Overview

Grading is the process of interpreting students' learning products and performance for the purposes of:

- reflecting where students stand in relation to an orderly development of competence
- informing students and teachers of students' current level of learning, and of what they need to do to improve it.
- combining a grade with other grades to meet administrative requirements for awarding grade levels for students' overall performance.

Grading is a high-stakes activity. Students use the results to define themselves as learners. Grading is also highly subjective; interpretation relies heavily on the wisdom of practice.

Expert assessors are highly skilled in interpreting and grading students' performances and products. They need to:

- possess detailed knowledge of their discipline, of curriculum intentions and of learners and their diverse backgrounds
- have detailed knowledge of assessment options, and understand the limitations of these options
- be very clear as to the purposes of the assessment
- have access to a repertoire of meaningful approaches that have been intentionally developed for the interpretation of students' learning performances
- be aware of contextual influences on their practice, the limitations of their own interpretations and judgments, and the ethical and practical implications of the way they conduct grading.

When to use

When planning an assessment-as-learning paradigm, you should consider the questions "When should I use grading?" and "How should I interpret learning, in this course?" Ought students' learning products and performances to be graded at all? Consider Alverno College's assessment-as-learning process—see the bottom of the Strategies section of this page. Do you want to implement something similar instead, scaling it to meet the requirements of your suite of assessments?

Once you have decided to use grading, it's also important that you address the question "How should students be
Assessment-as-learning challenges traditional assessment methods such as objective examinations. It recommends alternative assessment practices such as **authentic assessment**, **standards-based assessment** and performance-based assessment.

Assessment-as-learning grading entails taking account of:

- inclusive assessment
- graduate capabilities
- constructivism
- instructional strategies
- formative and summative evaluation
- rubrics
- peer and self-review and assessment.

**Benefits**

The benefits of standards based assessment have "strong educational and ethical underpinnings" (Sadler, 2005):

- Students are graded on the quality of their work against standards or criteria. Bias cannot arise from comparison with other students' work.
- Students know the criteria against which they will be judged. The primary purpose of criteria-based assessment and grading is to make students aware, before they complete assessable work, of the quality of the work expected.
- This awareness enables them to shape their work appropriately to the standards expected, and assures them that assessors will not be influenced in judging their work by how other students perform, or by their own previous performances.
- When standards are established and communicated before and during the course - to students, all assessors and anyone reviewing the grade distributions—they will function properly for both formative and summative assessment (Sadler, 2005).

**Challenges**

The four fundamental challenges to interpreting and grading assessment-as-learning include:

- understanding the concept of a standard
- developing the standards
- developing ways to explicitly communicate the standards to students and staff
- being a proficient user of standards based assessment and grading (Sadler, 2005).
Strategies

It's usually the course or class teacher who interprets and assesses students' learning performances. Teachers can, however, bring a number of alternative contributors to the process of assessing. For example, they might involve:

- external examiners
- expert professionals and community representatives
- computers (automated assessment)
- other teaching colleagues
- students assessing themselves
- students assessing their peers.

Students develop self-critical and independent learning when you involve them in interpreting and judging their own learning performance. With proper training and support, their interpretations and judgments about their own and their peers' learning performances are marginally more consistent and reliable than those of multiple sessional tutors. Being involved in the process of assessment is a useful learning activity in its own right.

The nature and purpose of the assessment task will determine who will contribute to the assessment, and how they will determine the final grade. A grade is a single indicator of the standard of a student's work, but multiple interpretations of a student's work can contribute to deciding it.

In the overall management of assessment processes, allow scope for moderating final grades to ensure that they accurately represent each student's demonstrated capabilities and performance.

Why do we grade?

We can organise grading processes around at least 3 different goals (Wolff 1969):

- **criticism**: analysing a product or performance for the purpose of identifying and correcting its faults or reinforcing its excellence
- **evaluation**: measuring a product or performance against an independent and objective standard of excellence, to indicate whether a person is professionally qualified
- **ranking**: comparing individual students' performances one with another to decide specific one-off questions such as who will receive scholarships.

