

Assessment criteria, standards and marking schemes

This document contains information on the following:

- **Assessment criteria and standards for course learning goals**
- **Criteria and standards for assessment tasks**
- **Marking schemes and standards for assessment tasks**



Assessment criteria and standards for course learning goals

Definition

The assessment criteria for a course ('course level criteria') are the dimensions with which you will judge how well a student has achieved the learning goals of a course. Criteria might be likened to yardsticks.

Standards are the actual levels which might be achieved by students against the criteria. Standards might be likened to measurements on the yardsticks.

Course level criteria

Explain

Often it seems completely clear how you will tell whether or to what extent a student has achieved a course learning goal. For example, if the learning goal were:

the student is able to explain Newton's laws of motion

then surely the criterion for assessing a student's achievement of this learning goal would be the number of laws explained correctly. The possible standards will then be one, two or three laws correctly explained.

However, questions and doubts arise immediately. What of the student who correctly states all three laws, and who paraphrases two of them? What of the student whose explanation of the laws consists of rewriting them as mathematical equations? What of the student who states a law incorrectly, but then proceeds to give a correct explanation of the law as it ought to have been stated?

Perhaps the assessment criteria need to be set out more explicitly:

- *In judging how well a student has achieved this learning goal the following will be considered*
- *the student's ability to state laws correctly*
- *the student's ability to state laws clearly in their own language - grammar and syntax will be considered **only** if their deficiencies make a student's meaning unclear*
- *the student's ability to illustrate the import of a law by applying it to an example.*

The standards which may be achieved on each criterion are still numerical, and are the same for each criterion:

- *one, two or three laws stated correctly*
- *one, two or three laws correctly paraphrased*
- *one two or three laws illustrated by application.*

How these standards might feed into a grading scheme is another issue.

The 'bigger' criteria

As they stand the criteria in the example do not amount to much more than a marking scheme. However, it is easy to see that they are examples of three more general criteria:

- the student's ability to correctly state definitions and rules
- the student's ability to rephrase definitions and rules in their own words
- the student's ability to set down simple examples illustrating the import of a definition or a rule.

When they are stated in this way it is clear that the criteria may be used to look at a student's achievement of a whole class of objectives.

Communicate

An example from language learning might be:

the student is able to communicate orally in various common situations.

A first try criterion might be the number of situations (taken, perhaps, from a list) in which a student's communication is adequate. The standards will then be numbers.

A second look at the learning goals might lead the teacher to look at:

- the variety of situations in which the student can be understood
- the variety of situations in which the student can understand another speaker.

A third look might lead the teacher to look at the 'bigger' criteria:

- comprehension (ability to understand)
- fluency (ability to express).

These criteria will now be applicable to other communication learning goals (written as well as oral, for example). The difference when they are applied to various course learning goals will lie in the standards.

Generic examples

Many criteria look as if they might be applicable in many course areas. Generally they are, but only if the terms used are interpreted appropriately according to context. Here are some examples:

- the quality of the student's argument
- the quality of the student's expression
- appropriateness of the data presented
- ability to define key terms
- application of theory to specific problems or in specific situations.

In a course dealing with literature, for example, 'appropriateness of the data presented' might mean 'appropriateness of the examples mentioned', or even 'appropriateness of the scholarly writing cited'. Given that the criteria must be interpreted for each specific course area, the main use of such generic examples is likely to be as a source of ideas from which more meaningful criteria might be derived.

Standards for course level criteria

The standards which students might achieve on course level criteria frequently are specified in rather vague terms. The real definition of these standards happens when you specify how you will judge the standard achieved on the basis of the student's attempts at the various assessment tasks. Sometimes the main role of standards for course level criteria is to help describe the qualities of achievement required for given grades in the course. These are referred to as grade descriptors (see the following PDF document). [Grades and feedback](#)

Download an example

The School of Geography, Planning and Architecture has devised a set of criteria and standards on which marking schemes for work in their area are likely to be based. The criteria are extremely detailed, but the issue of how to arrive at overall grades is left to individual course profiles. (This document, in rich text format, can be downloaded from the [Download area](#) of the Assessment section of the TEDI website.)

Criteria and standards for assessment tasks

Definition

Criteria and standards for assessment tasks are quite analogous to course level criteria and standards. Thus the definitions are almost identical.

Criteria, or, more correctly, the assessment criteria for an assessment task ('task level criteria') are the dimensions along which you will judge how well a student has carried out the assessment task. Criteria might be likened to yardsticks.

