

Towards a Learning Standards Framework

Learning and Teaching Standards (LaTS) Project: *Peer review and moderation in the disciplines*¹

Funded by DIISRTE Office for Learning and Teaching

*Professors Liz Deane and Kerri-Lee Krause² (University of Western Sydney)
in collaboration with the project team*

Preamble

This discussion paper arises from issues raised during the course of this national, cross-institutional project on approaches to monitoring and assuring learning standards. It is the result of collaboration among project team members, the project steering group, project participants and associates of the project. It is intended to contribute to the national discussion of approaches to monitoring and verifying learning standards. The views expressed in this report do not necessarily reflect the views of the funding body, the Australian Government Office for Learning and Teaching.

1. Context

1.1 For the past few years there has been rising global interest in academic quality and the development of valid indicators. In research this is being played out in a range of ranking systems whose merits, though hotly contested, are generally accepted. In education the situation is more ambiguous. Although the aforementioned ranking systems have sought to include indicators such as student survey responses or staff student ratios, these have often been seen as problematic and the relationship to educational outcomes is questionable.

1.2. Into this gap has stepped the discourse of determining the quality of graduating student outcomes. To date, this has been measured by proxy indicators such as surveys of student satisfaction and employment outcomes. More recently this conversation has expanded to include possible exiting testing of graduates. Attention has also turned to the more complex and nuanced area of determining and evidencing the learning outcomes of final year undergraduate and postgraduate coursework students where there is typically no integrative assessment such as a dissertation.

1.3. Internationally, the direction of this discourse is evident in the OECD AHELO project which, in two program areas of engineering and economics, is investigating the feasibility, utility and validity of both standardised exit testing and assessment of agreed, complex learning outcomes, discipline knowledge and its application. The Tuning project, currently being deployed in Europe, South and North America, is a further example of a cross-national approach to establishing consensus regarding descriptors for intended learning outcomes in undergraduate degrees.

1.4 In Australia there has been little support for standardised exit testing of generic skills and of discipline skills among graduating cohorts. The majority of academic effort has focussed on development of agreed discipline based approaches to determining student learning outcomes. The initial step along this path was the establishment of the ALTC/OLT Discipline Scholars. They have been working within and across their disciplines to develop agreed threshold learning outcomes (TLOs) or learning standards for graduates in respective discipline areas. To a significant extent this consultative, collegiate approach mirrors that undertaken in the UK Subject Centres of the previous decade.

1.5 Allied to this, and in the case of the Accounting discipline, evolving from it, have been three related but slightly different national projects:

- the OLT/ALTC supported Learning and Teaching Standards Project (LaTS) involving inter-university peer review and moderation in the disciplines;
- the Group of Eight (Go8) Quality Verification of Standards (QVS) project; and
- the Achievement Matters: External Peer Review of Accounting Learning Standards project (AMA) supported by the Australian Business Deans Council, professional bodies and OLT and benchmarking outcomes explicitly against the learning standards agreed by the accounting discipline in 2010 for bachelors and coursework masters degrees.

1.6 The TEQSA regulatory framework places a premium on proportionate management of risk, informed by quality information. In the case of the learning and teaching standards, the projects mentioned in 1.5 arguably play a key role in yielding robust, discipline-based evidence of approaches to monitoring and assuring academic standards at the discipline level within and across higher education providers. Considerable discussion of teaching standards (i.e., process or delivery standards) has occurred (e.g., see the Teaching Standards Framework) and they are well represented in the current threshold Provider (Course Accreditation) Standards. However, a learning standards (i.e., outcome standards) framework is yet to be proposed.

1.7 This paper presents an initial comparative review of three learning and teaching standards projects (see 1.5) with the intent of progressing towards a generally agreed, reliable and valid learning standards framework that is sustainable. The projects are reviewed in terms of their methodology, utility, validity, respective strengths and limitations.

1.8 The development of a learning standards framework, grounded in disciplinary contexts, would have value in many domains – in providing assurance to current and prospective students as to the quality of their award; to employers as to the relevance and competence of graduates and to government regulators as to attainment of currently ambiguously defined threshold standards of program delivery and outcomes.

2. Three Learning and Teaching Standards Projects

2.1 Table One provides a summary of the key features of three complementary national projects designed to provide evidence that learning and teaching standards are being monitored and assured by individuals, institutions and the sector.

