Guidelines on Moderation, Validity and Reliability of Assessment

Parent Policy/Procedures

Assessment in Coursework Units Policy

Assessment in Coursework Units: Grading and Marking Procedures

Purpose

To provide Chief Examiners with guidance regarding the process of moderation to ensure the design of assessment tasks and the subsequent marking processes are educationally and pedagogically sound while administratively feasible and sustainable.

Moderation, Reliability and Validity

Moderation is a quality assurance process that aims to maintain the validity and reliability of the assessment tasks and their marking, which can be understood as:

- Validity of the assessment tasks: Assessment tasks are designed to assess what they are supposed to assess, which are the learning outcomes of the tasks, unit and course. This means even if students find a particular assessment task engaging and performed well, if the task does not address the learning outcomes, it is not valid in the given context.

- Reliability of the assessment tasks: Assessment tasks are designed to be implemented consistently. This means even if the task itself addresses the learning outcomes (i.e. it is valid), if it is too complex in its nature and/or has its scope beyond what is expected to address the learning outcomes, the task may become unreliable as a means of assessment.

- Validity of marking: Assessment tasks are marked in response to what the assessment tasks are supposed to assess (i.e. criterion-based marking; see section 1 of Assessment in Coursework Units: Grading and Marking Procedures). This means even if the criterion-based marking is conducted by a single trained marker using a rubric to maintain the consistency of the marking, if the marking was undertaken against the criteria inconsistent with the learning outcomes, the marking is said to be invalid as assessment. This inconsistency includes the disproportionate allocation of marks for only specific aspects of the learning outcomes.

- Reliability of marking: Assessment tasks are marked consistently, correctly and fairly across different submissions/examinations being assessed by the single marker, as well as across different markers, cohorts and locations. Many assessment tasks require the markers to apply their expert discretion to generate a mark/grade and an effort needs to be made to make sure this subjective variation is eliminated as much as possible.

These guidelines refer to the Monash Assessment Vision and should be read in conjunction with Assessment in Coursework Units: Grading and Marking Procedures. Please note, faculties, schools or departments may provide further guidelines that extend or vary these, tailored to the particular discipline or learning context.

As part of the development of assessment tasks

In order to maintain the validity and reliability of the assessment tasks, Chief Examiners are encouraged to consider the following aspects of the tasks being developed.

Alignment of the assessment task within the unit and its contexts: In order to maintain the validity of the tasks, Chief Examiners are encouraged to first benchmark the learning outcomes of the task both: 1) in the context of the units itself to ensure that there is a suitable range and number of assessment tasks within the unit, and 2) in the context of other relevant units as well as courses in order to ensure that learning outcomes are appropriately aligned both vertically and

horizontally outside the unit. This is in accordance to point 2 of [Development and implementation of an Assessment Regime Procedures](#) set out by the University.

Alignment of the assessment tasks and the skills/knowledge assessed: It is also critical to ensure that a student's ability to complete an assessment task is contingent on the achievement of the desired learning outcomes. If the task can be completed without that learning, then the measure is invalid. For example, if the purpose of a task is to assess content knowledge, but the task's completion is heavily based on the student's' communication skills, then it lacks validity unless the communication skills are part of the learning outcomes and being taught as part of the unit.

Selection of the genre/format of the assessment tasks: In order to maintain the reliability of the tasks, both the students' familiarity with the assessment format/genre, as well as the logistics of the marking need to be considered at this stage. While innovative approaches to assessment are always encouraged to further enhance the quality of learning and teaching, a particular caution needs to be applied when the unit adapts in its assessment task a format of assessment with which not all students are familiar with (such as videos, infographics, posters, social media), unless the development of communication skills in these formats is central to the learning outcomes themselves.. A solid pedagogical and educational justification needs to be established before such a non-traditional assessment format/genre is implemented, so that this can be communicated to both the entire teaching staff members as well as the students (more below).

External factors that may affect the implementation of the assessment tasks: Even if the assessment task is valid and its format justified, if the task incurs unrealistic expectations on the markers' expertise (e.g. if using teaching associates to assess student work not all have the expertise to assess communication skills), or unrealistic logistical requirements for marking during the semester, it is difficult to expect reliable implementation of the assessment task. It is thus recommended that the marking scheme and regimen is developed alongside the development of the actual tasks and their specifications/instructions, and this marking scheme and regimen should include the development of appropriate marking instruments (such as rubrics), training/briefing methods for the markers, and also detailed plan of how long marking will take place, and who will conduct the marking.

As part of the unit's teaching

Since assessment tasks form part of the unit's teaching, it is important to introduce each task to the students in a pedagogically and educationally informed manner, using clear and timely communication methods. This will ensure that students are well-informed of: 1) the validity of the task they are to conduct, and also 2) how they can approach the task, therefore increasing the reliability of the task as a piece of assessment. In so doing, the Chief Examiners are encouraged to consider the following aspects as part of the unit's teaching.