Standards are the actual levels which might be achieved by students against the criteria. Standards might be likened to measurements on the yardsticks.

'Made up' examples of criteria and associated standards

Examples of criteria really need to be linked to the relevant assessment tasks. Some 'made up' examples are given below, while 'real' examples are given either for downloading or as external sites in the next section, [Marking schemes and standards for assessment tasks](#).

The standards which might be achieved on each criterion are labelled with possible levels of achievement on the criterion. Usually the performance needed to achieve such a standard is set out in words.

1. Task: A formal essay with abstract

Criterion: The quality of the abstract

Standard	Required performance
Very poor	no abstract
Poor	inaccurate or unduly brief abstract
Marginal	accurate abstract; slightly incomplete or rather wordy
Adequate	accurate and complete abstract; may be a little wordy
Good	accurate and complete; no unnecessary words

Criterion: The appropriateness of the literature cited

Standard	Mark (Max 20)	Required performance
Very poor	0 - 1	no literature cited
Poor	2 - 4	some literature cited; only some references to argument
Marginal	5 - 6	literature cited is appropriate; no citations at some crucial points
Adequate	7 - 8	literature cited is appropriate and covers all crucial points, however, it is limited to references supplied by teaching staff
Good	9 - 10	literature cited is appropriate and covers all crucial points, includes some references found by the student independent of the teaching staff

Criterion: Quality of the argument

Standard	Mark (Max 20)	Required performance
Very poor	0 - 3	there is no argument - isolated statements are made but not connected
Poor	4 - 9	there is an argument; however, it is logically flawed and unclear
Marginal	10 - 13	the argument seems logically correct but is in some important areas unclear
Adequate	14 - 16	the argument is logically correct and clear in all important aspects
Good	17 - 20	the argument is logically correct and clear in all important aspects; the argument is also surprising, original, or very elegantly put

Criterion: Quality of English expression

Standard	Mark	Required performance
Inadequate	no marks are given as inadequate work will receive 0 in the other categories	poor expression makes student's work unintelligible
Adequate	no marks are given as inadequate work will receive 0 in the other categories	student's work readily intelligible

2. Tasks: The student will design and carry out an experiment to determine whether

Criterion: Appropriateness of the experiment to the question posed

Standard	Mark (Max 15)	Required performance
Poor	0 - 2	no experiment proposed
Marginal	3 - 6	experiment only marginally relevant
Adequate	7 - 11	experiment appropriate, but a textbook standard experiment
Good	12 - 15	experiment appropriate and to some extent original

Criterion: Execution of the experiment

Standard	Mark (Max 15)	Required performance
Poor	0 - 2	no experiment carried out
Marginal	3 - 6	experiment only partially/very sloppily executed
Adequate	7 - 11	experiment completed, but not particularly well executed
Good	12 - 15	experiment completed, done well, tidily etc

Criterion: Interpretation of the results

Standard	Mark (Max 10)	Required performance
Poor	0 - 1	no results/no write up
Marginal	2 - 4	interpretation inaccurate or irrelevant to the question
Adequate	4 - 7	interpretation accurate, but relevance not made completely clear
Good	8 - 10	interpretation accurate, relevance completely clear

Criterion: Quality of English expression

Standard	Mark	Required performance
Inadequate	no marks are given as inadequate work will receive 0 for interpretation	poor expression makes student's work unintelligible
Adequate	no marks are given as inadequate work will receive 0 for interpretation	student's work readily intelligible

Criteria and assessment tasks

Normally, when giving students an assessment task to attempt, you will tell the students what criteria are to be used in judging their attempts. Ideally you also will give them at least an idea of the relative importance of the various criteria.

For the abstract/essay example above, the information for students when the task is set might say:

In evaluating your essay, the quality of the argument and the appropriateness of the literature you cite to support it will be of major importance. To gain a good grade you will need to show evidence of reading beyond the references supplied by the teaching staff. Also important will be your ability to summarise your essay in a concise abstract following the models seen in class. English expression will not be considered unless your expression is so poor that it makes your essay unintelligible to the marker.

Marking schemes and standards for assessment tasks

Definition

A marking scheme is a way of assessing and reporting the value of a student's attempt at an assessment task.

Profiles

Marking a student's attempt at an assessment task may simply mean indicating what standard the student has attained on each of the criteria. These 'indications' may simply be ticks in the boxes of a pre-printed feedback sheet, or they might be notes written by the marker. The student will then receive a profile – a list of standards achieved on the various criteria – as their 'mark' for the task.

