Interpreting and grading student learning: an overview

Grading as interpreting

Grading or marking are often terms that are used synonymously with assessing. Grading is a high stakes activity, the results from which students use to define themselves as learners. Grading is the process of interpreting students’ learning products and performance:

- to reflect where students stand in relation to an orderly development of competence
- to inform both student and teacher not only where the student currently is, but also what needs to be done to improve that position
- to be combined with other grades in order to meet administrative requirements for awarding of levels of pass, and the like

Not only is grading a high stakes activity, but it is also a highly subjective activity of interpretation that relies heavily on wisdom of practice. Expert assessment behavior, including a high level of skill in interpreting and grading students’ performances and products depends on assessors having:

- a detailed knowledge of their discipline, of curriculum intentions and of learners
- a detailed knowledge of assessment options and understanding of the limitations of these options
- clarity in understanding the purposes of the assessment
- a repertoire of meaningful approaches that have been intentionally developed for interpreting students’ learning performances
- awareness of contextual influences on their practice and of the limitations of their own interpretations and judgments

Who will grade?

Common practice often assumes that it is the teacher who will interpret and assess students’ learning performances. However, a number of alternative contributors to assessment are possible, for example:

- student self assessment
- student peer assessment
- external examiner
- expert professionals and community representatives
- computer aided assessment
- teachers

The involvement of students in interpreting and judging learning performance is fundamental to their development of self-critical and independent learning. There is evidence to suggest that, with proper training and well-supported execution, the interpretations and judgements made by students on their own and their peers’ learning performances are marginally more consistent and reliable than those given by multiple sessional tutors. Being involved in the process of assessment is a useful learning activity in its own right.

It is the nature of the task itself and its purposes that should determine who will contribute, and how, to make judgements about the final outcome. Based on information from sometimes multiple interpretations of a student’s work, the teacher assigns a grade. The overall management of assessment processes needs to allow scope for moderating final grades to ensure an accurate representation is made of each student’s demonstrated capabilities and performance.

Why do we grade?

Grading processes can be organised around different goals, or as Wolff (1969) coined these, three species of grading:

- **Criticism:** the analysis of a product or performance for the purpose of identifying and correcting its faults or reinforcing its excellence
- **Evaluation:** the measuring of a product or performance against an independent and
objective standard of excellence, oriented towards the professional goal of indicating if a person is qualified

- **Ranking**: the relative comparing of students’ performances to achieve specific one-off decisions such as who will receive scholarships.

Grading activity that focuses on the goal of ranking produces the greatest anxiety in university departments, provokes the most controversy and opposition, and yet results in the least, if any, advantage for learning.

A question that may be raised in planning assessment tasks is whether students’ learning products and performances ought to be graded at all. At Alverno College in the USA, assessment is performance-based with self-assessment being an integral part along with teacher, peer, and external assessment. Rather than using grades, performance standards are established in clearly defined profiles of desired learning outcomes organised according to Alverno’s eight “abilities”. These outcomes must be demonstrated in a digital portfolio of students’ ongoing achievements. Using non-graded assessment in this way was an institutional decision and clearly required considerable up-front planning but the pay-off has been a far less uncertain process for both students and those who will assess the work (Ehley, 2006).

**Points of reference and grading criteria**

Judging and grading implies that what is being observed is being compared with something. The meaning of grades is embodied in both the criterion and the points of reference and, theoretically, they are selected based on the purpose or intentions of a particular assessment task.

Points of reference can be of three types:

- **Pre-established criteria**: in which the assessor asks, “Did the student performance or learning product demonstrate or address the criteria for which the task was established?”

- **Pre-determined behavioural norms**: in which the assessor asks, “How does the student performance or learning product compare against established norms for this particular level of students?”

- **Ideographic**: in which the assessor asks, “How does the performance or product measure against this student’s earlier performances or products?”

Actual practice demonstrates that the points of reference that academics use are not always clear-cut and rational. Those identified in the cognitive processes of experienced academics as they assess include:

- products of other students
- recall of classroom events and conditions
- broad pedagogical objectives and the specific intended learning objectives
- knowledge of content
- recall of prior assessment events
- incrementally developed construct based on assessor’s perceptions of form, process and content cues in their students’ texts
- performance standards

**Using generic assessment rubrics**

Grading rubrics are helpful for articulating and communicating performance expectations and standards. Their use adds value in a number of ways:

- to guide the unit design
- to communicate expectations to students
- to provide students with an idea of where they sit in a framework of orderly development towards increased expertise in a learning domain
- as a peer and self evaluation tool
- as an aid to consistency and accuracy and representativeness in interpreting, grading and reporting learning outcomes across multiple markers.

A useful structure for establishing a rubric identifies the generic capabilities being assessed, and the differential levels of attainment for each, for example, developing, functional, proficient, and advanced.

