CONSISTENCY AND COMPARABILITY OF GRADING OUTCOMES AT UQ

ISSUES PAPER FOR THE ASSESSMENT SUB-COMMITTEE

Assoc. Professor Susan Rowland
Dr Christine Slade
Dominic McGrath

Institute of Teaching and Learning Innovation (ITaLI)
The University of Queensland

August 2016
EXECUTIVE SUMMARY

Data collected over the past eight years show that UQ grading is disparate from program to program and from course to course. In addition, the differences in grading between courses are increasing, rather than decreasing.

Inconsistencies in grading reflect disciplinary norms; however, grading inconsistency poses risks for UQ in terms of undergraduate reward and retention, recognition of students who sit at the ends of the student achievement curves, and graduate employability in a competitive marketplace.

In this paper, we address the current and historical variations in grading at UQ and use the lenses of discipline, year level, and class size to examine how inconsistencies manifest for our students. We examine the risks posed by grading inconsistency and the current regulatory pressures that drive the movement towards assured standards and qualities for graduate evaluation.

We then examine the literature on grading philosophy and practice and link this to grading practice at UQ. In particular, we focus on norm-referenced assessment, criterion-referenced and standards-based assessment, professional judgement, tacit knowledge and disciplinary differences in marking, moderation, and calibration and peer review of assessment practice.

We conclude by suggesting a range of approaches that may improve consistency of grading at UQ. Given that this might not be an acceptable or achievable goal at UQ, we also make suggestions about practices that can give both students and potential employers (i) a better sense of where students lie within their cohort and (ii) a more complete understanding of the skills and experiences obtained during a UQ education.
TABLE OF CONTENTS

EXECUTIVE SUMMARY ........................................................................................................ 3

TABLE OF CONTENTS ........................................................................................................... 4

INTRODUCTION ...................................................................................................................... 6

GRADING AT UQ: POLICIES, PRACTICES, AND POTENTIAL PROBLEMS ......................... 6
THE PURPOSES OF ASSESSMENT ......................................................................................... 6
THE COMPARABILITY AND CONSISTENCY OF GRADING AT UQ ........................................ 8
THE IMPACT ON STUDENTS OF DISCIPLINE GRADING INCONSISTENCIES AT UQ .......... 11
Undergraduate Scholarships ................................................................................................. 11
Honours and Postgraduate Scholarships ........................................................................... 11
University Graduate of the Year Awards .......................................................................... 11
Competition for Employment Beyond Graduation ......................................................... 11
Difficulties for Cross-Disciplinary Study ........................................................................ 11
Inequity for Students Who Apply for Program Transfer .................................................. 11
Course Shopping Based on Perceived Easier Grading .................................................... 12
COMPLIANCE WITH NATIONAL-LEVEL GUIDELINES AND STANDARDS ...................... 12

MAJOR CONCEPTS AND PRACTICES IN ASSESSMENT METHODOLOGIES ..................... 13
NORM-REFERENCED ASSESSMENT .................................................................................. 13
CRITERION-REFERENCED AND STANDARDS-BASED ASSESSMENT ............................... 14
PROFESSIONAL JUDGEMENT, TACIT KNOWLEDGE, AND DISCIPLINARY DIFFERENCES IN MARKING 15
MODERATION, CALIBRATION, AND PEER REVIEW OF ASSESSMENT PRACTICE .............. 16

GRADING PRACTICES AT UQ .............................................................................................. 17
THE USE AND APPLICATION OF UQ GRADE DESCRIPTORS ........................................... 17
THE GROUP OF EIGHT QUALITY VERIFICATION SYSTEM (QVS) ........................................ 18

AN ENVIRONMENTAL SCAN OF PRACTICES USED TO ACHIEVE GRADING EQUALITY AT OTHER UNIVERSITIES ACROSS AUSTRALIA ................................................................. 18

Batchelor Institute of Indigenous Tertiary Education (AUQA) ........................................ 18
The University of South Australia Assessment Moderation Toolkit .................................. 19
The Higher Education Achievement Report (HEAR) ....................................................... 19

SUGGESTIONS AND OPTIONS FOR ADDRESSING GRADING INCONSISTENCY AT UQ ........ 19
ADOPT A NORMATIVE APPROACH TO GRADING WITH A BELL-CURVE OF SET PROPORTIONS .... 19
SET QUOTAS FOR DIFFERENT GRADE OUTCOMES IN A “QUASI-NORMATIVE” APPROACH ...... 19
ASK DISCIPLINES FOR JUSTIFICATION OF THEIR GRADE SPREADS .................................. 20
PUBLICISE GRADE SPREADS FOR COURSES AND PROGRAMS ...................................... 20
MAKE ALLOWANCE FOR STUDENT GRADE OUTCOMES BASED ON THEIR DISCIPLINARY NORMS (OR THE NORMS OF THE COURSES THEY HAVE TAKEN) ...................................................... 20
PROVIDE COURSE GRADE OUTCOMES AND A FINAL TESTAMUR TO STUDENTS WITH THEIR PERCENTILE RANK INCLUDED ALONGSIDE THE GRADE ............................................. 20
PROVIDE OPPORTUNITIES FOR PEER REVIEW AND STANDARDS COMPARISON WITHIN AND BEYOND DISCIPLINES AT UQ ................................................................................................. 21
PROVIDE ALTERNATIVE AND ADDITIONAL NON-GRADED WAYS FOR STUDENTS TO EVIDENCE THE QUALITY OF THEIR ACHIEVEMENTS AT UQ ................................................................. 21
DO NOTHING ..................................................................................................................... 22

CONCLUDING COMMENTS ................................................................................................. 22
REFERENCES ........................................................................................................................................... 23
APPENDIX A: GRADING DATA FROM UQ ......................................................................................... 25
APPENDIX B: ASSESSMENT PHILOSOPHIES COMPARED ................................................................. 32
APPENDIX C: MODERATION FLOWCHART ..................................................................................... 33
APPENDIX D: UQ GRADE DESCRIPTIONS FROM PPL 3.10.07 ...................................................... 34
INTRODUCTION

A comparison of grade spreads between UQ programs from 2008 to 2015 indicates that there is significant disparity in the spread of grades from one discipline to another, from program to program and from course to course. Despite calls for more benchmarking of grading from the UQ Assessment Sub-Committee, the differences in grading between courses increased between 2008 and 2015, rather than decreased. Marking practice can vary widely between disciplines (as defined by the four letter course codes at UQ) (see data in Appendix A). For example, the average fail rate varies from 23% in some disciplines to only 0.1% of students in others. The percentage of high grades (6 and 7) is similarly variable; in some disciplines, it sits around 20% while in others almost 90% of students consistently achieve this mark of excellence. The outlying disciplines (categorised by four-letter course codes) are consistently the same from 2008 to 2015.

