Developing Assessment Criteria and Performance Standards

INTRODUCTION:
In course documents there is a hierarchy of:
- learning outcomes
- assessment criteria
- standards of performance.

There is a hierarchy of these three in the sense that the assessment criteria must align with the learning outcomes and not the other way round; and standards of performance must in turn elaborate differences in how well students do in terms of the assessment criteria.

LEARNING OUTCOMES AND CRITERIA:
Whereas the learning outcomes of a course describe the broad range of objectives for students, the criteria describe the aspects of performance that will be assessed. These must align\(^1\) with, and provide adequate coverage of, the learning outcomes. The standards then describe the characteristics of performance in terms of each of these criteria at each of the rating levels.

It is generally known that words alone do not adequately and completely define achievement standards. The words in the course document are a first (and important) element in the development and application of the standards to be applied to students’ performance.

Sadler (1987)\(^2\) defines these terms as:

**Criterion:** a property or characteristic by which the quality of something may be judged. Specifying criteria nominates qualities of interest and utility but does not have anything to offer, or make any assumptions about, actual quality.

**Standard:** a definite level of achievement aspired to or attained. Standards are about definite levels of quality (or achievement, or performance).

The assessment criteria capture the essence of what is important, critical and central about doing well at the particular discipline or area of study/activity covered by the course. The criteria should not include aspects of performance that are not essential and intrinsic to the nature of the activity/discipline. A course will usually have between 3 and 10 assessment criteria and these should be roughly of equal importance.

---

\(^1\)‘Align with’ does not mean that there is one assessment criterion for each learning outcome. It does mean that it should be clear that the set of assessment criteria provides a reasonable coverage of the set of outcomes. Such coverage is reasonable when it is clear that a student who does well at all of the assessment criteria is very likely to have achieved the learning outcomes of the course.

The criteria must be meaningful to the particular course; they should not be so generic that they could apply to any course. This does not mean that generic skills such as communication do not appear; it means that the criterion will be about communication in the context of that particular course in the particular way in which communication occurs in that discipline, subject or area of study or activity.

When developing assessment criteria it is helpful to review them in terms of:

• assessment criteria are clear and specific when it is obvious to a reasonable person from the statements of criteria alone (no course name) which discipline/study area they belong to.

Criteria are written in the form of:

The assessment for <name of course> will be based on the degree to which the learner can:

1. [brief description of the required performance, starting with a verb]

Examples of some clearly written assessment criteria from existing courses:

• observe and critically appraise drama works (starts with a verb, succinct, clear, related to the nature of the learning – drama)
• develop a personal movement vocabulary as a means of expression (again we have a clear picture of what the required performance looks like)
• create a design brief that incorporates design process and principles (here we know what the product will be and what its critical qualities must be)
• use probability and data analysis techniques to analyse distributions (we know what the task is and what it has to do).

Note that these criteria describe the actual performance required. Where possible avoid criteria which imply but do not describe the performance (e.g. demonstrate an understanding of....) Go straight to the point of what the demonstration would look like – is the student expected to describe, perform, analyse, draw, create?

PERFORMANCE STANDARDS:
Performance standards describe a level of achievement aspired to or to be attained.
Standards are about definite levels of quality (or achievement, or performance).
Standards have two dimensions:

1. What are they describing?
2. How many levels of performance do they attempt to describe?

What are they describing?
They can describe student performance either in total or in terms of each of the criteria individually.

(a) In total:
Here the standards attempt to describe the overall performance required. While the description will cover all the criteria, it will build a total picture of the student performance. Such a picture might read something like:
At this level:
The learner reads and understands basic scientific texts and selects the key ideas from texts, such as scientific journals. The learner gathers relevant scientific information and organises it in a meaningful way to present a case.

The learner explains fundamental principles of physics and chemistry. The learner designs a simple experiment to test a given hypothesis and conducts simple experiments, closely adhering to given practices including safety requirements.

The learner lists some of the applications of the physical sciences in their own world.

(b) Against each criterion individually (this is the approach required in externally assessed courses). We have numerous examples of these in current courses.

How many levels of performance do they attempt to describe?
Standards are the description of student performance at various levels. The levels can be:
(a) the award levels of the course – SA, CA etc. This model may be more appropriate where there are only a small number of awards, such as Satisfactory Achievement and High Achievement.
(b) a designated set of ratings that are then used to calculate an award. Traditionally TQA courses have used 3 ratings: (C, B and A). A ‘C’ rating is the minimum acceptable standard, an ‘A’ rating is the highest standard of performance and a ‘B’ rating is between these two. Even when the three ratings are used, it is not essential to try to describe performance at all three levels. A description of the performance at the outer levels – C and A – may be all that is needed, with B being obviously in between these two.