2.2 Methodology. The LaTS and QVS approach consists of external, discipline-led, academic peer review of final year undergraduate student outcomes. In each project, external reviewers:

- comment on the appropriateness and comparative quality of the specified learning outcomes, assessment tasks, assessment criteria and assessment processes set for samples of final year subjects; and
- report on the appropriateness of the grades awarded to stratified random samples of student work in these subjects.

These two projects differ in two key ways (see Table 1):- the LaTS method uses blind peer review of unit materials and requires peer reviewers to grade assessment tasks where the grade allocations have been removed. The QVS reviewers examine graded assessment items and indicate whether or not they verify the grade allocated.

The AMA approach also consists of external, discipline-led, academic peer review of final year student outcomes. However, AMA includes coursework postgraduate as well as undergraduate outcomes, recognising higher standards must be evident in the former. Second, it uses discipline standards as the benchmark, pre-agreed across the discipline community (ie. academics, practitioners, professional bodies). Recognising there is rarely an integrative assessment covering all threshold outcomes in coursework degrees, initially it has focussed on a subset of learning standards. Third, reviewers rate unmarked student work into one of two categories ('not meeting' or 'meeting' the threshold standard) rather than one of four grades (Fail, Pass, Credit, Distinction). Fourth, peer reviewers engage in a three-stage consensus moderation activity with the aim of achieving calibration prior to undertaking any external review. For example, thirty reviewers participated in the September 2012 pre-workshop review, workshop consensus, and post-calibration confirmation activities. This 'calibration' of academics across institutions (and with practitioners) relates to the specific learning standards in focus. Calibration revolves first around the validity of an assessment task to demonstrate the learning standards in focus; calibration on the standard achieved in a small random sample of pre-reviewed student work; and, calibration on new samples of student work. Fifth, to ensure there is little chance of data biasing reviewers, samples are randomly collected rather than stratified by grade bands. Sixth, private and TAFE providers of coursework degrees are engaged in the consensus process.

2.3 Valid and reliable assessment. All three projects are characterised by a discipline-based peer evaluation of assessment items, a clear recognition that the focus of the debate hinges on the nature, suitability and validity of assessment in a disciplinary context. All require external

reviewers to comment on the appropriateness and quality of desired learning outcomes for the subjects; assessment tasks; criteria and processes. All are focused on the subject level, albeit what would constitute endpoint/capstone subject. Both LaTS and QVS require peers to provide program level information to enable judgements to be made in the context of course learning outcomes. Where they differ is in the extent, timing and level to which assessment items are critiqued and intra- disciplinary conversations facilitated.

2.4 Disciplinary judgements. All three projects require assessors to review assessment items in the context of the desired learning outcomes as determined by the assessment providers. This approach allows for academic diversity to be preserved. Thus whilst the projects require comments as to the suitability of the intending learning outcomes for a final year subject, not all subjects are intended to be delivering the same curriculum. This contextual assessment allows for the flexibility needed in discipline areas where learning outcomes are not tightly proscribed by professional accreditation requirements.

The AMA project, in contrast, has *a priori* agreed on the desired learning outcomes being assessed but has not been prescriptive as to the nature of the assessment tasks – just that they are appropriate and valid.

2.5 Capacity Building. All projects provide some scope for building capacity and academic development of participants. This is implicit in the QVS and the benefits of being involved was acknowledged by reviewers, particularly the opportunity to see other institutions subject structures and differing modes of assessment. Post review dialogue between reviewers is more explicitly built into the LaTS and the AMA protocol, with the latter being much more extensive and focused and developing consensus around specified learning outcomes, albeit the differing pathways of delivery and assessment

2.6 Workload. Possibly the biggest potential impediment to systemic introduction of peer review processes are workload implications. QVS reviewers identified timing and time frame are problematic for teaching academics as were the costs associated with de-identification and tracking assessments, and coming to grips with institutional variation, understanding the context. In terms of reviewing workload, AMA reviewers incur an additional initial workload as they must participate in calibration activities.

2.7 Scaling up. There are two issues related to scaling-up. The first of these concerns an institutional agreement as to how much peer review is enough to achieve the desired outcome of assured standards. The other concerns scaling up from unit/subject to course level. Both the LaTS and QVS multi-disciplinary projects have used final year agreed indicator subjects as a proxy for course level assessment. The LaTS also asks reviewers to make judgements about unit level performance in the context of course learning outcomes provided by the home institutions.