Publication of a written specification of the assessment task and its marking instruments: When students are introduced to an assessment task in the unit, they should be provided with a clear written assignment specification in a timely manner, and this should not be limited to the instructions of the tasks, as specified under point 14.1 of [Development and implementation of an Assessment Regime Procedures](#) set out by the University. Ideally, the assessment instruments aligned with the criteria (including the detailed rubric) are also published with (or shortly after) the assignment instruction in order to provide the students with a clear understanding of the assessment mechanism, including what the assessment is supposed to measure, how the students can demonstrate such learning using the designated format of assessment, and how their deliverable will be assessed.

Formal and informal in-class discussion on the assessment tasks: The provision of a well-prepared assignment specification and the associated marking instrument(s) should also be complemented by further discussion in class, and the students will also need to be taught in a timely and educative manner how they can approach the task effectively, especially if the task expects/requires the students to acquire additional skills in order to complete it. While this can be done formally as part of class activities, such discussion and teaching also often take place in tutorials and laboratory classes, and it is therefore essential that all the teaching staff have clear understanding of both the educational and pedagogical basis of the given task, and be able to explain to the students on such basis when asked any questions.

As part of the marking

As stated earlier, moderation of the marking should be implemented regardless of the number of markers. While it is obvious that the consistency of the making needs to be maintained across markers (inter-marker reliability), the same is true even if there is only one marker, as one still needs to be able to provide evidence to support the claim that marks awarded by that person would be comparable to those if marked by others (intra-marker reliability). In so doing, Chief Examiners are encouraged to consider the following aspects as part of the marking of a given assessment task.

Development and provision of the marking tools and resources: Developing and providing an appropriate set of marking instrument/instruments and resources ensure the validity of marking, and they also help maintain the consistency of the marking. Recommended approaches include:

- developing robust assessment instruments (e.g. rubrics) to articulate what is to be observed as the basis of assessment, hence also minimising variation across the markers;

- involving markers in the design of assessment tasks and relevant assessment instruments, hence also achieving common understanding amongst markers of the educational and pedagogical basis of the assessment; and

- preparing sample answers/submissions from past semesters for consideration and discussion by all staff involved in marking, hence also establishing the consensus amongst markers of the expectations.

Identifying and benchmarking individual markers interpretations of criteria and standards before the marking starts: Once the students' work has been submitted, a set of sample submissions could be marked using the following method(s), in order to identify: 1) patterns and/or tendency unique to individual markers, 2) patterns and/or tendency unique to individual criteria, 3) different interpretations/application of the marking instruments/standards, and/or 4) practical issues with the instruments/scheme (such as time taken to mark each submission, and also the amount/quality of feedback given to each submission).

Even minor changes to the marking rubric can potentially result in the inconsistency of marking for pieces marked before and after changes, and due to different interpretations of the change made by each marker in the teaching team. The divergence of interpretation can occur regardless of the experience of markers so discussion and building of consensus of interpretations within the marking team is desirable. Different markers may also have different expectations about how feedback is to be provided in terms of quantity, style or focus of commentary so it is important to moderate the markers' feedback comments in order to maintain the integrity and pedagogical soundness of the feedback given to students (see Feedback Guidelines on Student Assessments [link to be provided] for more details). All markers should then agree on the ways feedback is provided before the marking commences, and this should constitute part of a consensus marking guide used by all the markers if applicable.

The marking of a set of sampled submissions using the following method(s) will provide a basis for further discussion to benchmark and calibrate each marker's expectations as well as the practical expectations on the markers in order to maintain the consistency and sustainability of their marking. This applies equally to a unit offered using multiple modes and/or at different locations.

- Double marking – where the second marker neither has access to the grades or the comments of the original marker.

- Second marking – where the second marker has access to the comments of the first marker and to the grades awarded by that marker.

- Sample marking – where markers mark a sample of the same assessment tasks independent of each other and then compare the results for consistency. This can be conducted before the marking of the entire submission starts in order to test the validity and the reliability of the marking instruments used, or after the marking, in order to validate the consistency of the marking conducted.

- Consensus marking or peer moderation – where markers come together and mark a sample of the same assessment tasks and address consistency issues as they arise.

- Self-moderation – where a marker is encouraged to check back over scripts that have been marked to ensure that scripts marked earlier and later have consistency in the marking and feedback provided.

Any discrepancies in marking between multiple markers should be discussed, and the outcome should be a refined understanding of the marking criteria or rubric evident in the assessment submission, or exam question response to inform further marking activity.