What makes the difference between different ratings
An analysis of standards statements used to date in course documents suggests that there are some key characteristics of performance that are used to discriminate one rating from another:
• does the learner know and can do only just enough or a lot? (breadth)
• does the learner carry out the required tasks only just adequately or very well? (quality)
• does the learner have only the basic knowledge and understanding or do they fully comprehend the required information and concepts? (comprehension).

In some courses, frequency of performance has also been used as discrimination between the ratings. This is often expressed as ‘often’, ‘sometimes’, ‘always’, ‘occasionally’ etc. This does not really tell us what the performance looks like and what you think is ‘often’ I might regard as ‘occasionally’. Frequency is usually a factor of the difficulty of the task. Describe the nature of the performance, not how often the student can do it.

---

3 The standards represent ‘anchor points’ along the continuum described by the criterion. It is more useful to have meaningfully described anchor points than it is to insist on a particular number of such points for all criteria. However, consistency is helpful unless it detracts from meaning and clarity.
Standards should be set to be reasonable – neither so high that most students will not achieve them nor so low that all will meet them. Even though TQA courses do not include a ‘fail’ or ‘unacceptable performance’ standard, it is often helpful to have one in mind so that the minimal acceptable performance (a ‘C’ rating) is put into sharper focus.

It is easier to achieve reliable and valid assessment in terms of ratings where the standards describe contextualised knowledge and skills rather than generic performances.

Performance standards are clear and specific when it is obvious to a reasonable person looking at the standards alone (no assessment criterion statement) which assessment criteria they belong to and what order they are in (no statement of A, B, C).

Consider the following to see if it would meet the above test. What is the probable criterion and which would be the ‘A’ rating, which the ‘B’ rating and which the ‘C’?

- set out the content of the report clearly and logically
- produce a report that contains an introduction, a body and conclusions
- structure the report following established conventions and justify the conclusions by the information in the body of the report

Some guidelines:

1. Write standards as simply as possible avoiding unnecessary words that detract from clarity

Rather than say: *demonstrate the ability to identify relevant sources* just say: *identify relevant sources*.

2. Write them as simply as possible avoiding unnecessarily complex structures and ones that include several aspects of performance. Can the standard below be expressed more clearly?

*Consider, select and use strategies to anticipate and solve minor problems; clearly identify more complex issues and consider, select and use strategies seeking to resolve them.*

3. Avoid phrases that don’t describe any particular performance. E.g.:
   - *Demonstrate a sound knowledge of…*
   - *Demonstrate extensive knowledge and comprehensive understanding of a wide range of…*
   - *Demonstrate some willingness to attempt*
   - ‘Attempt to…’
   - *Endeavour to…*

4. Avoid very subjective terms such as ‘satisfactory’, ‘good’, ‘acceptable’, ‘adequate’, ‘very well’. These mean different things to different people. Describe the actual performance you expect to see from the student.
5. Avoid comparative terms such as ‘less’, ‘more’, ‘better’. Again these can be interpreted very differently by different people.

6. Always try to describe the actual performance you expect to see. Some good practice examples from our current courses might be:
   - use relevant sources of information in the report
   - produce designs incorporating all key features
   - list the basic steps in the process
   - initiate conversations with toddlers
   - adapt technology/communication methods and styles to suit audience
   - given an unfamiliar computing task, apply top down design in order to identify the required subtasks.

7. Do not use the context of the performance as part of the standard, eg do not use such terms as “with teacher guidance”. If the independence of action is part of the standard, describe it using words such as: “when provided with a range of options, selects ...”.

8. Limit the number of elements of a standard to that which is sufficient but not superfluous (as a guide consider 2 to 5 elements per criterion). Writing standards is difficult so write a few very good ones rather than a lot of not very good ones. There does not have to be an element in each of the ratings. E.g. a particular behaviour might not be expected for a ‘C’ rating but is expected for a ‘B’ and ‘A’ rating. It might not be meaningful to try and describe the ‘B’ rating; it is in-between the ‘A’ and the ‘C’ rating.

9. Try to make sure the elements go to the heart of the criterion and are not focussing on peripherals. An example might be that the criterion is to do with reflecting on own creative performance but one of the elements of the standard is being able to write an essay. Is this central to the criterion?