These grade distributions occur under the current UQ criterion-referenced assessment system in which “judgements about the quality of students’ performance are made by reference to explicit or predetermined criteria and standards and not by reference to the achievement of other students” (UQ, PPL 3.10.02 Assessment). Despite the explicit statement in policy that students’ achievement should not be compared to other students, UQ uses assessment outcomes to rank students for entry to other programs of study and for awards. The UQ PPL also states that assessment should be “fair and equitable” such that “no individuals or groups of students are unfairly advantaged or disadvantaged”. Despite this statement, UQ students who study in programs with inconsistent average grade outcomes stand a significant chance of losing their GPA-dependent scholarship.

These examples show some of the academic risks to the University when we consider variable consistency and comparability of grading outcomes from one school and faculty to the next. Grading inconsistency also poses wider risks for UQ in terms of undergraduate reward and retention, recognition of students who sit at the ends of the student achievement curves, and graduate employability in a competitive marketplace. When we consider grading at UQ we need to ask whether the grades given to students truly reflect their levels of achievement in their courses in a way that is transparent, credible, fair and equitable, sustainable, and workable. We also need to consider whether UQ is using grades in a manner that is consistent with the PPL. These considerations lead us to a central question: Should we respond to reduce grading inconsistency at UQ and if so, how can we do so appropriately?

GRADING AT UQ: POLICIES, PRACTICES, AND POTENTIAL PROBLEMS

THE PURPOSES OF ASSESSMENT

Assessment as a core element of academic study has multiple purposes – it fosters learning, promotes student engagement, examines and certifies learning, and provides quality assurance. These multiple purposes mean that assessment items and the marks given for them need to serve students, academics, the university, the community, and the government at the same time; which is a difficult position description. This situation is, however, applicable at UQ and helps explain the issues around assessment at the University.

The different purposes for assessment can influence how assessment is designed and marked. These assessment processes are also influenced by the two, competing core philosophies of assessors,
which can be described as positivist vs interpretivist (Elton and Johnston, 2002), scientific measurement vs judgement (Hager and Butler, 1996) and realist vs relativist (Yorke, 2010) (see Tables 1 and 2 in Appendix B). The analysis provided by Yorke (2010) indicates that science-based subjects more frequently use “realist” assessment while “relativist” assessment is more commonly associated with the arts, humanities, and social sciences. Realist assessment is characterised by the measurement of achievement against objective, detached, and explicitly defined criteria. In contrast, relativist assessment incorporates a more holistic view of marking the work by taking into account the students’ situation, the particular values of the marker, and the various ways to evidence excellence. Both philosophies influence assessment practice and policy at UQ and it is important to acknowledge that different aspects of both will be more workable and palatable for different disciplines. Fundamental and immutable epistemologies and ways of knowing that are inherent to different disciplinary practices influence workability and palatability (Dunn, 2016). As a consequence of this, asking people to change the assessment of skills and knowledge cuts to the core of their identities and belief systems.

The UQ PPL 3.10.02 (see below) describes assessment as a developmental activity in which both students and academics play a responsible role. This is a particularly important description, because it implies (i) that assessment should not necessarily be a single-point activity during a course and (ii) that students have the opportunity to develop as part of the assessment activity. This description leans towards the relativist, interpretivist, and judgement-dominated approach to assessment. Relativist assessment is also well suited to the process of assessing for work readiness (or work capability), innovation and entrepreneurial capacity, as well as the ability to create change – all of which are pillars of the UQ Student Strategy. These implications suggest that as UQ enacts the Student Strategy it may be necessary to consider carefully how our current assessment policies and procedures serve student learning and the development of our desired graduate attributes.

The UQ PPL 3.10.02 Assessment makes the following statements about assessment:

“Assessment is making judgements about how students’ work meets appropriate standards and drawing inferences from these judgements about students’ attainment of learning outcomes. At The University of Queensland assessment is used to achieve the following purposes:

- engage students in productive learning;
- inform teaching and learning decision-making;
- provide evidence of course- and program-level learning outcomes and graduate attributes;
- provide comprehensive, accurate, consistent and dependable certification of student achievement; and
- maintain professional and disciplinary standards.”

The UQ PPL also indicates that assessment

- is a developmental learning activity
- involves mutual responsibility
- is criterion-referenced
- is transparent
- is credible
- is fair and equitable
- is a purposeful, professional activity
- is sustainable and workable
THE COMPARABILITY AND CONSISTENCY OF GRADING AT UQ

This paper discusses grading comparability and consistency, with particular focus on the issue of grading variation at UQ. Figure 1 illustrates the extreme examples, both high and low, on the UQ course-level grading continuum. From these data, we can see that courses with a SLAT course code have a median 80% of students who gain a grade of 6 or 7. In contrast, a median 20% of the students in IBUS courses gain a 6 or 7 grade.

This grading variation is consistent with the PPL, because the PPL does not indicate that grade moderation is necessary across courses or across disciplines. The PPL does, however make mention of “quality assurance of assessment at program- and course-level, and within schools and faculties” while also suggesting that academics should engage in “collaboration with colleagues to provide the whole-of-program approaches that make assessment a coherent experience for students; and ongoing revision and enhancement of assessment practices.” These policies indicate the need for regular effort to examine the rigour and practice of grade allocation at UQ (hence this issues paper), but they stop short of recommending the necessity of moderation of grade averages or spreads so that different disciplines conform to a standard mean, median, or spread of grades.

![Box plots showing grade distribution](image)

Figure 1: Grade distributions for study areas with the top four (left) and bottom four (right) proportions of 6 and 7 grades in 2012-15. The figure represents the outliers from 127 study areas (i.e., 127 four-letter course codes), each of which has more than four individual courses with average enrolment >10 students. The vertical box-whisker plots show the proportions of students in courses with this code with a grade of 6 or 7 (green), another passing grade (orange) and a failing grade (blue) are shown (for example, a far higher proportion of students fail FINM courses than SLAT courses). For each box-whisker plot the box represents the middle two quartiles in each distribution, the line within the box is the median, and the whiskers represent the full extent of the distribution. The bars emanating from the Y axis show a normalised distribution of courses by grade distribution. For example, around half of the SLAT courses have between 80 and 90% of students achieving a grade of 6 or 7 and 100% of SLAT courses have fail rates of less than 10%. In contrast, the vast majority of IBUS courses have 65% to 75% of the students achieving “another passing grade” (not a grade 6 or 7) while around 15% of IBUS courses have a fail rate of more than 10%.

