Assessing the assessments: evidencing and benchmarking student learning outcomes in chemistry

Final Report 2018

Lead Institution: The University of Sydney

Partner Institutions: The University of Adelaide, Curtin University, Deakin University, Macquarie University, Queensland University of Technology, University of Wollongong

Project Leader: Associate Professor Siegbert Schmid

Team Members: Associate Professor Glennys O'Brien, Professor Simon Pyke, Dr Samuel Priest, Dr Madeleine Schultz, Dr Daniel Southam, Professor Adam Bridgeman, Associate Professor Gwen Lawrie, Dr Simon Bedford, Associate Professor Kieran Lim, Dr Ian Jamie

http://chemnet.edu.au/assessment

Support for the production of this report has been provided by the Australian Government Department of Education and Training. The views expressed in this report do not necessarily reflect the views of the Australian Government Department of Education and Training.

2018

ISBN 978-1-76051-369-6 [PRINT]
ISBN 978-1-76051-367-2 [PDF]
ISBN 978-1-76051-368-9 [DOCX]

Acknowledgments

We are grateful to the members of the chemistry community who submitted assessment items for evaluation. We also acknowledge the people who attended workshops run as part of this project.

We acknowledge the invaluable work of our External Reference Group: Associate Professor Mark Freeman (The University of Sydney), Professor Liz Johnson (Deakin University), Professor Will Price (University of Wollongong), Professor John Rice (Australian Council of Deans of Science) and Dr Chris Thompson (Monash University).

Special thanks go to our External Evaluator, Emeritus Professor Carmel McNaught, The Chinese University of Hong Kong, for all her advice, especially in the early stages of the project, and her support throughout.

We also would like to acknowledge the support of the Office for Learning and Teaching in the Australian Government Department of Education and Training.

List of acronyms used

ChemNet	Chemistry Discipline Network
CoP	Community of Practice
CTLO	Chemistry Threshold Learning Outcome
HESF	Higher Education Standards Framework
LTAS	Learning and Teaching Academic Standards
OLT	Office for Learning and Teaching
PRoAN	Peer Review of Assessment Networks
RACI	Royal Australian Chemical Institute
TEQSA	Tertiary Education Quality Standards Agency
TLO	Threshold Learning Outcome

Executive summary

The primary aim of this project was to develop a tool that enables academic staff to evaluate the fitness for purpose of their assessment tasks. This was important because measurement of the capacity of a task to allow students to prove their achievement of threshold learning outcomes is critical within the new Australian tertiary education regulatory framework. Development of the tool was carried out in multiple stages over the duration of the project and the tool itself, available on the project website, forms an important output.

The tool first classifies the level of engagement of an assessment task with a learning outcome at one of three levels—A (addressed), meaning that the task exposes students to some aspect of the learning outcome; D (demonstrated), meaning that students are explicitly directed to demonstrate their achievement of the outcome; or C (credited), meaning that students are given some credit or marks for their achievement of the outcome. For tasks that achieve a C for a particular learning outcome, a much more detailed analysis is then performed examining the depth and breadth to which the learning outcome is credited, and whether the student could pass or not without achieving at that level. After the lengthy iterative process of developing and refining the tool, the final set of 45 assessment tasks was re-evaluated with the final version of the tool. At this stage, the large project team was critical to the success of the project because the evaluation process is demanding. Even at this final stage, consensus between team members was sometimes difficult to reach regarding a task's rating. This illustrates how much personal experience and values can influence views on the functioning of assessment tasks.

The process of building this tool involved many stages of testing against sets of assessment tasks, individual and group reflection and discussions, refinement of the tool to combat observed problems with its application, further testing with other assessment tasks and further refinement. The final tool is sophisticated and can be applied to evaluate any assessment task against any desired learning outcome. During the process, the project team experienced a transformation in understanding of how assessment can and should operate, and this is an important outcome of the project.

The project outcome of increased awareness of, and deeper thinking about, assessment extends beyond the project team because the chemistry academic community in Australia was involved in the project at many different stages. First, all chemistry academics were invited to submit assessment tasks to help with testing and refinement of the tool, and around 12 were generous with their time and shared their tasks. During the submission process their attention was drawn to some of the same issues encountered by the project team. Later, chemistry academics were invited to participate in free workshops during which the current version of the tool was discussed and again they were exposed to the critical issues being discussed within the project team. Additional assessment tasks were evaluated during the workshops and tasks contributed by workshop participants were included. Feedback from members of our community indicate that these workshops were transformative for some participants and all have deepened their appreciation of the importance of assessment design.