10. Try to make the distinction between the ratings as clear as possible. You do not need to use the same form of words. For example:

<table>
<thead>
<tr>
<th>‘C’ rating</th>
<th>‘B’ rating</th>
<th>‘A’ rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner:</td>
<td>A learner:</td>
<td>A learner: trouble-shoots common system malfunctions</td>
</tr>
<tr>
<td>lists the basic</td>
<td>describes the function of the major systems</td>
<td></td>
</tr>
<tr>
<td>components</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11. The distinction between the ratings can sometimes be best described by using different verbs. Possible examples might be:

<table>
<thead>
<tr>
<th>‘C’ rating</th>
<th>‘B’ rating</th>
<th>‘A’ rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>use</td>
<td>modify</td>
<td>design</td>
</tr>
<tr>
<td>list</td>
<td>describe</td>
<td>analyse*</td>
</tr>
<tr>
<td>identify/select</td>
<td>describe</td>
<td>evaluate*</td>
</tr>
<tr>
<td>participate</td>
<td>initiate</td>
<td>modify</td>
</tr>
<tr>
<td>‘C’ rating</td>
<td>‘B’ rating</td>
<td>‘A’ rating</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>state</td>
<td>formulate</td>
<td>predict</td>
</tr>
<tr>
<td>copy</td>
<td>draft</td>
<td>design</td>
</tr>
<tr>
<td>respond</td>
<td>guide</td>
<td>facilitate</td>
</tr>
</tbody>
</table>

* Analysis and evaluation are typical of TQA 3

12. The distinction might be best described in terms of the complexity or sophistication of the task. E.g.:

<table>
<thead>
<tr>
<th>‘C’ rating</th>
<th>‘B’ rating</th>
<th>‘A’ rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-on-one</td>
<td>small groups</td>
<td>large groups</td>
</tr>
<tr>
<td>brief notes</td>
<td>memos and letters</td>
<td>reports</td>
</tr>
<tr>
<td>familiar operations</td>
<td>using whole numbers and fractions</td>
<td>using decimals, fractions and percentages</td>
</tr>
</tbody>
</table>

13. The distinction might be in terms of the quantum of response provided by the learner:

<table>
<thead>
<tr>
<th>‘C’ rating</th>
<th>‘B’ rating</th>
<th>‘A’ rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>basic, few, limited, most obvious</td>
<td>a range of, the most important</td>
<td>comprehensive, extensive</td>
</tr>
</tbody>
</table>

14. The distinction might be in terms of the quality of response provided by the student, particularly if describing a performance or a product:

<table>
<thead>
<tr>
<th>‘C’ rating</th>
<th>‘B’ rating</th>
<th>‘A’ rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>usable</td>
<td>finished</td>
<td>polished</td>
</tr>
<tr>
<td>clear</td>
<td>clear</td>
<td>expressive</td>
</tr>
<tr>
<td>functional</td>
<td>structurally sound</td>
<td>creative, innovative</td>
</tr>
</tbody>
</table>

15. It is probably a good idea to start with the ‘C’ rating – what is the minimum acceptable performance? Then go to the ‘A’ rating – what would a very good performance look like? Only then come back and think what an in-between performance might be.
Some examples of good practice standards from our existing courses:

Example 1
Criterion  Use basic research skills to locate and record information

<table>
<thead>
<tr>
<th>Rating ‘C’</th>
<th>Rating ‘B’</th>
<th>Rating ‘A’</th>
</tr>
</thead>
<tbody>
<tr>
<td>The learner:</td>
<td>The learner:</td>
<td>The learner:</td>
</tr>
<tr>
<td>• uses a limited range* of provided information sources</td>
<td>• locates and uses a range of information sources</td>
<td>• locates and uses a wide range of information sources</td>
</tr>
<tr>
<td>• uses tools and strategies as directed to collect and organise information</td>
<td>• uses tools and strategies to collect and organise information</td>
<td>• effectively uses a range of tools and strategies to collect and organise information</td>
</tr>
<tr>
<td>• sorts information into simple categories (such as relevant/irrelevant to task)</td>
<td>• sorts information into appropriate categories</td>
<td>• sorts information into appropriate categories and identifies the relative significance of information to task</td>
</tr>
</tbody>
</table>

* Range of sources: has dimensions of type (primary and secondary), number (how many sources) and scope (books, academic articles, internet, film/video etc.)

Example 2

CRITERION:  USE DIFFERENTIAL CALCULUS IN THE STUDY OF FUNCTIONS
Rating ‘C’
A learner:
• calculates an average rate of change from given information (graph or table), with suitable units

Rating ‘B’
In addition to the standard for a C rating, a learner:
• interprets the average rate of change from given information, including the case where there is no change

Rating ‘A’
In addition to the standard for a C and B rating, a learner:
• sketches a graph of the rate of change of a function whose graph is given.
## Example 3

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Satisfactory Achievement</th>
<th>High Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrate basic numeracy skills relevant to the workplace</td>
<td>Perform a limited range of familiar and predictable calculations of whole numbers with the 4 operations (+, -, x, ÷)</td>
<td>Correctly carry out the four calculations with whole numbers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correctly carry out simple additions with decimals, fractions and percentages.</td>
</tr>
<tr>
<td>Measure familiar and predictable quantities using simple and routine measuring instruments</td>
<td>Correctly measure length, mass, capacity, time, temperature using simple instruments graduated in familiar units</td>
<td></td>
</tr>
</tbody>
</table>