However, one of the challenges identified in the LaTS project has been the need to engage academic staff in course-level thinking in order to make judgements about exit standards. In

contrast, by benchmarking against discipline standards, AMA is explicitly course level focussed.

2.8 Deliverables. All three projects have the potential to identify, critique and consolidate the indicators that can be used to set the standards for teaching standards (e.g., assessment items, marking criteria, and guidelines for moderation and calibration) and learning standards (i.e., learning outcomes).

3. Lessons learned and next steps

3.1 Arguably, academic staff in the disciplines are well placed to monitor and assure learning standards. Universities have well-established quality assurance and policy frameworks requiring regular moderation of assessment at the department level, benchmarking and cyclical review.

3.2 However, in general, academic staff in the disciplines have not been called upon to evidence their standards-setting practices and learning outcomes to the extent that is now required to provide public assurance that learning and teaching standards are being met and exceeded. Three lessons derived from the LaTS project may help to inform next steps.

- Lesson 1: shift from unit to course level thinking. In order to make judgements about exit standards, academic staff need to be able to demonstrate that they are interpreting unit-level learning standards in the context of course-level learning outcomes. This requires them to articulate the extent to which students' performance on unit-level assessment tasks meets course or discipline-level learning standards.
- Lesson 2: articulate reference points. In the LaTS project, interviewees had difficulty articulating how they arrived at academic judgements about the standard of work that warranted a Fail/Pass/Credit grade. One strategy for addressing this is to ensure that academic staff are able to articulate various disciplinary reference points and how these inform decisions about learning standards. Reference points might include:- the Australian Qualifications Framework; discipline standards; accreditation and professional body requirements; institutional graduate attributes and capabilities; input from external advisory committees; results of cyclical review; and practices of peers in other universities.
- Lesson 3: engage in 'calibration' and capacity building among academic peers. This project argues for the value of 'calibrating' academic staff in the discipline. In the context of grading this means academic staff 'tuning' their 'judgement-making ability' to ensure that grading is valid, reliable and self-regulated.

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. Furthermore, the way in which academic achievement standards are assured needs to be transparent to colleagues, students, quality assurance agencies and the wider society.' (Sadler, 2012, p.14).

4. Towards a learning standards framework: putting the parts together

Figure 1 represents one approach to designing a purpose-driven framework for assuring learning standards.

- Level 1 (department level) focuses on assuring learning standards through the use of moderation and calibration activities among marking teams (e.g., among teams of sessional staff at the unit level) prior to marking, and during or after marking, moderation activities. This should take place every time a unit is offered.

Purpose: to assure validity and reliability of assessment practices through ongoing calibration of markers at the unit/subject level.

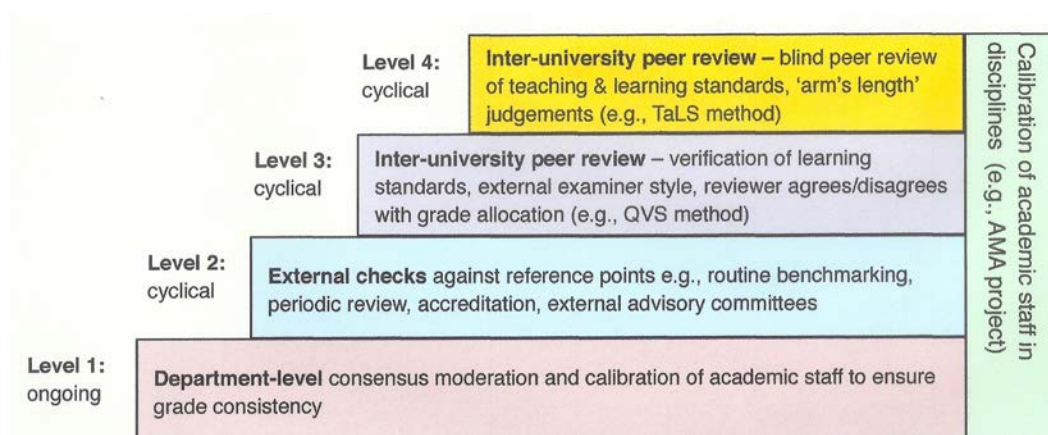


Figure 1. A framework for monitoring and assuring learning standards in the disciplines