Identifying unusual mark/grade distributions across individual markers: Once all the submissions are marked, prior to the release of marks/grades to students, it is recommended that all the marks generated by each individual marker are reviewed to identify unusual mark/grade distributions across individual markers. Common methods of identifying such distributions include:

- descriptive statistics – reviewing any unusual variation/distribution of marks;

- spot checks – selecting papers at random and checking their marks;

- examining outlier marks – selecting samples from HDs and Ns; and

- checking borderline assessments (e.g. borderline pass-fail results).

Within the context of criterion referenced assessment it is unrealistic to expect an even bell curve distribution from every single marker, and in some instances it is quite possible to have a particularly high number of HD (or N for that matter) in one or more cohorts where context may vary. Unusual mark/grade distributions across individual markers should trigger a discussion into the context of the cohort and interpretation of assessment standards. They should be validated by the Chief Examiner in consultation with the markers such that a distribution is not contingent on the individual marker. This applies equally to a unit offered using multiple modes and/or at different locations, and the actual marks must be reviewed across the different cohorts of students.

Alternative methods of moderation: In some cases where the above approaches may not be feasible due to the nature and/or the format of the task, the following approaches may be taken to maintain the consistency of the marking.

- Using panel marking to minimise subjectivity, especially for oral, performance, practical or laboratory assessment

- Engaging external moderators (i.e. individuals external to the teaching/marking team) to review a sample of marked assessment

- Assigning markers to mark the same question in assignments, texts or examinations composed of multiple sections

As part of the release of the marks

Returning marks to the students is part of an educative process, and this can be a stressful experience for some students. Hence, marks need to be communicated to the students in an educationally justified and pedagogically sound manner. In so doing, Chief Examiners are encouraged to consider the following aspects.

Consistency in the provision of feedback that accompanies the mark: Effective feedback (see Feedback Guidelines on Student Assessments [link to be provided]) makes the mark meaningful, and it communicates clearly why a student has obtained the mark given, and also how they could improve their work further in future. It is part of the educative process, and plays a significant pedagogical role to shape the students' experience of the teaching and learning in the unit. It is therefore encouraged that the Chief Examiner:

- recognises the consistent provision of feedback both in terms of the amount and quality as part of the moderation of the assessment tasks;

- makes all the teaching staff aware of the educational and pedagogical importance of the feedback that accompanies the mark;

- implements appropriate methods to provide such to the students (e.g. using a well-designed assessment instrument, providing group feedback in class, providing individual consultation hours); and

- actively communicates to students the availability of feedback, the importance of feedback, and how to make use of such feedback. (See Feedback Guidelines on Student Assessments [link to be provided] for more details.)

Teaching staff's understanding of the marking process: The Chief Examiner should ensure that all teaching staff involved are aware of the unit-level protocols to be followed. This includes handling processes such as student complaints or a request for special consideration. These matters need to be dealt with in accordance with the University's and Faculty's relevant policies and procedures in order to maintain the integrity and fairness of the assessment process as a whole. Having a shared understanding of how to respond to these requests amongst the teaching staff prior to the release of the marks assists in maintaining consistency in the assessment process.

As part of the review of the unit

Every assessment task, its marking and the marks should be reviewed generally to further improve the next iteration of the unit. In so doing, Chief Examiners are encouraged to consider the following aspects.

Appropriateness of the assessment task (e.g. level, content and genre/format): This is often indicated by the overall performance of the students for a specific assessment task (e.g. too many Ns), and this also applies to individual criteria. For example, the students may perform well overall in a given assessment task, but they may collectively not perform well against certain criteria. These indicators, together with feedback from students and staff, would inform future refinement of the complexity, scope and alignment of the assessment tasks and/or future adjustment in the coverage and modality of the teaching in the unit in order to address those criteria.

Clarity of assessment tasks and their criteria: This is often indicated by the number and nature of questions the teaching staff commonly received from students during the semester in regard to assessment tasks. If the assessment task's requirements and/or its criteria are unclear, they naturally attract questions from students to seek further clarification. This can also further be validated by the students' performance under the specific criteria which attracted the questions from the students. Such indicators, together with feedback from students and staff, would again inform future refinement of the way the assessment tasks are introduced and communicated to the students. This may also inform future adjustment in the coverage and modality of the teaching in the unit in order to address those criteria.

Effectiveness and sustainability of the marking and feedback mechanism: Feedback from students and staff would inform future refinement of the way the assessment tasks are marked and their results communicated to the students. This would also implicate to the resourcing and training of the markers involved in order to maintain the consistency within the team across different semesters.

The outcome of such a post-assessment review should inform the assessment design and the criteria for marking for subsequent study periods/unit offerings.

Content enquiries: policy-education@monash.edu

Document history

Version number	Date	Description
1.0	February 2018	Noted at LTC meeting 1/2018 item 10.1