The outcome of expanded appreciation of the importance and complexity of assessment was achieved in parallel, and intertwined with, the process of developing and refining the evaluation tool. The significance of chemistry academics taking ownership of a new view of assessment cannot be emphasised too strongly. The adoption of the new federal regulatory regime in Australia requires that students must be provided with tasks that allow them to prove their achievement of diverse learning outcomes beyond subject matter knowledge. To adequately design and assess such tasks, academic staff must appreciate the inadequacy of their previous modes of assessment. Within the new regime, assessment is more precisely specified through marking criteria, making it fairer and more reliable. Detailed checking of assessment tasks against desired learning outcomes also ensures their validity. We are pleased with the progress in this area but also aware that a single project is not sufficient and ongoing professional development will be required to sustain this outcome. The Chemistry Discipline Network gives a platform to continue this work.

A secondary aim of the project was to collect a set of exemplar assessment tasks for different learning outcomes that can be adopted or adapted by others for their own use. This was achieved, although not every chemistry threshold learning outcome (CTLO) is covered within the set of exemplars so far collated. The exemplars are available within a searchable part of the project website. This output is available to the community and is also of interest internationally.

A final output that was not in the initial project design is an online version of the workshop that was built in response to the positive feedback received about them. This will allow people to experience the transformational learning experience that applying the tool gave to workshop participants.

Table of contents

Tables and figures

Tables

Figures

Chapter 1: Background

In Australia, and increasingly worldwide, higher education institutions define attributes that graduates are expected to attain through their education (Barrie, 2004, 2007). Such attributes aim to describe, in the most general terms, what a graduate of that institution knows, understands and can do. In most cases, they include both academic and societal aspects such as community responsibility and ethical behaviour (Hager, Holland & Beckett, 2002; Macquarie University, n.d., The University of Sydney, n.d). Although frequently aspirational, graduate attributes or capabilities illustrate the philosophy of each institution and, to some extent, inform the curriculum as a series of outcomes (De la Harpe & David, 2012; Hughes & Barrie, 2010). Such outcomes can be aligned to shared national (American Chemical Society, 2015) or international (Pinto, 2010) normative practices and present a complex array of imposts on curriculum.

The elucidation of outcomes within a curriculum is seen as an essential cue to both learner and teacher about the intention of a learning environment (Biggs, 1996). Effective alignment of the objectives of the teacher to the outcomes for a learner is mostly influenced by what students do (Biggs, 1999) and, within the learning environment, this is most frequently measured through assessment (Meyers & Nulty, 2009). Assessment of such outcomes, especially those that are shared (Hager, 2006) or are transferable across discipline boundaries (Kemp & Seagraves, 1995), presents challenges (Green, Hammer & Star, 2009). These challenges emerge from a shared understanding of the outcomes between teacher and learner as well as from those outside the immediate learning environment.

Assessment aligned to desirable outcomes can establish lifelong learning (Boud & Falchikov, 2006) and skills necessary for employability (Jackson, 2015). In the science curriculum, there is frequently a gap between what is intended by teachers and what is actually achieved by learners (Van den Akker, 1998) and, by extension, there is a gap in how the learner is assessed (Bryce & Robertson, 1985). This is especially true when intangible outcomes are reduced to facile assessment practices (Bowman, 2013). Thus, reforming assessment is necessary to improve outcomes in higher education by designing better tasks that clearly identify thresholds, and specifying how these tasks contribute towards the attainment of a degree (Sadler, 2015).

In this context, the corresponding regulatory requirements for Australian institutions have been expressed in the Higher Education Standards Framework (HESF, Australian Government, 2015, p. 4) as follows:

1. *The expected learning outcomes for each course of study are specified, consistent with the level and field of education of the qualification awarded and informed by national and/or international comparators.*
2. *The specified learning outcomes for each course of study encompass discipline-related and generic outcomes, including:*
    - *specific knowledge and skills and their application that characterise the field(s) of education or disciplines involved*
    - *generic skills and their application in the context of the field(s) of education or disciplines involved*

- *knowledge and skills required for employment and further study related to the course of study, including those required for registration to practise if applicable, and*
- *skills in independent and critical thinking suitable for life-long learning.*

3. *Methods of assessment are consistent with the learning outcomes being assessed, are capable of confirming that all specified learning outcomes are achieved and grades awarded reflect the level of student attainment.*

4. *On completion of a course of study, students have demonstrated the learning outcomes specified for the course of study, whether assessed at unit level, course level, or in combination.*

Statements 3 and 4 above indicate that for institutions to satisfy these requirements their methods of assessment must be evaluated to ensure they allow demonstration of learning outcomes. In addition, they imply that all required learning outcomes must have been demonstrated by every graduate. To ensure that this is the case, the degree structure must be conditional on attaining the corresponding learning outcomes. Hence, individual assessment tasks must be structured to facilitate the explicit assessment of these outcomes.

