"The aim of teaching is simple: it is to make student learning possible."
Paul Ramsden, Learning to teach in higher education (1992:5)
Content

CONTENT ........................................................................................................................................... 1
PURPOSE AND PROCESS OF ASSESSMENT .............................................................................. 2
FAIR ASSESSMENT ..................................................................................................................... 2
QUANTITATIVE ASSESSMENT STANDARDS .............................................................................. 3
    1. Validity ................................................................................................................................. 3
    2. Reliability ............................................................................................................................ 3
QUALITATIVE ASSESSMENT STANDARDS ............................................................................... 4
    1. Credibility ............................................................................................................................ 4
    2. Dependability ....................................................................................................................... 4
    3. Confirmability ...................................................................................................................... 4
MODERATION PROCEDURES ......................................................................................................... 4
MODERATION STRATEGIES ........................................................................................................... 5
    1. Make marking/assessment criteria explicit ................................................................. 5
    2. Distribute marking keys/guides ....................................................................................... 5
    3. Use objective test questions ............................................................................................ 5
    4. Implement single marker assessment ........................................................................... 6
    5. Implement second-marking .............................................................................................. 6
    6. Employ double-blind marking ....................................................................................... 6
    7. Employ panel marking ....................................................................................................... 7
    8. Implement anonymous assessment ................................................................................ 7
    9. Apply statistical moderation ............................................................................................ 7
   10. Undertake external moderation ....................................................................................... 7
STATISTICAL MODERATION ........................................................................................................... 7
ASSUMPTIONS AND FURTHER PROVISOS ............................................................................. 10
LIST OF REFERENCES ................................................................................................................... 10
Assessment of student work has a number of purposes, one of which is to differentiate among students in relation to their learning and achievement of the learning outcomes. This purpose is one reason for summative assessment and is the focus of this document.

Whether summative assessment is criterion-referenced (judged against a set of absolute standards) or norm-referenced (judged against the performance of the whole group) the instruments and processes must be reliable. That is, the instruments and processes must consistently identify and differentiate higher achieving students from lower achieving students.

Moderation means a set of processes which aim to ensure that:

- Assessment tasks reflect the learning outcomes and are set at the correct level, and
- The marks awarded are accurate, and consistent with the assessment criteria and marking scheme.
- This document specifically refers to the second of these processes. The appropriate alignment of assessment tasks with learning outcomes is dealt with in the booklet, Teaching and Learning at Curtin (http://lsn.curtin.edu.au/publications/tlc.html).

Assessment must be fair and equitable i.e. all students must be given an equal chance to have their achievement of learning outcomes recognised. Some of the principles involving fairness for all are:

- Students can expect fair reward for quality time on task. The weighting of assessment tasks, therefore, reflects both the importance of, and the time expected for students to achieve, the learning outcomes.
- Student and staff workloads are given due consideration in the scheduling and design of assessment tasks.
- Plagiarism is minimised through explicit education, appropriate monitoring of academic honesty and careful design of assessment tasks.

Fairness in terms of assessment validity is also important but only briefly dealt with in this document.

Equitable assessment means that:

- Assessment practices accommodate individual and group diversity.
- Assessment comprises multiple and diverse tasks to cater for students’ different learning styles. Where possible and appropriate, there is provision for student choice in assessment tasks.
- Reasonable accommodations are extended to students with a verified disability, medical or other condition. Accommodations can include extra time, ergonomic provisions, scribes or alternative assessment tasks that measure achievement of the learning outcomes.
Assessment can be unfair if there is differential treatment in marking and grading. This can occur in a number of ways:

- Markers apply marking guidelines inconsistently.
- Students are marked on criteria that are not explicit in the marking guidelines.
- Markers gradually change or deviate from standards of marking over time.
- Markers are systematically biased i.e. always harder or easier than others in interpreting the marking guidelines.

Whatever the cause, the result is that some students will be favoured and others suffer adverse discrimination. The assessment is said to lack comparability.

Moderation is the process by which fair assessment is assured. It is therefore a responsibility of the unit coordinator (in particular) and all staff marking students’ work to implement strategies that ensure assessment comparability. In most cases, just a few strategies can make a large difference to the quality of marking practices, and to staff and students’ perceptions of fair treatment.

