarity of voice;
- quality of oral communication;
- cheerfulness;
- manifest willingness to help; and
- patience in the face of customer stupidity or other difficulty;

… and so on.

To make a rubric that could assess affective achievement, we need to identify what we are concerned about. We define these as clearly as possible. We describe the distinguishing characteristics, so we can say, when we see behaviour X, we know that LITTLE of characteristic P is present, when we see behaviour Y, we know that SOME of P is present, and when we see behaviour Z we know that a LOT of P is present. Then we explain our criteria to our students, ahead of time, so they know how we are going to assess them (and they can’t complain when we do, because we were fair, and open, and warned them).

Assessing Student Attitudes and Values - Can Rubrics Help?
On a related matter we might ask: “Does students’ attitude steer us in our decisions on their cognitive assessments, tests or assignment marks?”

Sometimes it might. Sometimes it is reasonable to do this. But if the assessment criteria being used do not include student attitude, it is better to try to be impartial, and judge the cognitive behaviour independently of the affective.

Similar questions arise with values. In Victoria this is becoming a focus of attention as the nearly decade old collection of Curriculum and Standards Frameworks (CSFs: formerly published by the Board of Studies) for eight Key Learning Areas (KLAs) are augmented and re-framed in terms of so called Victorian Essential Learning Standards (VELS). (For example, at the internet homepage of the Victorian Curriculum and Assessment Authority (VCAA), http://www.vcaa.vic.edu.au, and related web-pages, such as: http://www.vcaa.vic.edu.au/prep10/CRP/index.htm)

The main point here is that the CSFs are now being taken as the curriculum content, for each of the KLAs. But generic skills, held to be “essential” and applying across most KLAs and at many levels of schooling, are being identified, for special (renewed?) attention. This parallels the emphasis already given to values and attitudes by the subject-association “Standards” movement in the USA, where “standards” have never been restricted to content-based “learning outcomes”, but also attempt to address ideals such as social tolerance, constructive work practices, intellectual honesty, and so on.

On the other hand, if attitude IS one of the criteria we are trying to teach, or foster, and assess, then we are entitled to “mark” it, and be clear about doing so. We might note that Student X obtained a test score of 85% (cognitively commendable), but the work-focus was weaker,
and contributions to small-group projects weaker still. Hence an overall "grade" for learning and study might be less than the impressive 85% registered for cognitive achievement alone.

Practical Consequences of Assessment Judgments
What we should do about one student who gets 60% and another gets 95%? Is public acknowledgement of differences in achievement fair? Should it be reported? Yes, if the assessment is well constructed. This should mean that the student with the higher mark is more competent. Otherwise the difference in the marks is meaningless.

Overall, SHOULD we be striving for (academic) excellence? Might this disadvantage some students? That depends on the task, or the curriculum goals. Moreover, what is "excellent" for one student (a "personal best") might be "mediocre" ("could try harder") for another, more able, but lazy student.

Consider learning to drive a car as an indicative example. Yes, we want ALL drivers on the road to be as competent as possible. But realistically we accept that people vary; their ability to learn to drive varies. So we settle for a MINIMUM - competence for ALL, and hope this will be good enough. Sometimes that is the best we can do. The natural road attrition may sort out the really ghastly drivers (survival of the fittest - but at whose cost?). Similarly the market-place will sort out the weakest trained-Tourism-professionals, or the weakest skilled supermarket check-out operators. Customers vote with their feet. Dangerous drivers have accidents. Poor employees (ex-students) find their own level of employability - an inverse version of the famous Peter Principle (that people naturally rise through employment ranks to the level of their incompetence). Weak mathematics students harden their low self-concepts, and antipathies towards mathematics, and move to a life-long avoidance of anything that smacks of mathematics. A minimum goal may not be good enough if we do not tackle unproductive attitudes.