In university departments, ranking-focused grading activity produces the greatest anxiety and provokes the most controversy and opposition. Yet it advantages learning the least, sometimes not at all. Boud and Associates (2011) argue that "while marks and grades may provide a crude tracking measure of how well students are doing, they do not help students move beyond their present standard of performance."
Points of reference and grading criteria

When you interpret and grade, you are comparing what you observe with one or more criteria, and points of reference, based theoretically on the purpose or intentions of a particular assessment task.

Points of reference can be of three types:

- **pre-established criteria**: "Does the student performance or learning product demonstrate or address the criteria for which the task was established?"
- **pre-determined behavioural norms**: "How does the student performance or learning product compare with established norms for this particular level of students?"
- **ideographic**: "How does the performance or product compare with this student’s earlier performances or products?"

In practice, experienced academics' points of reference are not always clear-cut and rational. They can include:

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

Representativeness, accuracy and consistency

Bachor et al. (1994) suggest that assessors should focus not on the grading validity and consistency of a single test, but on the achievement overall of representativeness, accuracy and consistency. When they seek:

- **representativeness**, assessors question the meaningfulness of the information the student has generated and the extent to which it reveals the student's cognitive activities
- **accuracy**, they map a student's typical performance against clearly outlined criteria
- **consistency**, they use consistent, established criteria, but in tasks that best suit individual students, acknowledging that not all students demonstrate their learning in the same manner.
Generic assessment rubrics

Use grading rubrics to articulate and communicate performance expectations and standards. They add value in a number of ways. They can:

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

To begin creating a rubric, identify the generic capabilities being assessed and the differential levels of attainment for each, as shown in Figure 1.

Figure 1: Example levels of attainment in a rubric

<table>
<thead>
<tr>
<th>DEVELOPING</th>
<th>FUNCTIONAL</th>
<th>PROFICIENT</th>
<th>ADVANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not yet at the basic level of expectations</td>
<td>Meets basic requirements at pass level</td>
<td>Exhibits independence, translation, integration and application (relational knowledge)</td>
<td>Performance beyond core expectations</td>
</tr>
<tr>
<td>Some minimal desired features may be present but not enough to pass</td>
<td>Can be carried out in part without support but some guidance may still be needed</td>
<td>Competently analyses and applies conceptual knowledge to novel contexts</td>
<td>Highly independent, creative, critically reflective, generative and transformative</td>
</tr>
<tr>
<td>May be enough to ask for further work and resubmission</td>
<td>There may be a high degree of reliance on authority</td>
<td>May correspond to a credit grade</td>
<td>Uses evidence to formulate defensible personal viewpoints and hypotheses and generate new ideas</td>
</tr>
<tr>
<td></td>
<td>Little translation or integration of concepts into students' own language or existing knowledge schema</td>
<td></td>
<td>May correspond to a distinction or high distinction</td>
</tr>
</tbody>
</table>
the statements serve as descriptors of the level of attainment for each criterion. In the example row (from Murray-Harvey, Silins & Orrell, 2003), descriptors relate to each level of attainment for the broad attribute "discipline knowledge and understanding".

**Figure 2: Example levels of attainment for one criterion**

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th>DEVELOPING</th>
<th>FUNCTIONAL</th>
<th>PROFICIENT</th>
<th>ADVANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discipline knowledge and understanding</td>
<td>Limited understanding of required concepts and knowledge</td>
<td>Encyclopaedic discipline knowledge; accurately reproduces required facts and definitions</td>
<td>Exhibits breadth and depth of understanding of concepts in the knowledge domain</td>
<td>Exhibits accurate and elaborated breadth and depth of understanding of concepts in the knowledge domain</td>
</tr>
<tr>
<td></td>
<td>Inaccurate reproduction of text and lectures</td>
<td>Shows adequate breadth, but limited depth of understanding of basic concepts</td>
<td>Uses terminology accurately in new contexts and has transformed the ideas to express them appropriately in own words</td>
<td>Shows understanding of how facts are generated</td>
</tr>
<tr>
<td></td>
<td>Does not translate concepts into own words</td>
<td></td>
<td>Aware of limits of own understanding</td>
<td>Appreciates the limited and temporary nature of conceptual knowledge in the discipline or field</td>
</tr>
</tbody>
</table>