<table>
<thead>
<tr>
<th>developing</th>
<th>functional</th>
<th>proficient</th>
<th>advanced</th>
</tr>
</thead>
</table>
| - not yet at the basic level of expectations  
- some minimal desired features may be present but not enough to pass  
- may be enough to ask for further work and resubmission.  
- meets basic requirements at pass level  
- can be carried out in part without support  
- but some guidance may still be needed  
- may be high degree of reliance on authority  
- little translation or integration of concepts into students own language or existing knowledge schema  
- exhibits independence, translation, integration and application (relational knowledge)  
- competently analyses and applies conceptual knowledge to novel contexts  
- may correspond to a credit grade  
- performance beyond core expectations  
- highly independent, creative, critically reflective, generative and transformative  
- uses evidence to formulate defensible personal viewpoints and hypotheses and generate new ideas  
- may correspond to a distinction or high distinction. |  |  |  |  |
A typical way of setting out a rubric is in a grid format relating levels of attainment to attributes (such as “discipline knowledge and understanding” and “psychomotor skills and procedures”). Within each of the “cells” in such a grid, the statements serve as descriptors of the level of attainment on each criterion. As an example, the excerpt above (from the comprehensive generic rubric developed by Orrell, 2003) provides example descriptors relating to each level of attainment for the broad attribute of “discipline knowledge and understanding”.

### Consistency, accuracy and representativeness in judging and grading

Bachor, Anderson, Walsh & Muir (1994) suggest that rather than being dominated by a concern for grading validity and consistence on a single test at one moment in time, the focus should be an overall concern to achieve:

- **Representativeness**: this concerns questioning the meaningfulness in the information the student is required to generate and the extent to which the task reveals the student’s cognitive activities
- **Accuracy**: this concerns mapping a student’s typical performance related to clearly outlined criteria
- **Consistency**: This concerns using consistent established criteria but in tasks that best suit individual students, acknowledging that not all students demonstrate their learning in the same manner

### Factors that interfere with judgement and interpretation

Even where assessors have been trained to recognise the subjectivity in their grading processes, and to ignore influences that might interfere with making good judgements, factors such as the following have been shown to have a profound impact on the grades assigned to students’ learning products and performances:

- the visual appeal of the assignment’s presentation
- the quality and legibility of the student’s handwriting and drawings
- the correctness of grammar and spelling
- the quality of the introductory paragraph alone
- the quality of the other papers being assessed (especially the five preceding papers)
- the teacher’s own knowledge and expectations of particular students based on classroom events
- the teacher’s own “assessment personality”, for example, the tough grader or the encourager of students
- the teacher’s own beliefs about grading and education
- the teacher’s experience in grading, for example, less experienced assessors tend to focus on transmission of content, whereas more experienced assessors tend to focus on learning and transformation

### Strategies to enhance grading reliability

The reliability of grading can be significantly improved with careful planning ahead aimed at reducing the confounding effect of some of the above interferences. Some strategies to improve reliability are:

1. **Consistency in grading criteria**
2. **Training and calibration of assessors**
3. **Use of criteria sheets**
4. **Standardised response options**
5. **Feedback and reflection on grading processes**

---

Podcast: Dr Mitch Parsell talks about using instructional rubrics to engage students in assessment.

Visit the Macquarie University Learning and Teaching Centre website at [www.mq.edu.au/learningandteachingcentre/for_staff/engage_students/parsell2/index.php](http://www.mq.edu.au/learningandteachingcentre/for_staff/engage_students/parsell2/index.php)
• multiple marking of the same paper by either the same assessor or by two different assessors
• blind marking of papers, that is, marking papers without knowing the name of the student
• establishing and maintaining standards through the use of model answers
• annotating model answers to identify performances of different levels on specific criteria
• assigning markers to mark the same question in assignments or tests composed of multiple sections
• involving neutral external examiners and assessors
• using computer-aided marking, for example, with machine-readable multiple-choice quizz sheets, or online automated marking
• avoiding sorting assessment products into predicted grade categories prior to marking and assigning grades

Managing grading

Effective management of grading is very important, and should include at least the following:
• a failsafe procedure for recording the lodgement of student learning products
• clear statements to students about their responsibility to keep a copy of all work submitted until the grading process is concluded for the unit
• an orderly filing and storage system to manage students' submitted work
• a failsafe system for storing student's grades for their work (for example, with a template on your computer or in a record book which could be audited if required)
• your own allocation of time to reviewing the grade distribution and your impressions of how students have managed the task immediately after you have finished grading (individually and in your team/department)
• a routine recording of reflections from reviewing the grade distribution: For example, you might consider what can be learned to improve it for next time. What factors have influenced any unexpected results? What other information should the head of department, head of Faculty or examinations committee have to understand the grade outputs for the unit?
• a program-based or department-based moderation process which is collegial, educative, developmental, not punitive, and focusing on successes and effective practices as well as providing support for improvement where practices have not been so effective.

References and resources