Using the first example in the previous section on [Criteria and standards for assessment tasks](#) we might report:

In this essay your argument was rated as adequate, while the appropriateness of the literature you cited was good. Your abstract was rated marginal as it was very wordy. Your expression was adequate for this task. Please see the attached sheet for a definition of the standards used above.

A criteria and standards summary sheet would then be attached to the student's essay.

Marks

Marks may be associated with a student's attempt at an assessment task in the same way that grades may be associated with a student's overall attainments in a course. A mark may be a number or a label of a category.

The overall mark is the total of component marks – overall standard expressed as a number. The simplest method of allocating a mark to a student's attempt at an assessment task is to allocate marks for the standard reached on each criterion and then to add those marks up.

To back up this method, marks first need to be associated with the standards achieved. The relative importance of the various criteria is signified by allocating more or fewer marks to achievements against them.

Here is the essay example given above. Now marks have been included and the overall mark is the total of the marks on the criteria:

Your **mark** for this assessment task is the **total** of the marks awarded for the abstract, the literature and the argument.

Criterion: The quality of the abstract

Standard	Mark (Max 10)	Required performance
Very Poor	0 - 1	no abstract
Poor	2 - 4	inaccurate or unduly brief abstract
Marginal	5 - 6	accurate; slightly incomplete or rather wordy
Adequate	7 - 8	accurate and complete abstract; may be a little wordy
Good	9 - 10	accurate and complete abstract; no unnecessary words

Criterion: The appropriateness of the literature cited

Standard	Mark (Max 20)	Required performance
Very poor	0 - 1	no literature cited
Poor	2 - 4	some literature cited; only some references to argument
Marginal	5 - 6	literature cited is appropriate; no citations at some crucial points
Adequate	7 - 8	literature cited is appropriate and covers all crucial points, however, it is limited to references supplied by teaching staff
Good	9 - 10	literature cited is appropriate and covers all crucial points, includes some references found by the student independent of the teaching staff

Criterion: Quality of the argument

Standard	Mark (Max 20)	Required performance
Very poor	0 - 3	there is no argument - isolated statements are made but not connected
Poor	4 - 9	there is an argument; however, it is logically flawed and unclear
Marginal	10 - 13	the argument seems logically correct but is in some important areas unclear
Adequate	14 - 16	the argument is logically correct and clear in all important aspects
Good	17 - 20	the argument is logically correct and clear in all important aspects; the argument is also surprising, original, or very elegantly put

Criterion: Quality of English expression

Standard	Mark	Required performance
Inadequate	no marks are given as inadequate work will receive 0 in the other categories	poor expression makes student's work unintelligible
Adequate	no marks are given as inadequate work will receive 0 in the other categories	student's work readily intelligible

Here the quality of the abstract and that of the literature cited are together worth about the same possible marks as the quality of the argument. This indicates that quality of the argument is the most important criterion. Consistent with the specifications given to students, quality of expression is not awarded any marks as inadequate expression leads to no marks being awarded on the other criteria.

Here is the experiment example given above. Now marks have been included and the overall mark is the total of the marks on the criteria:

*Your **mark** for this assessment task is the **total** of the marks awarded for the design of the experiment, the execution of the experiment and your written interpretation of the results.*

Criterion: Appropriateness of the experiment to the question posed

Standard	Mark (Max 15)	Required performance
Poor	0 - 2	no experiment proposed
Marginal	3 - 6	experiment only marginally relevant
Adequate	7 - 11	experiment appropriate, but a textbook standard experiment
Good	12 - 15	experiment appropriate and to some extent original

Criterion: Execution of the experiment

Standard	Mark (Max 15)	Required performance
Poor	0 - 2	no experiment carried out
Marginal	3 - 6	experiment only partially/very sloppily executed
Adequate	7 - 11	experiment completed, but not particularly well executed
Good	12 - 15	experiment completed, done well, tidily etc

Criterion: Interpretation of the results

Standard	Mark (Max 10)	Required performance
Poor	0 - 1	no results/no write up
Marginal	2 - 4	interpretation inaccurate or irrelevant to the question
Adequate	4 - 7	interpretation accurate, but relevance not made completely clear
Good	8 - 10	interpretation accurate, relevance completely clear

Criterion: Quality of English expression

Standard	Mark	Required performance
Inadequate	no marks are given as inadequate work will receive 0 for interpretation	poor expression makes student's work unintelligible
Adequate	no marks are given as inadequate work will receive 0 for interpretation	student's work readily intelligible

Note that these marking schemes for the individual criteria do not, and need not, make marking mechanical. They are to some extent generic and need to be specified further for specific essay or experiments in particular courses.