Making and Using Rubrics in Mathematics Assessment
A "rubric" is a way of assessing what an individual student has learned about a particular topic. A rubric uses a checklist made of the sub-tasks or components within the task, with graded descriptions of how well a student can DO or has LEARNED the task.

For example, we might have a rubric for learning about early ideas of algebra.

Here the sub-tasks might include such algebra-skills as:

- using a letter as a pronumeral to stand for:
  - an unknown number;
  - a variable;
  - a set of numbers;
- distinguishing the value of a variable from the number of times the variable is used;
- translating a rule or law or relationship from words in a sentence into mathematical expression or formula;
- substituting a numerical value into a formula, and using ordinary arithmetic to calculate the resulting value of the substituted formula;
- transforming a rule or formula so that the formula is re-expressed in terms of a different variable or letter or pro-numeral;
- solving a simple equation, when some numerical values are known;
- making a simple graph that represents or displays a mathematical formula;
- interpreting a simple graph in terms of a mathematical formula.

And so on.

Any one of these sub-tasks can be described, using typical examples, so that we can see different ways that a student might possibly be able to work with the sub-task.

For example, with the first sub-task ("using a letter as a pronumeral to stand for an unknown number, or, a variable, or, a set of numbers") we can describe this as follows:

A student can do this sub-task when, for example, the student can:

- write a sentence such as, "Given some mathematical problem about an unknown number, let the letter X represent the unknown number", as in a "Think of a Number" problem;
- when working with a variable quantity such as Length, or Angle, or Time, the student can choose a suitable letter to represent the varying quantity;
- when a letter represents some varying quantity, such as P = the number of professors in a university, understanding that 3P represents THREE times that many professors;
• understanding that a letter can represent some collection of numbers, such as P standing for the set of prime numbers, or E representing the set of even numbers.

Given descriptions, such as these, for a sub-task, we can then use a simple number-grading, to rate or score the student's ability to do this sub-task. For example:

0 (zero rating) - means the student shows no understanding or the sub-task, or is unable to do any version of this sub-task correctly when asked to;
1 means the student has some familiarity with the ideas and processes of the sub-task, but makes frequent mistakes in calculation, substitution, or application (many errors, a few correct answers);
2 means the student knows about the ideas of the sub-task, and is usually able to correctly answer questions on this sub-task, but makes occasional errors (not many errors, mostly correct answers);
3 means the student correctly answers all questions on this sub-task.

We can use a rubric like this as a marking scheme for grading a student's performance on a worksheet, test, or project.

We construct the rubric:
- listing all sub-tasks in the task being assessed
- fitting each sub-task with a practical description about how the student might do this sub-task
- making a graded numerical score for different levels of skill with the sub-task, from NO SKILL, through to 100% CORRECT.

Then we identify within the worksheet, test, or project:
- which parts of the student's work correspond with which sub-tasks in the rubric; and
- how well (according to the graded numerical scoring) the student handled the sub-task.

Finally, we draw overall conclusions about how well the student handled the task, altogether, while also identifying those sub-tasks where the student performed less well, and still needs practice (and possibly re-teaching).

We might, for example, have 5 sub-tasks for a rubric, and each sub-task has a possible graded score from 0 to 3. We can work out the graded score for each sub-task, and add these together, to make a TOTAL GRADED SCORE of some number out of the MAXIMUM POSSIBLE TOTAL OF 15 (this total of 15 comes from 5 sub-tasks x maximum sub-task score of 3).

We might then look at the range of TOTAL RUBRIC SCORES for all students, and identify those who have achieved significantly lower than the “average” score.

Alternatively, we might look at the range of SUB-TASK scores, across all the students in the class, to identify those particular sub-tasks that a large number of students are still struggling to learn effectively.

In either case, we know WHICH students to give extra teaching to, and which particular parts of the topic need further teaching, clarification, and practice.

As already noted, the National Council of Teachers of Mathematics (NCTM) in USA has recently started using rubrics a lot as a new way of assessing students’ learning.

References and Further Reading


