Moderation
CLD

What is it about?
Moderation is about ensuring consistency and fairness of the marking and grading processes. It involves scrutiny of assessment tasks and marking criteria in relation to learning outcomes, as well as scrutiny of samples of student work across the range of marks. Moderation assists markers to align their marking standards with those of other markers. It may involve adjustment of individual student marks as a result of remarking, or a general adjustment of a cohort’s marks by applying a scaling process.

Why is it important?
Moderation enhances confidence in assessment outcomes. As assessment defines academic standards and is used to certify student achievement, it is vital that we have a quality assurance process in place to ensure consistency between assessors and to provide equity for students. The quality of our assessment practices ultimately defines the quality of our graduates.

What does the ECU policy say?
The ECU [Moderation of Unit Outcomes Policy] specifies that, for one unit with multiple markers:

1. the unit coordinator should provide all markers with:
   a. a sample marking experience; and
   b. comprehensive marking keys for each assessment showing marks allocations.
2. acceptable strategies to enhance consistency and fairness include:
   a. allocating questions to markers so that the same person marks a particular question for all papers and every paper is marked by more than one marker; and
   b. applying a check marking process.
What are the key principles?

Results from assessment tasks are used to infer achievement. Moderation processes enhance confidence in assessment practices and ultimately in certification of achievement. 

Good moderation processes should result in improved assessment tasks, marking guides and professional marking judgement.

How to carry out moderation

Pre-assessment moderation:
Assessment tasks should be subjected to routine pre-assessment review to ensure: alignment with the unit learning outcomes that are being sampled; focus on higher-order learning; and clarity of marking criteria. Ideally a third person should be asked to read the task and to outline what a good response might look like, and how marks might be allocated. This should then be checked against the marking guide. It is surprisingly easy to inadvertently allocate marks to something that we haven’t explicitly asked for, thereby testing our students’ ability to read our minds before they can answer the question!

Review before marking:
Once student work has been collected a few sample pieces of work can be randomly chosen and marked by all markers. Markers then meet and discuss any discrepancies in marks, adjusting and clarifying the marking guide (and often the task) for future use.

Review during marking (before work is returned to students or marks published):
It is often too late or too awkward to change marks after all the marking is done. It is better to monitor and refine marker performance during the marking. While marking a few samples at the start can eliminate many discrepancies, markers will invariably come across a response that doesn’t quite fit the norm. These student responses should be marked by another marker, preferably without the second marker seeing the first marker’s work. Again, discrepancies are discussed and resolved, and notes taken for future use of the task and marking guide.
Review after marking (before grades are finalised):

If the previous processes have not achieved consistency and fairness, moderation may require scaling of student marks.

Suppose, as a unit coordinator, you receive marks for the same assessment(s) from two different tutors. One has a mean score of 76 (of 100) and the other 82 (also of 100). It might seem that one class had more able students than the other or that one tutor may have marked harder than another. It is also possible that the standard of teaching may have been better in one class leading to the higher mean mark. In such situations, it is difficult to compare the scores.

As a coordinator you may feel tempted to just add or subtract marks from one class or the other to bring them into some comparability. This simple solution distorts the distribution of marks. Similarly just reducing top scores or inflating bottom scores to conform to some pre determined distribution is also not valid. So how do we compare the scores and if necessary adjust them?

If we knew the mean and standard deviations of the two distributions, we could compare these scores by comparing their Z-scores. A Z-score quantifies the original score in terms of the number of standard deviations that that score is from the mean of the distribution. This calculation is easily accomplished with statistics programs like PASW (SPSS). However, it is important to note that a Z-score transformation changes the central location of the distribution and the average variability of the distribution. It does not change the skewness or kurtosis. Once the Z-score is found for each mark the marks can be scaled by nominating a new mean (and keeping the same standard deviation if desired) and multiplying this mean by each Z score to create the new mark.

This method maintains the ranks of scores, enables valid comparisons and does not change the “shape” of the distribution of scores. There are other more complicated statistical procedures that will accomplish the same end but this Z-score transformation is straightforward.