### Quantitative Assessment Standards

1. **Validity**

   Assessment of a student’s knowledge and skills usually results in a mark or a grade that represents the knowledge, skills and abilities being assessed. Validity refers to the extent to which that mark (grade) measures what it claims to measure. Does it measure the student’s achievement of specific learning outcomes? Is it a measure of the current state of their knowledge? For example, a mark based on a student’s recall of content knowledge is not a valid measure of the student’s higher order skills, such as knowledge application. An essay examination might be a measure of students’ essay writing skills rather than their ability to apply knowledge of the discipline. In education, the purpose for which the mark (grade) is being used should also be submitted to validation processes. For example, a mark that represents a student’s knowledge of the discipline content is not valid as a measure of their future ability to work, for example, in a research team.

2. **Reliability**

   Reliability refers to both the accuracy and precision of measurement. Different tests of a student’s particular knowledge or skills, if administered independently of one another, should all give the same result. Similarly, two different assessors should arrive at the same conclusion about a given student’s learning. The three facets of reliability refer to (1) the basic rationale (can the knowledge and skills to be assessed be translated into a measurement?), (2) the procedures for data collection (the assessment tool, its administration and marking) and (3) the statistical procedures following (what we subsequently do with the numbers).
Qualitative Assessment Standards

The following standards are adapted from Guba & Lincoln’s (1989) Fourth Generation Evaluation standards:

1. Credibility

Assessment is credible when the form of assessment is aligned closely with learning outcomes being assessed. Authentic assessment, where students carry out a task that represents a real-world situation, is preferred to one in which there is little correspondence between the two. This is the counterpart of validity referred to above in that it pre-supposes that assessment is focused on student demonstration of specific learning outcomes. Assessment is credible where it is based on detailed evidence, judgements made by two or more people, or evidence derived from different but contributing tasks.

2. Dependability

Dependability refers to the assessment process. Assessment is dependable when subjective assessment methods are applied consistently and are stable over time. That is, the assessment method must be applied as much as possible in the same way and under the same conditions for the duration of the assessment period. Assessment is also dependable if those participating in it (staff and students) agree that the process is a fair and reasonable test or assessment of achievement of the learning outcomes. The use of scoring rubrics can promote dependable assessment.

3. Confirmability

Confirmability refers to the quality of judgements derived from the assessment process. Assessment is confirmable when an audit trail (documentary or electronic) is maintained to enable back-tracking to original criteria-based judgements. Scoring rubrics, analytical or holistic, are a primary means of doing this. Other records may include observational notes, annotated documents or other records of participation or achievement. Records or evidence must conform to the required format and be collated in the same way and to the same level of detail for each student. Confirmability is enhanced where there is student agreement that the judgement about their performance is fair and accurate.

Moderation procedures

Schools/departments should clearly document procedures for moderation. The key elements of the information should also be included in student handbooks. These procedures should specify:

- The method/s of moderation to be used (which should be appropriate to the potential for variability among markers).
- Particular arrangements for moderation of practical assessment such as oral examinations, presentations, music or drama performance, and laboratory work etc. Where opportunities for effective moderation of practical work are limited, the weighting given to this form of assessment in the overall programme assessment strategy should be carefully considered.
• The basis for sampling assessed work for moderation where a large cohort has been assessed.
• Particular arrangements for moderation of work that is marked by sessional staff or postgraduate research students, or any new members of academic staff.
• The procedure for resolving significant discrepancies between multiple markers.
• The provision and retention of evidence to demonstrate that internal moderation has taken place either through comments made on scripts or recorded separately.

---

**Moderation strategies**

It is not necessary to employ every strategy below, but to consider what are the likely causes or contributors to a lack of comparability in your area and to make changes that are likely to have the greatest impact. What is your situation?

• Assessment tasks are not common to all students
• Assessments involves multiple markers
• Assessment is largely subjective
• One person marks many papers
• Assessment occurs on different campuses

1. **Make marking/assessment criteria explicit**

   Assessment criteria describe the key characteristics of differing standards of performance and are usually standards defined by marking or classification bands. The development and dissemination of appropriate assessment criteria is a key element of consistent assessment procedures. For criterion-based assessment, grading criteria (broad statements of student achievement associated with each grade band and/or level) must be established prior to students attempting the assessment. Provide students with clear marking criteria or scoring rubrics prior to them attempting any assessment task.

2. **Distribute marking keys/guides**

   Marking keys must be distributed to all markers for comment prior to the assessment being attempted by students, in time for feedback and concerns to be addressed.