Figure 1 shows the most extreme examples of grading variation at UQ, but this figure does not establish an overall picture of UQ grading outcomes. To better understand grading at UQ (and possibly to find pockets of grading concern) we examined the effects of course level and course size (Figures 2 & 3 respectively) on grading outcomes.
The data in Figure 2 show that grading distribution changes by program year. In undergraduate programs (years 1-3 as defined by course codes) early-year students have poorer grade outcomes overall than students in later years. It is more complex to define trends for the grading outcomes in years 4, 5, and 6 because these course codes serve a mixture of undergraduate, honours, and masters courses.

**Figure 2: Grade distributions by year level (course code) in 2012-15.** The plots are drawn and interpreted as described in Figure 1. The relationship between course code and year level is explained here [http://www.uq.edu.au/startingatug/course-codes](http://www.uq.edu.au/startingatug/course-codes). Course codes beginning with 1, 2, and 3 are components of undergraduate programs. Course codes that begin with 4 can be (i) components of undergraduate programs which are more than four years in length or (ii) specialist honours-level coursework offerings. Honours research projects begin with 6, as do introductory Masters' program courses. Other Masters' program courses begin with 7. Course codes beginning with 8 are PhD-level offerings.

The data in Figure 3 show that:

(i) course grade distributions are more uniform for courses with over 250 enrolled students;

(ii) there is a noticeable trend towards a tighter grade distribution and overall lower course scores as the size of a class increases; and

(iii) very small classes (n<10) produce the highest overall proportion of students who attain a 6 or 7 grade. These trends are expected, as large classes are most frequently taught in first year to mixed-interest and mixed-ability cohorts. In addition, academics may find it more difficult to give low marks to students who they know personally in a small class and students may benefit from the additional guidance associated with a low student: teacher ratio.

Previous studies found that diverse factors contribute to students’ grades including gender, number of hours worked, SAT scores, number of missed classes, recommending the course to a friend, instructors, being a junior, course, and interest in the course (Kara et al, 2009). We do not currently have the data to consider the impact of all the factors known to contribute to grading outcomes at UQ. We can say, however, that the diversity of grade distributions is consistent over the last four years (Figure 4) with the range of distributions increasing in the four years prior (2008-2011) and some significant changes in distributions between course offerings.

Diversity in grade distributions is expected and desirable, since the students who enter the University will achieve varying levels of excellence. The extent of the diversity of grade distributions does indicate, however, that the standards for UQ grades are not consistent across UQ. This poses a
risk, particularly for undergraduate students, as an assumption of comparable grade standards across UQ underlies UQ GPA requirements for entry to programs, university awards, and scholarships. Inconsistency in grading may also unfairly affect students in their progression at UQ and employment related opportunities (e.g. selection for graduate recruitment and internships).

Figure 3: Grade distributions by class size in 2012-15. The plots are drawn and interpreted as described in Figure 1. Numbers of enrolled students for the courses in each category are at the bottom of each plot. The designation $n<10$ represents grade distributions only for courses with fewer than 10 students enrolled. The designation $n<5000$ represents grade distributions for all courses at UQ.

Figure 4: Grade distributions at UQ over time (2012-2015). The plots are drawn and interpreted as described for Figure 1.
THE IMPACT ON STUDENTS OF DISCIPLINE GRADING INCONSISTENCIES AT UQ

There are several important consequences for students of current inconsistent and incomparable grading practices at UQ outlined below.

**Undergraduate Scholarships**

Students on scholarships usually need to maintain a minimum GPA to continue their candidacy. A student enrolled in a program with a tradition of low grades may not be able to maintain a required GPA, even if they are “high-achieving” student within the discipline, and consequently they may lose their scholarship.

**Honours and Postgraduate Scholarships**

Honours and postgraduate scholarships are awarded on a competitive basis. GPA is a major contributor to the criteria used for decision-making around these scholarship awards. “Low-grading” disciplinary norms mean that students will be less competitive in comparison to others on a university-wide or national level. “High-grading” disciplinary norms mean that the best students cannot be distinguished from other students when awards are given.

**University Graduate of the Year Awards**

Engineering and science students win the majority of UQ University Graduate of the Year Awards, which probably does not reflect the distribution of quality students across UQ. Since 1999 Science, Engineering, IT, Biotechnology, Applied Science, and MBBS Graduates have won 27 awards. In the same period, students in the humanities won one award, and students in commerce won one award; in both cases, these students were completing dual degrees with Science. No students from any other UQ disciplines have ever won a Graduate of the Year Award. The list of awardees can be found at: [http://www.alumnifriendsuq.com/awards-scholarships/graduate-of-the-year-2/](http://www.alumnifriendsuq.com/awards-scholarships/graduate-of-the-year-2/)

**Competition for Employment Beyond Graduation**

Students from UQ are competing with graduates from other universities in the job market and in some situations, with fellow UQ students from other disciplines. Low GPAs due to low grading disciplinary norms do not serve our students well. Excessively high grading disciplinary norms that do not allow quality students to shine are also not helpful. If employers are unable to use an academic transcript to gauge accurately the skill and achievement level of an applicant, the reputation of UQ and all of its graduates can suffer.

**Difficulties for Cross-Disciplinary Study**

Differences in implicit disciplinary norms for grades are a particular challenge for students studying across disciplines. Differences in expectations and practices between disciplines may result in student confusion and dissatisfaction with the course and University. Students frequently study elective or service courses that have different norms from their core disciplines.

**Inequity for Students Who Apply for Program Transfer**

Students who do not have the entry requirements for their desired programs often study another program for a period to improve their entry scores. UQ specifies a minimum UQ GPA requirement [http://www.uq.edu.au/study/docs/domestic/uq-cutoffs-commonwealth-supported-places-](http://www.uq.edu.au/study/docs/domestic/uq-cutoffs-commonwealth-supported-places-).
for program transfers. Students who are studying some disciplines may be disadvantaged through this process, particularly as these entry cut-offs do not consider disciplinary relevance or any other criteria such as percentile rank in a cohort.

Course Shopping Based on Perceived Easier Grading
When grading is not consistent and comparable across the University, students may choose courses or programs with a tradition of higher grading. Variable norms throughout the University create competition between courses for enrolments. In some cases students may choose courses that do not fit the best with their interests or their course requirements due to the way in which a course is graded. It is common to see students seeking known “easy” elective courses to boost their GPA (e.g., https://www.reddit.com/r/UQreddit/comments/iyemx/looking_for_uqs_easiest_minimal_effort_subject/)

COMPLIANCE WITH NATIONAL-LEVEL GUIDELINES AND STANDARDS

The threshold standards require universities to “ensure the integrity of student assessment” (Section 4.3, Chapter 3) and “provide current, accurate, adequate, and openly accessible information for prospective and enrolled students on [...] content and assessment for each unit in the course of study” (Section 6.3, Chapter 3). In addition, the Standards for each higher education award include “Assessment is effective and expected student learning outcomes are achieved” (Standard 5, Chapter 3). Operationalisation of this Standard is described in detail in the HES document and a particularly important point is the requirement that course management includes “moderation procedures” to “ensure consistent and appropriate assessment” (Standard 5.3, Chapter 4).