In the essay, there is still a call to judge whether, for example, the literature cited is appropriate and whether citations have been given at all critical points. If the literature cited is judged appropriate, but it is judged that citations are not given at some critical points, then the marker still has to decide whether to allocate 5 marks or 6.

In the experiment, staff designing the marking scheme must still decide, for example, which experiments are relevant to the question and to what extent. Where an interpretation is accurate, but the relevance not completely clear, markers will need to decide how to award a mark in the range 4 to 7 and so on.

If there is more than one marker then measures still will have to be taken to ensure consistency. When feedback is provided to students it may still be necessary to provide paradigm examples (exemplars) of work of various standards to make clear to students why they were judged to have met specific standards.

Overall standards

You do not need to award numerical marks either for the standards reached on the criteria or for the overall assessment task. Here is the example shown above, using a recipe for determining the overall mark.

Essay example

Possible overall standards on this assessment task are: poor, marginal, adequate, and good.

Your overall standard is derived from the standards you achieved on the criteria of Abstract, Literature, Argument, and Expression as follows:

- Inadequate expression means poor overall, since the essay is incomprehensible.

- The overall standard for the assessment task cannot exceed the standard for Argument (with very poor or poor argument leading to poor overall).
- The overall standard is reduced by one level from the Argument standard for every level worse the standards for Abstract and Literature fall below that of Argument.
- The overall standard is increased by one level for every level better the standards for Abstract and Literature exceed that of Argument (but only up to the same standard as for Argument).

This means, for example, that:

- A profile (Abstract: marginal; Literature: good; Argument: adequate; Expression: adequate) gives an overall standard of adequate.
- A profile (Abstract: marginal; Literature: marginal; Argument: adequate; Expression: adequate) gives an overall standard of marginal.

Experiment example

Possible overall standards on this assessment task are: poor, marginal, adequate, and good.

Your overall standard is derived from the standards you achieved on the criteria of Appropriateness of the experiment, Execution of the experiment and your written Interpretation of the results as follows:

- If either the Appropriateness or the Execution is judged poor then the overall standard is poor.
- The overall standard is the lesser of the standards achieved on Appropriateness and Execution unless
 - either the standard of the Interpretation is good, in which case the overall standard is one higher
 - or the standard on Interpretation is poor, in which case the overall standard is one lower.
- If Expression is inadequate then Interpretation is poor.

This means for example that:

- A profile (Appropriateness: adequate; Execution: marginal; Interpretation: poor; Expression: adequate) gives an overall standard of poor.
- A profile (Appropriateness: adequate; Execution: marginal; Interpretation: good; Expression: adequate) gives an overall standard of adequate.
- A profile (Appropriateness: good; Execution: poor; Interpretation: good;

Expression: adequate) gives an overall standard of poor.

- A profile (Appropriateness: marginal; Execution: good; Interpretation: good; Expression: adequate) gives an overall standard of adequate.
- A profile (Appropriateness: marginal; Execution: good; Interpretation: poor; Expression: adequate) gives an overall standard of poor.
- A profile (Appropriateness: marginal; Execution: good; Interpretation: adequate; Expression: adequate) gives an overall standard of marginal.

Examples downloadable or accessible on the Web

See Cynthia Mitchell's course profile and resume criteria (in the [Download area](#) of the Assessment section of the TEDI website). They contain an assessment task, the task level assessment criteria and standards, and a marking scheme set out on a sheet suitable for returning to students as feedback.

lean Russell's QM231 assignment document (available from the [Download area](#) of the Assessment section of the TEDI website) includes an assessment task and the task level criteria and standards. No marking scheme is provided here.

Simone Kelly's CO353 document (in the [Download area](#) of the Assessment section of the TEDI website) includes an assessment task, together with task level assessment criteria, as well as a marking scheme on a sheet which can be returned to students as part of the feedback. Standards for the criteria are not specified in this document.

A generic grading scheme for projects is shown at:

<http://www.waukesha.tec.wi.us/busocc/blaw/proj.html>. The grading scheme here goes directly to grades ('A' to 'F'). The performances required to attain each grade are set out. However, the criteria and standards used are left implicit in the descriptions.