3. **Use objective test questions**

   Objective test questions are those with answers that are unique or have a limited range of possible answers. Examples include multiple-choice questions involving calculations, where only a correct numerical answer is required, and questions with single word or simple phrase answers. Despite compelling views that no assessment task can be completely objective, the use of tests or items within tests that are ‘objective’ promotes comparability through reducing marker bias or inconsistency, even with multiple markers.
4. Implement single marker assessment

All assignment submissions from students in a particular unit are marked by the same marker according to the approved marking scheme\(^1\). This method assumes that all students complete the same assessment task. It can also be applied to individual components of assessment tasks provided that all students complete the same components (for example, if all students are expected to attempt all questions in an assignment or examination).

From time to time through the assessment process, the single marker should review papers marked earlier (e.g. the day before) to ensure that the marking standard has not changed.

When a single marker is new to the task or not part of an assessment team, for example a sessional academic, the unit coordinator or another member of the assessment team should independently re-mark a representative sample of submissions marked by the single marker from across the grade ranges (for example two from the each of the pass/fail boundary, the middle range and the top of the range) in order to confirm that the marker is marking according to the agreed standard. If a single marker is not maintaining a consistent standard, the unit coordinator should ensure that any work already assessed by that marker is re-marked to the agreed standard and that ongoing assessment of submissions assigned to that marker are also marked to the agreed standard.

5. Implement second-marking

Second-marking involves a sample of marked assessment items being marked a second time by another person. It may involve remarking or simply a verification of the marks allocated by the first marker. It is used to validate assessment standards across a group of markers. The unit coordinator or other experienced marker may second-mark a sample of the assessed work from each marker in a team across the mark range (for example two from the each of the pass/fail boundary, the middle range and the top of the range).

The number of items chosen for second-marking should be representative. The usual is to choose a number equal to the square root of the number of assessed items, or five items, whichever the larger (e.g. if there are 36 students, second-mark six of them, if there are 16 students, second-mark five of them). There should be clear criteria for remarking all the work if there are sufficient inconsistencies or inaccuracies within a sample.


In this method, two markers independently mark the work of each student. This is preferred for larger assessment tasks such as research projects or where the assessment involves more subjective judgements. Consistency limits and ways of resolving differences must be prescribed. The following is an example:

*If the marks given by the two markers for a particular student are within 10% of each other, the final mark is taken as the average. If the marks differ by more than 10%, a third marker examines the work. The markers may then discuss the marks to arrive at a consensus result or the median of the three marks is taken as the final mark.*

---

\(^1\) Marking schemes are discussed elsewhere in this document.
7. **Employ panel marking.**

Panel marking involves independent but concurrent assessment by two or more markers. It may be used for oral presentations, performances, exhibitions or other transient assessment work. Audio-visual recording of the assessment might also be undertaken. Comparability of assessment is enhanced if the composition of the panel is the same for all students doing a particular assessment task. If the panel is not the same for all students then one member of the assessment panel should chair all panels in order to promote comparability.

8. **Implement anonymous assessment**

Anonymous assessment attempts to remove marker bias. It only works where the identity of students is not revealed in other ways e.g. through writing styles, references in the text etc. Bias can result, unconsciously or not, from prior dealings with different students and be based on attitude, behaviour, gender, race, disability etc.

9. **Apply statistical moderation**

This process involves:

- checking that different assessors have marked students’ work in accordance with predetermined criteria and marking guidelines, and
- where different markers have been shown to have employed different marking standards, adjusting students’ marks accordingly.

(See section below)

10. **Undertake external moderation**

Consistency and quality of assessment standards can be established by having some assessment tasks common to those in another course and/or university. Collaborative or cross-marking with assessors in other courses or universities contributes to maintenance of assessment standards as well as students’ perceptions of impartiality. Employment of external markers can also ensure comparability of standards.

---

**Statistical moderation**

At times, despite having clear marking criteria, different markers can be systematically biased i.e. be hard markers or easy markers. This leads to student mistrust of the fairness of a marking process. While statistical moderation is a recognised process, it should be a last resort when moderation procedures have failed or have simply been impractical to implement. How, then do we ascertain whether the differences in marks distribution of different classes/tutorials are due to actual differences in performance of the students, or differences between markers in applying the marking criteria?