Descriptions of required learning outcomes to which this regulatory framework applies have been developed by discipline communities through the Learning & Teaching Academic Standards (LTAS) project (Ewan, 2010). That project was established in 2009 by the Australian Learning and Teaching Council to facilitate and coordinate the definition and implementation of academic standards by discipline communities. The Science LTAS project developed overarching threshold learning outcomes (TLOs) for bachelor degree graduates (Jones, Yates & Kelder, 2011; Kelder & Jones, 2015). The science TLOs, and their derivatives, all contain a common structure, grouped around a series of broad outcome statements (first tier) that are the bases for the more functional statements at the stem (second tier). Read together, each base and stem embodies a particular aspect of knowledge, skills and/or attributes that every graduate of the discipline will have explicitly demonstrated upon graduation through assessment.

Within chemistry, the discipline community developed chemistry-specific TLOs as a derivative of the science outcomes—the chemistry threshold learning outcomes, the CTLOs (Buntine, Price, Separovic, Brown & Thwaites, 2011). The CTLOs (Appendix L) were further elucidated through a series of meetings organised by the Chemistry Discipline Network (ChemNet), some of which were sponsored by the Royal Australian Chemical Institute (RACI; Pyke, O'Brien, Yates & Buntine, 2014; RACI; Schultz, Mitchell Crow & O'Brien, 2013; Schultz & O'Brien, 2017).

As a result of the ChemNet meetings in 2012 and 2013, further levels of detail have been elucidated for CTLOs 2.1 and 3.3 as a third tier. Within the third tier, CTLO 2.1 has been expressed as a list that constitutes the core principles and concepts of chemistry, while CTLO 3.3 lists the practical techniques and tools considered fundamental to this science.

With the regulatory framework and CTLOs established, an approach to determine whether methods of assessment are adequate to confirm attainment of CTLOs was required. The design of assessment tasks is critical because appropriate assessment task design can optimise student learning (Boud et al., 2010; Gibbs & Dunbar-Goddet, 2009; Gibbs & Simpson, 2004; Rice, 2011; Sadler, 2015). In contrast, poor task design may, for example:

- prevent a 'good' student from demonstrating a high level of capability
- prevent an 'average' student from meeting minimum performance requirements
- allow a 'poor' student to obtain a passing grade without having met required threshold outcomes.

Much has been written about assessment design (Boud & Falchikov, 2006; Boud et al., 2010; Gibbs & Dunbar-Goddet, 2009; Gibbs & Simpson, 2004; Nicol, 2009; Sadler, 1989; Scouller, 1998), yet academic staff who are designing assessment tasks are often unaware of those reports. This makes review of assessment tasks against shared outcomes by peers challenging, requiring extensive consultation to permit effective benchmarking (O'Connell et al., 2016). Two recent projects in Australia in other disciplines have addressed the utility of peer review of assessment in the context of comparability and standards when implementing TLOs (Booth, Beckett & Saunders, 2016; Krause et al., 2014). Both found that if the design of assessment tasks and their review against a priori shared outcomes is to be effective, a shared understanding must be developed within an agreed framework.

A notable issue for the appropriateness of current assessment practices is the imposition of the new paradigm of confirming attainment of the TLOs on to existing assessment practices. In the context of this project, assessment confirming attainment of a TLO, a useful definition of 'assessment' can be stated as '*a **measurement or judgement** of student capability with respect to a specific **learning objective'***. Therefore, student capability must somehow be judged so that their attainment of the TLO (or other stated learning objectives) can be confirmed. The fact that the stated TLOs are now mandatory is a fundamental change to many existing assessment paradigms where the desired learning objectives have historically been mainly aspirational. The TLOs are now mandated *outcomes*, not idealistic *objectives* that may or may not be attained by any given graduate. This places the focus on what is observable in student *outputs*, rather than on ensuring teachers provide a sufficient number of *inputs* for learning. This radical shift means that many tasks that may have reasonably been claimed as an assessment of a stated learning objective in the past may be found to be insufficient for confirming attainment of the TLO.