The way moderation should be achieved is unclear in the HES document, but the HES Framework’s guidance notes for implementation make specific repeated mention of “external referencing” (TEQSA, 2015c). There is now also a TEQSA requirement that universities conduct “periodic overall reviews of courses of study” that are “informed and supported by more frequent monitoring of course performance at unit level” (p. 3). Provider’s review activities are now “expected to encompass external referencing against comparable courses (including student performance data) and to be informed by student feedback” (p. 3). This requirement echoes the currently used program review process at UQ but crucially, TEQSA defines “external referencing” as comparison of “internal courses and activities with others within or beyond the institution” (p. 1). This means that a comparative grading outcomes process across courses and programs both within and outside of any given institution may become part of the mandated standards quality assurance process at universities.

It is also important to note that the “Qualification Standards” in the HES Framework discuss program articulation and recognition of prior learning for students and includes the following statement:

“In determining credit towards awards, the higher education provider ensures it takes into account the comparability and equivalence of the learning outcomes, volume of learning, program of study including content, and learning and assessment approaches” (Section 3.5, Chapter 4).
This statement has the capacity to drive significant examination of assessment practice and grade comparability both within and between universities.

MAJOR CONCEPTS AND PRACTICES IN ASSESSMENT METHODOLOGIES

To address the complexity of assessment and to provide students, academics, the institution, and the community with a reliable “outcome” from education, Universities use two primary mechanisms of grade allocation: (i) norm-referenced assessment and (ii) criteria- and standards-referenced assessment.

In this section, we discuss these two assessment methods and examine the impact of disciplinary and individual expectations on marking. We also discuss models and mechanisms for moderating marks; some of which are inherent to the method of allocating grades. Frequently, disciplinary practice affects mark moderation behaviour – consequently the discussions of assessment methodologies, disciplinary practice, and moderation practice are somewhat interwoven.

NORM-REFERENCED ASSESSMENT

Norm-referenced assessment measures a student’s performance in comparison to the performance of same-cohort students on the same assessment items. The assumption is that all assessment items are marked in a consistent manner; that some students will perform better than others; and that ranking students can occur. Normative scoring uses a bell-curve as its foundation, meaning only half of those tested can score above the 50th percentile (Figure 5). This means that rank and position in the cohort underpin the awarding of student marks rather than on the raw scores that the students have gained (Mc Daniel, 1994). Consequently, norm-referenced assessment always means that students compete with each other for their position in the group, rather than with themselves for achievement against a set standard.

![Figure 5: The Bell or Normal Curve (adapted from www.AssessmentPsychology.com).](image)

The distribution of grades is specified in advance in this type of assessment and such distribution can be controlled at a course, program, faculty, or university level. Norm-referenced assessment allows complete comparability and control of the numerical grades given to students; which means it can create a standardised grading system across an entire institution. It does not make allowance, however, for the differences between cohorts, course difficulty, teaching quality, or assessment activities with the consequence that students of vastly different capacities can achieve the same grade if their cohorts also differ significantly. In addition, a student who has ably achieved all of the required graduate attributes of a course can may achieve a poor grade if they are even slightly less “competent” than other members of their cohort. In this situation, the definition of “competence”
and the metrics and methods used to quantify and calculate it become the subject of intense scrutiny (Crocker and Algina, 1986, p. 432).

UQ has chosen not to use this method of grade allocation. This choice is in line with the current practice in many other tertiary institutions both nationally and internationally as they recognise multiple disadvantages and limitations of norm-based assessment. Yet, even when academics use a criterion-referenced system, norm-referencing often tacitly influences their marking when they sense it is should be possible to rank students’ assessment items (Yorke, 2009). This sense can be magnified by the fear of institutional repercussions should too many students do too well (or too poorly) in a course.

CRITERION-REFERENCED AND STANDARDS-BASED ASSESSMENT

Criterion-referenced and standards-based assessment examines a student’s performance based on mastery of a specific set of skills. This type of assessment measures what the student does and does not know (or what they can and cannot do) at the time of assessment. The student’s performance is not compared to the performance of their cohort on the same assessment.

The criteria are the main categories of performance assessment (e.g., for a writing task the criteria might include “quality of argument”, “grammatical correctness”, and “use of references”). The standards are defined as “a definite level of excellence or attainment or a definite degree of any quality viewed as a prescribed object of endeavour or as the recognised measure of what is adequate for some purpose, so established by authority, custom, or consensus” (Sadler, 1987).

Unlike norm-referenced assessment, criterion- and standards-based assessment allows the student to strive for attainment of mastery goals that are consistent with their own level of skill and input. There are, however, some limitations to the relationship between skill, input, and outcome - some of these are related to the criteria and standards themselves. Sadler (2005) describes the four fundamental challenges facing university educators who wish to grade students according to true standards as:

- coming to grips with the concept of a standard
- working out how to set standards
- devising ways to communicate standards to students and colleagues
- and becoming proficient in the use of standards (p. 191)

These challenges are consistent and difficult to address and Sadler (2014, p. 273) goes so far as to describe “the futility of attempting to codify academic achievement standards”. In practice, both criteria and standards are open to interpretation by both markers and students – this means that it is practically impossible to fully define the levels of achievement required to gain particular marks (note Table 2 Appendix A). Academics attempt to overcome this limitation through a range of practices such as the provision of increasingly detailed rubrics, discussion of criteria with students, provision of exemplars, asking students to mark model items, and co-construction of the assessment criteria with students. These practices are variably effective. Moderation of marking and grading helps to facilitate consistency, however there is still a level of subjectivity associated with both working to and marking to criteria that invites students to argue about their grading. In some cases, the contention forces academics to return to assessing content recall rather than criterion mastery because content recall can easily be marked as right or wrong.
PROFESSIONAL JUDGEMENT, TACIT KNOWLEDGE, AND DISCIPLINARY DIFFERENCES IN MARKING

Despite the introduction of criteria and standards at universities, it is fair to say that assessment is “largely dependent upon professional judgement” (Price et al., 2008). As already discussed (Appendix A) there are different philosophies of assessment and these will affect the types and amounts of assessment that is set for students. There are also disciplinary differences in the way people mark. Academics generally learn to mark and employ “appropriate” disciplinary standards through an informal process of marking alongside disciplinary colleagues (Shay, 2005). This training in disciplinary expectations and tacit knowledge often begins when the academic is working as a teaching assistant graduate student.