**Factors that interfere with judgment and interpretation**

Even where assessors are trained to recognise the subjectivity in their grading processes, and to ignore influences that might interfere with their making good judgments, the following factors have been shown to affect profoundly 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 5 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

You can significantly improve the reliability of your grading if you plan how you will reduce the effect of some of the above factors. Some strategies are:

• establishing and maintaining standards by using model answers to benchmark standards at different grades
• annotating model answers to identify performances of different levels on specific criteria
• avoiding sorting assessment products into predicted grade categories prior to marking and assigning grades
• blind marking of papers, that is, marking papers without knowing the name of the student
• multiple marking of the same paper by either the same assessor or by two different assessors
• 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 quiz sheets, or online automated marking.

Managing grading

To effectively manage grading, at the very least you should develop and implement:

• a failsafe procedure for recording the lodgement of student learning products. For example, formal date-stamping on receipt of hard copies submitted to the departmental office, or use of an online submission process
• clear statements to students about their responsibility to keep a copy of all work submitted until grading is concluded for the unit
• an orderly filing and storage system for students’ submitted work
• a failsafe system for storing students’ grades, such as a spreadsheet or record book that can be made available for audit if required
• a system for allocating time, immediately after grading, to review the grade distribution and any impressions of how students managed the task. Do this individually and as a team or department.
• a program-based or department-based moderation process that is collegial, educative, developmental rather than punitive, and focuses on successes and effective practices as well as providing support for improvement where practices have not been so effective
• a routine recording of reflections after review of the grade distribution. For example, respond to these
questions:
○ What can we learn to improve the assessment process for next time?
○ What factors have influenced any unexpected results?
○ What other information should we provide to the head of department, head of Faculty or examinations committee, so that they can understand the grade outputs for the unit?

Use technology

Many learning technologies support the interpretation and grading of student work. On the UNSW TELT Gateway you can find many practical guides to particular technologies, such as the grading functions within the Moodle LMS.

Standard word processing tools can support online marking. For example, assessors can develop and share standardised feedback comments relating to an assessment rubric, and select and insert these comments using auto-text and keyboard shortcuts. (Turnitin's GradeMark tool is an automated version of this.)

You can put to new uses technologies you are already using to support student learning engagement and assessment. For example, audience response systems (clickers) have been used effectively by teaching teams in large class settings to achieve better grading consistency when applying assessment criteria to examples of student work. At a group marking session, assessors can use clickers to assign ratings, display all ratings anonymously and then discuss any discrepancies (Cathcart & Neale, 2010).

An example of effective use of negotiated outcomes

At Alverno College in Wisconsin, assessment is performance based; self-assessment is integral to the process, as is teacher, peer and external assessment. Instead of assigning grades, teachers and students establish performance standards in clearly defined profiles of desired learning outcomes, organised according to Alverno's eight "abilities". Students must demonstrate their achievement of these outcomes in an ongoing digital portfolio.

Using non-graded assessment in this way was an institutional decision and clearly required considerable up-front planning. The pay-off has been a far less uncertain process for both students and assessors.

Linda Ehley's 2006 doctoral thesis about the Alverno system is available online.

Additional information

Further readings


**Acknowledgments**

The contributions of staff who engaged with the preparation of this topic are gratefully acknowledged.

This topic was inspired by a resource developed for the Macquarie University Assessment Toolkit.