- Level 2: External checks take place on a cyclical basis as a way to benchmark learning and teaching standards. Meeting the requirements of accreditation and professional bodies may be included in Level 2.
Purpose: to benchmark processes and outcomes and to address external accreditation requirements.
- Level 3 involves inter-university peer review and verification of grades and standards using an external assessor approach. The identity of the external assessor is known, no effort is made to engage in blind peer review and graded assessment items are shared for verification purposes (i.e., the external assessor either agrees or disagrees with the grade allocated).
Purpose: to verify learning standards across institutions by agreeing/disagreeing with grades allocated to final year assessment items.
- Level 4 involves blind peer review where two external peer reviewers receive deidentified unit materials and ungraded assessment items (i.e., the identity of the institution and the unit is not divulged). The two reviewers grade assessment items using criteria provided by the home institution. Feedback and graded items are returned to the home institution via a third party (project officer). The home institution receives feedback from two partners

to inform practice. The identity of partners may be divulged by mutual agreement to enable further discussion.

Purpose: to provide 'arm's length' assurance of learning standards across institutions through blind peer review and grading of final year assessment items.

- Discipline-based approaches to calibrating academic staff continue (e.g., the Achievement Matters project).

Purpose: to calibrate academic staff across institutions, using external course-level reference points such as discipline standards

5. In summary

5.1 This paper argues for the value of developing a learning standards framework based on academic judgements and peer review of standards in disciplinary contexts.

5.2 Rather than advocating for a single approach to peer review of standards, a multi-level, holistic approach is advocated, based on the purpose of the activity. At the heart of this framework is the need to ensure that academic staff engage in regular calibration activities through the use of peer review of inputs (teaching standards) and outputs (learning standards). Calibration activities should include articulation of reference points that shape decisions about academic standards in the discipline, along with consideration of how judgements about unit-level learning standards are made in the context of course learning outcomes.

5.3 Inter-university peer review may comprise both verification and arm's length blind peer review of standards, depending on purpose. An institution may use multiple forms of peer review and calibration to provide evidence that it is monitoring and assuring learning standards. Each year, an institution may publish (e.g., on the MyUniversity website) information about types of peer review, peer review partners (including overseas reviewers), names of units/courses reviewed annually, steps taken to ensure that peer reviewer feedback is actioned, evidence of success in meeting and exceeding learning standards (e.g., feedback on cyclical reviews, feedback from industry, external advisory committees, accreditation outcomes).

For further details contact Project Leader: Professor Kerri-Lee Krause (k.krause@uws.edu.au).

References

AHELO Web site: www.oecd.org/edu/ahelo

OLT Threshold Learning Outcomes:

<http://www.olt.gov.au/resources/1134,4060?solsort=score%20desc>

Sadler, D.R. (2012). Assuring academic achievement standards: from moderation to calibration, *Assessment in Education: Principles, Policy & Practice*, DOI:10.1080/0969594X.2012.714742.

Teaching Standards Framework:

http://staff.mq.edu.au/teaching/projects/deewr_online_teaching_standards_framework_project/ TEQSA Higher Education Standards Framework: <http://www.teqsa.gov.au/higher-education-standards-framework>

Tuning Europe Web site: <http://tuning.unideusto.org/tuningeu/>

Status of Discussion Paper

This discussion paper arises from issues raised during the course of this national, cross-institutional project on approaches to monitoring and assuring learning standards. It is the result of collaboration among project team members, the project steering group, project participants and associates of the project. It is intended to contribute to the national discussion of approaches to monitoring and verifying learning standards. The views expressed in this report do not necessarily reflect the views of the funding body, the Australian Government Office for Learning and Teaching.

Acknowledgements

- 1 Learning and Teaching Standards (LaTS) Project: Peer review and moderation in the disciplines. Led by the University of Western Sydney with partner universities: Charles Darwin University, Griffith University, La Trobe University, Macquarie University, Queensland University of Technology, The Australian National University, The University of Melbourne. Funding for this project has been provided by the Australian Government Office for Learning and Teaching.
- 2 We gratefully acknowledge the invaluable feedback and insights of Professor Mark Freeman, University of Sydney, in this paper.