Shay (2005) suggests that it is not an “error” when people mark differently, but rather is simply an outcome of the multiple different perspectives that people bring with them to marking. These perspectives include “lecturer’s values, specialist knowledge, socialisation processes, relationships with students, and their previous experience” (Fry et al., 2015, p. 118, citing Bloxham, 2012). It is fair to say that each marker has a unique lens through which they judge student performance and they construct their own Standards Framework (Ashworth et al., 2010) to which they mark. Studies show that (i) this framework is frequently holistic and (ii) that markers often allocate a mark to a piece without using the published criteria and standards. In such cases, markers use the published criteria and standards after marking to check their judgement, to justify their decision, and to frame feedback so that it is most useful to students (Bloxham et al., 2011). Although the use of rubrics to frame feedback is important, the different (and personally valid) perspectives of different markers do still lead us to two questions. First - “Is the marking of assessment items fair?” and second, “Assuming the marking is not completely fair, how much unfairness are we and the students willing to tolerate?”

As previously mentioned, academics hold discipline-specific views of their students and their expected individual achievements around assessment and this will contribute to differences in marking practice and grading outcomes between disciplines. Consider the variance in the level of achievement needed to work in health (where practitioners need to get 100% of a prescribed treatment for a patient correct), and in science (where a 50% recall of content may be sufficient to work in a group environment where a new scientist is mentored by more senior staff and relevant content is freely available on the internet). In some disciplines it is also easy to define a desired or required standard (e.g., the ability to fully and correctly complete a set of diagnostic steps) while in others it may be quite difficult to define exactly how “good” an item needs to be to gain top marks. The marking of assessment will reflect these differences and make allowances for them.

Take the example of a physiotherapy student: to be passed as competent this student would likely need to attain marks in the 6 or 7 range to pass their appropriate mastery tests. There is little room for error or creativity in their practice and this student should complete all components to a high standard in order to satisfy accreditation standards. In contrast, students in other disciplines may pass with a wider range of grades because they are marked to a different set of standards for a different range of discipline-appropriate skills and attributes. Although industry and professional accreditation will drive marking practice, one could argue that very high grading norms are not just the result of accreditation requirements; it is possible that other practices within industry-accredited teaching programs also contribute to high grades.
Changing the disciplinary perspective of what is appropriate and acceptable for student achievements is very difficult. Such a change asks people to disengage from their own discipline and their own identity and establish a new (and foreign) set of standards that may conflict with their established values. The new set of standards may also conflict with the requirements for legal accreditation, professional credibility, and employment capability in an academic’s discipline. Knotts et al. (2009) sum up the tension between assessment standardisation and professional competence with this reflection around assessment in their own discipline:

As greater pressure is brought onto teachers to standardize and use so-called “objective” assessment connected to Student Learning Outcomes (SLOs), the theater history classroom begins to lean toward reflexive history and away from the kind of doing actors practice in their training and performance. [...] Although crossing the divide, within and across disciplines, is desirable, SLOs become barriers when the skill sets students are expected to gain are knowledge-based and discipline specific. (p. 194)

MODERATION, CALIBRATION, AND PEER REVIEW OF ASSESSMENT PRACTICE

Individual academics have different ideas of the required standards expected in an assessment task and thus vary in their opinions as to whether individual students have met their view of the criteria. Moderation does help us get a collective understanding of standards; the process is “intended to ensure that the mark a particular student is awarded is independent of which marker does the marking” (Sadler, 2013, p. 5). There are several avenues for grade moderation - a common component in these processes is the open communication, discussion, and critique of assessment criteria and standards.

Bloxham and Boyd (2007) offer us a visual depiction of moderation (see Appendix C). In this process, there are several communication events. These include an initial scrutiny of assessment items by colleagues; communication of the criteria and standards to students (presumably with associated discussion); an inter-module moderation process after the initial marking; the engagement of an external examiner to look at both the assessment items and the marking; and the further consideration of marks by an examination or assessment board. All of these steps are important with many of them conducted at UQ; however the feasibility and sustainability of this complete process is debatable, particularly in the modern era of tight turnaround requirements for marking and feedback during a semester that is threatened with trimesterisation. Traditionally academics rely on the re-use of tried and tested assessment items to obviate the need for some of these steps. However, the UQ requirement that assessment be 80% different for each iteration of a course\(^1\) means it is very difficult to standardise and compare the difficulty of assessment, and the levels of student achievement from one year to the next. Of course, similar assessment from one year to the next also creates benchmarking issues because assessment recycling makes it easy for students to obtain answers from students who have taken the course previously.

It is, however, common at UQ for groups of markers to meet before they mark and engage in “social” or “consensus” moderation (Sadler, 2013; Linn, 1993). In this process the marker group reach a shared understanding of the standards, thus calibrating their judgement in preparation for marking. Consensus moderation is effective for marking complex assessment tasks in which there is no one right or wrong answer. “Calibrated” academics also provide a repository for information

\(^1\) UQ PPL 3.10.02b Assessment, part 4.3.2: http://ppl.app.uq.edu.au/content/assessment-procedures
about disciplinary standards over time and institutions (Sadler, 2013 p. 17-18) – often these people are course coordinators who work with a group of tutors to establish an appropriate marking scheme for an assessment item.

Calibration is a particularly salient feature of the definition of standards offered by Sadler, in that it emphasises the centrality of peer and professional judgement in the marking process. Sadler goes on further to state, “the goal is for academics to be confident in their own informed and calibrated judgements, and able to trust their colleagues’ abilities to make routine appraisals of student works with an appropriate degree of detachment and self-regulation” (p. 14).

O’Connell et al. (2016) propose a further process of academic calibration in which assessors from multiple universities “come together to debate and apply standards to student work” in a large workshop format so that they “can achieve consensus in applying threshold learning standards to students’ work” (p. 334). Certainly, peer engagement in moderation through consultation and some form of calibration of peer judgements around exemplar assessments is “generally agreed to be desirable” (Freeman and Ewan, 2014, p. 5). O’Connell et al. (2016) found that marker grading variability was reduced (but not eliminated) after the group workshop, which lends further weight to the idea that facilitated group reflection and comparison of marking can be effective as a moderation tool. One can envisage that this process might help close the gap between disciplinary norms, but only if academics from multiple disciplines were all included in the same calibration workshop. In contrast, discipline-specific workshops and calibration exercises are likely to reinforce the extant disciplinary grading practices at UQ.

**GRADING PRACTICES AT UQ**

**THE USE AND APPLICATION OF UQ GRADE DESCRIPTORS**

The UQ PPL 3.10.07 provides a set of standards for grades to apply to all courses at all levels of study (See Appendix D Table 3). The challenge of implementing these standards in disciplinary contexts and across the different levels of education at UQ is managed locally at the course, program, or faculty level. UQ does not have a default set of percentage cut-offs for different grades and there is little consistency across UQ in how cut offs are decided upon or applied. There is also little common practice around moderating grades from course to course, from discipline to discipline, and from year to year. The requirement consistently to change assessment items contributes to this issue, as academics are now less likely to have a “bell weather” assessment item to give a sense of how an extant cohort compares to previous cohorts.