The following statistical moderation procedure pre-supposes that there is at least one **substantial task that is assessed using recognised and effective moderation procedures**. This is usually an examination. The marks distribution of this assessment task is used as the standard against which to moderate (adjust) the distribution of other assessment marks for each class.
Example:
Assume a unit has three tutorial classes, 1, 2 and 3. The mark distribution for an examination (moderated) and an assignment (un-moderated) are shown below:

<table>
<thead>
<tr>
<th>Teacher/class</th>
<th>Class exam mean (%)</th>
<th>Class exam standard deviation</th>
<th>Class assignment mean (/10)</th>
<th>Class assignment standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>13</td>
<td>5.3</td>
<td>1.3</td>
</tr>
<tr>
<td>2</td>
<td>79</td>
<td>11</td>
<td>7.7</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>10</td>
<td>7.7</td>
<td>0.90</td>
</tr>
</tbody>
</table>

The box and whisker plots that follow show the three classes’ distribution of marks for the exam (Figure 1) and the assignment (Figure 2). Class 3 has assignments marks that appear to be too high (i.e. teacher 3 has probably marked the assignments too leniently). In this case statistical moderation of marks may be appropriate. In this process, ALL students’ assignment marks will be adjusted so that each class’s assignment mark distribution matches the respective class’s exam mark distribution.

Figure 1: Box and whisker plots showing examination mark distributions for three classes (1, 2 & 3).

Figure 2: Box and whisker plots showing assignment mark distributions for three classes (1, 2 & 3).
The following algorithm is applied to each student’s assignment mark to either increase or decrease it. Each student’s adjusted assignment mark \((A') =\)

\[
\left( \text{Ass/t mark} - \text{Class ass/t mean} \right) \left( \frac{\text{Class exam st.dev}}{\text{Class ass/t st.dev}} \right) + \left( \text{Class mean exam} \right) \times \left[ \frac{\text{Ass/t total}}{\text{Exam total}} \right]
\]

i.e \(A' = \left( A - \bar{A}_c \right) \left( \frac{\text{SD}_{c,E}}{\text{SD}_{c,A}} \right) + \bar{E}_c \) \times \left[ \frac{A_{\text{tot}}}{E_{\text{tot}}} \right]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ass/t mark</td>
<td>Each student’s assignment mark</td>
</tr>
<tr>
<td>A'</td>
<td>Adjusted assignment mark</td>
<td>Each student’s adjusted assignment mark</td>
</tr>
<tr>
<td>(A_{\text{tot}})</td>
<td>Ass/t total mark</td>
<td>Max number of marks for the assignment (in this example 10)</td>
</tr>
<tr>
<td>(E_{\text{tot}})</td>
<td>Exam total mark</td>
<td>Maximum number of marks for the examination (in this example 100)</td>
</tr>
<tr>
<td>(\bar{E}_c)</td>
<td>Class exam mean</td>
<td>Average mark for that class on the examination (different for each class)</td>
</tr>
<tr>
<td>(\bar{A}_c)</td>
<td>Class ass/t mean</td>
<td>Average mark for that class on the assignment (different for each class)</td>
</tr>
<tr>
<td>(\text{SD}_{c,E})</td>
<td>Class exam st.dev</td>
<td>Standard deviation of the examination marks for the class (different for each class)</td>
</tr>
<tr>
<td>(\text{SD}_{c,A})</td>
<td>Class ass/t st.dev</td>
<td>Standard deviation of the assignment marks for the class (different for each class)</td>
</tr>
</tbody>
</table>

Using this algorithm for the example described previously, the assignment marks for class 1 will be adjusted upwards by a small amount, class 2 will have minimal adjustment, and the assignment marks for class 3 will be adjusted downwards. The rank order of marks within each class is preserved. Figure 3 shows the adjusted assignment mark distributions, which now closely resemble the examination mark distributions. This final process must be completed to check that no errors have been made in the re-calculation of marks.

*Figure 3: Box and whisker plots showing adjusted assignment mark distributions for three classes (1, 2 & 3).*
This process is much simpler if both exam and assignment marks have the same total i.e. are percentages. The algorithm then reduces to:

\[ A' = (A - \bar{A}_c) \left( \frac{SD_{E,A}}{SD_{E,A}^2} + E_c \right) \]

### Assumptions and further provisos

- If other moderation procedures have been employed successfully, statistical adjustment of marks should not be necessary. **Statistical moderation should be a last resort.**

- This process assumes that the difference between the classes’ initial assignment marks arise from differences between the different markers’ interpretation or application of the same marking key.

- The process assumes that the examination performance for each class is not atypical and the examination is well-moderated.

- The process is sensitive to small class numbers. As a rough guide, class numbers should be 15 or more.

- This process may produce adjusted marks outside the limits i.e. a student may get an adjusted mark of 10.5 out of 10!

- If it has been necessary to implement statistical moderation, the unit coordinator must initiate a review of moderation procedures prior to the next time the unit runs.

- Seek advice with this procedure if you are unsure about the calculation of adjusted marks.

### List of references