Feature	LaTS	QVS	Achievement Matters: Accounting
discipline focus	multiple disciplines across universities		discipline-focussed, multiple institutions
method: key points of similarity/difference	<ul style="list-style-type: none"> • blind peer review, deidentified unit materials using feedback sheet • assessment samples provided in 4 grade bands • all <u>grades removed</u> • <u>peer reviewer grades</u> 4 pieces of work using home university criteria • peer reviewers expected to make judgements <u>in the context of external reference points</u> (eg discipline standards, AQF) but these are not made explicit 	<ul style="list-style-type: none"> • unit materials provided to peer reviewer • graded assessment samples provided • <u>grades provided</u> • peer <u>reviewer verifies/agrees/disagrees</u> with grade allocated by home university 	<ul style="list-style-type: none"> • <u>double-blind</u> peer review, deidentified assessment samples and input materials • assessment samples <u>randomly</u> drawn across all grades • all <u>grades and markings removed</u> • <u>two peer reviewers rate</u> work using nationally agreed discipline standards and rate validity of task • in groups prior to review, calibration occurs to achieve <u>consensus on assessments (not) meeting national standard and assessment design (not) valid</u> • <u>masters</u> as well as bachelors
unit/mode of comparison	<ul style="list-style-type: none"> • unit-level • assessment items re-marked/graded 	<ul style="list-style-type: none"> • unit level • grade/mark verified but not re-marked 	<ul style="list-style-type: none"> • unit level in context of discipline TLOs • assessment items re-marked against national standard (ie. not meeting to meeting continuum) <p style="text-align: right;"><i>continued over page</i></p>

sampling	<ul style="list-style-type: none"> • stratified random sampling of assessments from final year students in selected course • one sample from each grade band for one assessment task (total of 4 samples) 	<ul style="list-style-type: none"> • stratified random sampling of assessments from final year students in selected course • 5% from each grade band • maximum 25 items from large classes 	<ul style="list-style-type: none"> • institution nominates task that best demonstrate TLOs in the discipline, project manager chooses samples • 5 random samples
peer reviewers	<ul style="list-style-type: none"> • <u>two partner institutions</u> review same material • reviewers to be experienced in the discipline, not sessional staff, preferably unit coordinator • blind assignment of reviewers by project officer 	<ul style="list-style-type: none"> • <u>one reviewer</u> • specified as Level D or above (not always) • selected from a panel rather than blind assignment 	<ul style="list-style-type: none"> • <u>two</u> reviewers filtered for substantial experience • all reviewers calibrated to national standard, with practitioner participation in calibration • blind assignment of reviewers, once calibrated, by project manager
basis of comparison	<ul style="list-style-type: none"> • <u>teaching standards</u> reviewed through user guide and feedback sheet – feedback on unit content, assessment design, criteria • <u>learning standards</u> reviewed – grades allocated by two partners for the purposes of comparison, with rationale 	<ul style="list-style-type: none"> • <u>teaching standards</u> reviewed through general comment on unit, no guided feedback sheet • <u>learning standards</u> reviewed – grades verified (i.e., agreed/disagreed) 	<ul style="list-style-type: none"> • <u>teaching standards</u> reviewed through <u>online</u> user guide and feedback form – feedback on assessment design, criteria • <u>learning standards</u> reviewed through online user guide and feedback form – rating allocated by two, unknown, calibrated external reviewers and calibrated reviewer from home institution; third external reviewer moderates consensus if first two disagree on rating ('not meeting' to 'meeting' continuum) <p style="text-align: right;"><i>continued over page</i></p>

<p>result of comparison</p>	<ul style="list-style-type: none"> • home university receives graded assessment items and feedback on teaching standards from two partner universities/peers on 4 pieces of work • identity remains unknown unless partners agree to discuss outcomes 	<ul style="list-style-type: none"> • home university receives comments on learning outcomes and assessment tasks from one peer reviewer • confirmation or disagreement on graded assessment items 	<ul style="list-style-type: none"> • home university receives comments and ratings online on learning standards and teaching standards when published after deadline (unless disagreement between externals requires a third external to reach consensus on final judgement of (not) meeting national standard) • comments on learning outcomes and assessment tasks in context of discipline standards • identity remains unknown
-----------------------------	---	---	---