Schools and faculties frequently have their own rules and practices for grade allocation and in many cases, these are based on “feel”, tradition, or a sense that a student “deserves” a particular grade. The lack of rigour is particularly apparent for students who attain a mark that falls close to an established grade cut-off. At examiners’ meetings, it is not unusual for the most argumentative person in the room to get their way about application of grade boundaries; in some cases, this can affect the outcomes for all of the marginal student grades on a particular day.

Although most academics and educational administrators try to do the best, most conscientious, and most student-friendly job they can around grade allocations there is a consistent question and unease about the “right” proportion of a class in each grade bracket. Heated arguments can ensue in
examiners’ meetings about very low (or very high) numbers of students in the most extreme grade categories. There are also questions about the level of achievement required to attain each grade hurdle in assessment.

While we do not have a simple way to quantify the proportions of the different grading approaches used at UQ, we suggest percentages (with or without item-specific hurdles) are used in the majority of courses. However, Yorke’s extended exploration of grading practices (Bridges et al, 1999; Yorke, 2011) identified diversity of distribution between disciplines using a percentage scale and issues in the underlying standards of percentile grades. Yorke (2000, p. 2) also suggests that grade heterogeneity “may be a greater problem where percentage scales, rather than other approaches to grading, are used”. We do not believe there is enough evidence to recommend an institutional move away from percentage-based means for allocating grades, but UQ may wish to promote other options for courses in which students may benefit from a different type of grading outcome.

THE GROUP OF EIGHT QUALITY VERIFICATION SYSTEM (QVS)

A part of the Group of Eight (Go8), UQ runs the Quality Verification System (QVS) each year. The QVS is designed to:

• demonstrate the appropriateness of the standards of learning outcomes and grades awarded in Go8;
• universities;
• maintain and improve the academic standards of Go8 universities;
• enable comparisons of learning outcomes in similar subjects across Go8 universities; and
• promote discussion on good practice in teaching and learning in the Go8 universities

Although participation in the QVS is an appropriate practice, the effect of participation on UQ assessment outcomes is debatable. It is possible that the individual academics nominated as assessors gain knowledge from seeing work from other universities. The course coordinators who read the feedback on their assessment items are also likely to gain insight into the opinions of academics from other universities. There is little evidence that this knowledge propagates further into UQ and impacts our practice.

AN ENVIRONMENTAL SCAN OF PRACTICES USED TO ACHIEVE GRADING EQUALITY AT OTHER UNIVERSITIES ACROSS AUSTRALIA

An environmental scan by the authors of practices used to achieve grading equality at other universities yielded few results. The current focus of assessment centres on discipline driven inter-institutional peer evaluation rather than inter-institutional practices. While as explained previously in this paper, scholarly literature provides conceptual discussion of pertinent topics such as different grading systems, measurement versus judgement of assessment, calibration and moderation, exemplars of best practice are harder to find and rather sporadic in their content topics as demonstrated in the examples below.

BATCHelor INSTITUTE OF INDIGENous TERTIARY EDUCATION (AUQA)

This example provides insights into a multi-pronged institutional moderation process that begins before the distribution of assessment tasks to students, and then then continues post-moderation
and finalisation of grades, to a final review of processes by a committee. For further information see: http://www.auqa.edu.au/gp/search/detail_print.php?gp_id=2967

THE UNIVERSITY OF SOUTH AUSTRALIA ASSESSMENT MODERATION TOOLKIT

This website (https://lo.unisa.edu.au/course/view.php?id=8539) provides resources to institutions with particular focus with transnational education to promote good assessment moderation practices. The foundation research comes from the ALTC project on moderation for fair assessment in transnational learning and teaching.

THE HIGHER EDUCATION ACHIEVEMENT REPORT (HEAR)

This report allows institutions in the UK to offer a more comprehensive e-record of individual student achievement as a companion to their graduation certificate. It includes achievement in areas of academic study, co-curricular or volunteer activities, awards and placement. The University of Sheffield’s website (https://www.sheffield.ac.uk/hear) provides information for staff and students about how to best use this tool. It also details the different sections that make up the report.

SUGGESTIONS AND OPTIONS FOR ADDRESSING GRADING INCONSISTENCY AT UQ

In presenting this paper, the authors take the stance that grading inconsistency is unavoidable unless the performance of a student’s peers influences their grade.

We also note that grading inconsistency does not suggest:

(i) that students are not meeting the required graduate attributes in UQ courses;
(ii) that teaching at UQ is poor; or
(iii) that the type or amount of assessment given to the students is inequitable, excessive, insufficient, or inappropriate.

The following suggestions may address grading inconsistency at UQ. In most cases these suggestions outline practices that can help make grading practice and outcomes more publicly visible to academics, to students, and to the community.

The authors of this paper are not necessarily recommending that any action is appropriate or any one practice is superior to another. In addition, the authors caution that UQ does not have the data in hand to establish a watertight case for wholesale reform of assessment practice. Some of the suggested practices would require UQ to amend its PPL on assessment.

ADOPT A NORMATIVE APPROACH TO GRADING WITH A BELL-CURVE OF SET PROPORTIONS

As discussed earlier, normative grading allows the University to define exactly which proportions of each cohort obtain different grades. This blanket approach is unlikely to be an acceptable way to achieve grading consistency at UQ, as:

(i) the institution is committed to criterion-referenced assessment;
(ii) norm-referenced assessment has many limitations; and
(iii) norm-referenced assessment contravenes the PPL.

SET QUOTAS FOR DIFFERENT GRADE OUTCOMES IN A “QUASI-NORMATIVE” APPROACH
The University could set minimal quotas for grade outcomes from courses (e.g., a course-grading outcome must have at least 10% of the students with a grade 7 and another 20% of students with a grade 6). Setting maximal quotas would be unacceptable with academics and students alike, however a case for a minimal quota that could be exceeded at the discretion of the course coordinator may be feasible. Setting a minimal quota may provide some clarity about the acceptable minimum threshold for student outcomes – this is currently unclear for many academics at UQ and central guidance may be enough to significantly improve outcomes and expand the grading spread. We provide an example of potential central guidance wording below.

“It would typically be expected that the proportion of grades of 7 would be around 10-15%, grades of 6 around 15-20%, etc.”. In any semester, grade cutoffs advertised in the ECP can be lowered, but cannot be raised. It is recognised that there are good reasons why grading outcomes may fall outside these ranges. In such cases, course coordinators should write a brief justification, to be considered by the School TLC or chief examiner. In some cases, assessment expectations may need to be reconsidered prior to subsequent offerings of the course.”

ASK DISCIPLINES FOR JUSTIFICATION OF THEIR GRADE SPREADS

UQ could ask disciplines to provide a justification of their grade spreads and their median grades. There is a possibility that academics will see this as yet another administrative burden and an attack on disciplinary autonomy, however disciplines should be able to justify the risks to their students that are associated with traditionally low marks. If disciplines can demonstrate that low marks are part of the disciplinary landscape both within Australia and internationally, and that students are not suffering as a result, then the University could consider the risk acceptable and sustainable.

PUBLICISE GRADE SPREADS FOR COURSES AND PROGRAMS

Some UQ course coordinators are unaware of the grade spread that typifies their disciplinary courses and it is likely that most academics do not know about the grading profiles seen in other disciplines. It is fair to say that the opportunities to compare one’s own course and discipline grading outcomes to those in other courses are limited. Grade spreads for courses and disciplines could be publicised on a website (placed behind an UQlogin). Alternatively, academics could have the spread of grades in their course sent to them in conjunction with appropriate grade distribution data from their own discipline and from other disciplines within the University.

MAKE ALLOWANCE FOR STUDENT GRADE OUTCOMES BASED ON THEIR DISCIPLINARY NORMS (OR THE NORMS OF THE COURSES THEY HAVE TAKEN)

A major risk associated with grading inequity is the vulnerability of students who are required to maintain a particular GPA. Since University academic prizes refer to GPA, students from certain disciplines are privileged over others. The system at UQ could be changed, however, and students could be awarded prizes, scholarships, bursaries, transfers to programs, and positions in postgraduate programs, based on their percentile rank in the cohort, rather than their GPA. If this method was used students from disciplines with low average grades would no longer be disadvantaged and students would probably be less likely to “shop” for courses that traditionally give high grades.

PROVIDE COURSE GRADE OUTCOMES AND A FINAL TESTAMUR TO STUDENTS WITH THEIR PERCENTILE RANK INCLUDED ALONGSIDE THE GRADE
Truly excellent students are currently unable to evidence how their performance rates against other students in their cohort. Although we are not recommending a norm-based system of assessment, UQ could provide students with both their grade and their cohort percentile position (both of which are gained through criterion-referenced assessment). Students could also request that their final testamur contain percentiles (i) for all of their course outcomes, (ii) for select courses, or (iii) for none of their courses. This would allow high-quality students to evidence their outcomes more thoroughly while weaker students could display only their grades. Students could also use these percentile values to evidence their quality to potential employers and to potential research supervisors at undergraduate, honours, or PhD level. Inclusion of a percentile does not take away a high or a low grade from a student; however it does give a sense of the “value” of that grade.

The inclusion of percentile values would likely be less contentious if done using a “broad brush” with categories of percentile (e.g., top 5% of a class of 700 or top 40% in a class of 20). Use of fine-grained percentiles (e.g., top 4.7% in a cohort of 200) would likely lead to intense competition between students for positions in the course. Fine-grained use of percentiles would also necessitate consistent alteration of multiple students’ percentiles if an individual student obtained a re-mark or resolved an incomplete grade after the conclusion of semester.

This practice would likely prompt a revision of the UQ PPL to formalise the idea that assessment gives a loose ranking to students in a class. Although this is likely to be a contentious idea, it would bring the UQ policy into line with the current practice, in which GPA essentially ranks students when considered for awards and other preferment at the University.

**PROVIDE OPPORTUNITIES FOR PEER REVIEW AND STANDARDS COMPARISON WITHIN AND BEYOND DISCIPLINES AT UQ**

UQ could develop a system to facilitate peer review of assessment standards both within and external to disciplinary silos. Academics could meet, for example, to explain and justify what makes a student’s submission worth a particular grade. This information can be discussed across disciplines and staff could use a range of taxonomies as anchors for discussion; these taxonomies include the UQ grade standards, the AQF descriptors (Australian Government Department of Education and Training, 2016), and the SOLO taxonomy (Biggs and Collis, 1982).

Moderation of course outcomes could also be modelled on the processes used in High Schools, where samples of student work are compared across courses. While there are huge disciplinary differences, this process provides opportunities to reflect on rationales for grades and compare assessment practices across disciplines.

**PROVIDE ALTERNATIVE AND ADDITIONAL NON-GRADED WAYS FOR STUDENTS TO EVIDENCE THE QUALITY OF THEIR ACHIEVEMENTS AT UQ**

Other universities in Australia provide students with additional ways to evidence their achievements. These methods include badging and grouping of particular graduate attributes to officially “flavour” a student’s qualification; the provision of individualised videos to students that they can use to advertise their personal brand; e-Portfolios; and additional credentials that sit alongside the official testamur to give a sense of a student’s involvement in the life of the University. UQ is also using or investigating some of these options (the UQ Advantage Credential and the HABS e-Portfolio project are examples). Such “point of difference” offerings are rapidly becoming the norm in Australian higher education and it is important that UQ remains aware of the competition in this space.
DO NOTHING
The simplest option for addressing grading inconsistency is to do nothing and assume that grading inconsistency is a necessary and natural outcome of assessment at a large university that offers diverse educational programs.

CONCLUDING COMMENTS
As an institution UQ provides the opportunity for a large and very diverse population of students to choose from a rich selection of study programs and pathways – this diversity makes UQ an engine for innovation and an exciting place for students and staff alike. The variety of disciplines and programs available for study at UQ means that the University will have a concomitant variety of assessment and grading practices that reflect the different learning outcomes and graduate attributes that are appropriate for our students.

The differences in grading outcomes between courses are a natural result of the scope and variety of UQ’s offerings. Our duty of care to students, however, provides us with an imperative that we must offer our students a grading mechanism that allows them to properly evidence their skill sets and the quality of their work. UQ must also decide whether we are using our assessment in a manner that is consistent with the PPL and with TEQSA requirements. UQ’s current practice of using assessment outcomes to rank students based on GPA is inconsistent with our stated developmental policy and largely relativist philosophy. Some students are unfairly advantaged or disadvantaged by the use of assessment (and GPA) to rank students, simply because of the disciplinary norms that drive grade allocation in their chosen UQ program.

As explained previously, the diversity in grading at UQ is not all bad – it reflects standards and practice that typify and even define disciplinary identities. The differences between disciplinary grade allocation profiles are, however, extreme in some cases. This suggests that UQ should look at the standards and practices used across the University and take active steps to reduce the grading inequalities experienced by students. We suggested a range of methods that could be used to address this issue; some of which are likely to be more palatable to staff and students than others. In addition to moderating grades, it is also possible that UQ could consider an additional method (or methods) for evidencing the value of student achievements such as percentile provisions, enhancement and expansion of the UQ Advantage recognition scheme, or personalised student artefacts such as videos and ePortfolios.

There is no simple solution to the issue of grade comparability and evidencing achievement at UQ however, left unattended, grading inequality is unlikely to resolve itself miraculously. Central leadership will be required for UQ to address this; it is likely that both funding and considered policy revision will be needed to guide and support staff as they (i) make grading fairer and more equitable for our students and (ii) provide opportunities for students to evidence their learning and achievements in diverse ways.
REFERENCES


Appendix A: Grading Data from UQ

NOT PUBLICLY RELEASED
Appendix B: Assessment Philosophies Compared

Table 1: Assessment principles (from Hager and Butler, 1996)

<table>
<thead>
<tr>
<th>Scientific measurement model</th>
<th>Judgemental model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assess tasks remote from the world outside of classrooms</td>
<td>Assess tasks that reflect outside context</td>
</tr>
<tr>
<td>Assess solutions to problems</td>
<td>Assess processes by which problems are solved, as well as the solution</td>
</tr>
<tr>
<td>Assess simplified, discrete tasks</td>
<td>Assess performance of holistic tasks in their actual context</td>
</tr>
<tr>
<td>Assess individuals only</td>
<td>Assess group work as well as individual work</td>
</tr>
<tr>
<td>Emphasise one right solution</td>
<td>Emphasise alternative ways to reach acceptable solutions</td>
</tr>
<tr>
<td>Assess tasks directly from the curriculum as taught</td>
<td>Assess tasks that are relevant to the curriculum but expand on it</td>
</tr>
<tr>
<td>Assess discrete tasks one by one</td>
<td>Assess performance on holistic tasks as well</td>
</tr>
<tr>
<td>Teachers rigidly prescribe nature and form of assessment tasks</td>
<td>Learners help to design nature and form of their assessment tasks</td>
</tr>
</tbody>
</table>

Table 2. Assessment, as viewed from realist and relativist standpoints (from Yorke, 2011)

<table>
<thead>
<tr>
<th>Realist</th>
<th>Relativist</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standards are objectively defined</td>
<td>Standards are normative and consensual</td>
<td>The realist’s objectivity reflects underlying values</td>
</tr>
<tr>
<td>Performances can be measured against these standards</td>
<td>Performances are assessed with reference to these standards</td>
<td>The difference here is between measurement and judgement</td>
</tr>
<tr>
<td>The assessor is objective and detached</td>
<td>The assessor interprets the extent to which the performance relates to the standards</td>
<td>The underlying distinction is between context-free assessment (the realist’s position) and context-relevant assessment (the relativist’s position)</td>
</tr>
<tr>
<td>Values play no part in the assessment</td>
<td>Value positions are embedded in the norming of standards</td>
<td></td>
</tr>
<tr>
<td>The situation of the student is not taken into account</td>
<td>The context of the assessment is taken into account</td>
<td></td>
</tr>
<tr>
<td>Explicit criteria and rubrics are invoked</td>
<td>There are broad statements of expectations</td>
<td>The issue is the degree of detail that can be encompassed by criteria</td>
</tr>
<tr>
<td>Measurements are taken as true and reliable representations of achievement</td>
<td>Assessments are judgements of the extent to which achievements fit with expectations</td>
<td>The distinction is between measurement and judgement</td>
</tr>
<tr>
<td>Tasks are set by assessors</td>
<td>Tasks may be selected by students to suit their strengths and interests</td>
<td>The selection of project topics is a feature of both the sciences and the social sciences</td>
</tr>
</tbody>
</table>

Further resources around assessment can be found here: http://www.adm.heacademy.ac.uk/news/sector-news/review-of-the-ltsn-generic-centre-guides-and-briefings-on-assessment/index-18507.html
Appendix C: Moderation Flowchart

Moderation flowchart

1. Lecturer(s) design assessment items and write assessment criteria (and marking scheme)
   Consider learning objectives, fairness, contextualisation and clarity

2. Scrutiny of assessment items by colleagues/course team/external examiner

3. Communication of assessment criteria and marking scheme to markers and students

4. Students complete assessment items

5. Lecturers mark completed assessment

6. Second marking
   Representative sample √ n + fails (min. 5)
   OR
   Double marking (e.g. dissertations and high value assignments)

7. Markers agree on final marks (with third marker involved if dispute)

Intra-module moderation
Checking for consistency of marking and standards across modules which are offered more than once during the year or at different campuses

9. External examiner receives an agreed sample of items for all modules (not usually applicable to Year 1 works)

Examination/assessment board
Considers and agrees all marks. Reviews mark profiles and consider unexpected or inconsistent patterns in student performance across modules and student groups.

11. Marks/grades released to students

Review of assessment with reference to examination board data and examiner, staff and student feedback to inform future assessment processes

### Appendix D: UQ Grade Descriptions from PPL 3.10.07

<table>
<thead>
<tr>
<th>Final Grade</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fail. Fails to demonstrate most or all of the basic requirements of the course.</td>
</tr>
<tr>
<td>2</td>
<td>Fail. Demonstrates clear deficiencies in understanding and applying fundamental concepts; communicates information or ideas in ways that are frequently incomplete or confusing and give little attention to the conventions of the discipline.</td>
</tr>
<tr>
<td>3</td>
<td>Fail. Demonstrates superficial or partial or faulty understanding of the fundamental concepts of the field of study and limited ability to apply these concepts; presents undeveloped or inappropriate or unsupported arguments; communicates information or ideas with lack of clarity and inconsistent adherence to the conventions of the discipline.</td>
</tr>
<tr>
<td>4</td>
<td>Pass. Demonstrates adequate understanding and application of the fundamental concepts of the field of study; develops routine arguments or decisions and provides acceptable justification; communicates information and ideas adequately in terms of the conventions of the discipline.</td>
</tr>
<tr>
<td>5</td>
<td>Credit. Demonstrates substantial understanding of fundamental concepts of the field of study and ability to apply these concepts in a variety of contexts; develops or adapts convincing arguments and provides coherent justification; communicates information and ideas clearly and fluently in terms of the conventions of the discipline.</td>
</tr>
<tr>
<td>6</td>
<td>Distinction. As for 5, with frequent evidence of originality in defining and analysing issues or problems and in creating solutions; uses a level, style and means of communication appropriate to the discipline and the audience.</td>
</tr>
<tr>
<td>7</td>
<td>High Distinction. As for 6, with consistent evidence of substantial originality and insight in identifying, generating and communicating competing arguments, perspectives or problem solving approaches; critically evaluates problems, their solutions and implications.</td>
</tr>
</tbody>
</table>