When developing the project scope, the team considered the current body of knowledge that existed in terms of peer review of assessment, benchmarking of assessment standards and what has become known as external referencing. This included both national and international perspectives for assuring that methods of assessment could allow students to demonstrate the required learning outcomes. We looked at several Australian models of external referencing of achievement standards established within the higher education sector. These were at various stages of development, testing and implementation. Some models are similar in that they examine the validity and reliability of teaching and learning standards, yet they differ in their methodology with regard to the depth and breadth of the review (Deane & Krause, 2013). The *Teaching and learning standards project* process requires assessment grading of clean copies of stratified randomly sampled assessment tasks across four grade bands. The methodology of the Group of Eight Quality Verification System (QVS) system requires that reviewers judge the accuracy of the marked assessment items from a stratified random sample across five different grade bands. The *Academic calibration project* of the innovative research institutions is closely aligned to the QVS system. The *Peer review of assessment networks project*, with its sector-wide options for calibrating and assuring achievement standards within and across disciplines and other networks (PRoAN, University of Tasmania), demonstrated a need to support the higher education sector, particularly those institutions that are non-aligned to university networks, in using peer review to enhance and assure the quality of both the inputs to and the outcomes of assessment (Booth, Beckett & Saunders, 2016).

Other recent projects have focused on discipline standards. For example, the *Achievement matters: external peer review of accounting learning standards* project is a double-blind process focused on consensus moderation on the achievement of course-level learning outcomes of randomly sampled student work drawn across all grades. The external examining system in the UK is under review in 2016–18 under the leadership of the Higher Education Academy, and is tasked to consider what, if any, further changes might be required to improve the assessment standards. This project has taken a similar approach using a mix of peer review of assessment standards and discipline standards and concluded that generic staff development as well as precalibration events or consensus moderation within disciplines are needed.

In this context, in 2014, the Office for Learning and Teaching funded this project *Assessing the assessments: Evidencing and benchmarking student learning outcomes in chemistry* (OLT ID14-3652) with primary goals of establishing a common understanding of how the CTLOs can be applied and developing a tool to allow academic staff to evaluate the ability of their assessment items to show demonstration of particular TLOs.

Chapter 2: Approach/Methodology

The project approach centred on team members and academics in the tertiary chemistry community developing shared understandings of assessment standards within an action research framework (Appendix C). A series of concurrent activities were used to develop a framework or procedure for evaluating the fitness of assessment tasks to provide evidence of attainment of a CTLO. The interconnected activities involved collecting and evaluating tasks, developing the framework via an iterative process using the tasks, and providing professional development workshops to the tertiary chemistry community. In the workshops, assessment tasks were collected and the framework was trialled at various developmental stages. Further evaluation of the tasks by the team was used to identify exemplar tasks and potential exemplars. Thus, the elements of Sadler's requirements for developing assessment standards were applied: examples were collected (as opposed to Sadler's exemplars) and shared judgements of those tasks were used in developing the framework; team meetings, especially workshops, were used to provide ways of opening up tacit knowledge to be shared among academics (Sadler, 2009).

2.1 Online portal for lodging assessment tasks

The first stage of this project required preliminary evaluation of several diverse assessment tasks to obtain an overview of what would be involved in the evaluation process. It was critical to not only have access to the assessment task as given to students, but also to associated documentation (assessment brief), including the criteria for assessment as well as samples of student work to see how the criteria were applied (student performance standards). The first assessment tasks evaluated were provided by members of the project team from their own teaching practice. Over a series of meetings, the team developed and refined a proforma to gather required information, along with the assessment task and associated documents, via an online submission portal (Appendix D).

The portal gathered data in sections via questions and checkboxes—first by collecting information about the task, including how it fitted into the assessment pattern of the unit or subject of study in which it was used, what it aimed to do, and whether it was compulsory for all students enrolled in a chemistry major. The portal also asked what types of knowledge and which, if any, skills and cognitive processes were required to complete the task based on Bloom's revised taxonomy (Krathwohl, 2002). The next section allowed the submitter of the CTLOs' self-evaluation of what was demonstrated (fully or partially) by successful completion of the assessment task. Finally, submitters were asked what quality assurance processes were used in developing the task and how they ensured that the assessment task was valid and reliable.

2.2 Workshops: conversations with colleagues

In parallel with collecting assessment tasks through the portal, a series of professional development workshops was held around Australia to inform the chemistry community about the project, to trial and disseminate the evaluation process (see Appendix E), and to improve assessment practice. The workshop elements were (i) Introduction to the learning and teaching landscape and HESF, (ii) exercise in and discussion of purpose of assessment,

(iii) exercise in and discussion of Bloom's revised taxonomy, (iv) collection of tasks via the portal, and (v) evaluation of sample assessment tasks in small groups (see Appendix F and G).

Discussions at these workshops were guided by the project team's current suggested evaluation process (which changed over time) and served to increase awareness within the chemistry education community of the importance of careful assessment design and the application of the CTLOs. The workshops highlighted key issues in the design of assessment tasks to meet the emerging regulatory requirements. The discussion also contributed to the ongoing development of the project team's evaluation procedure. Further details of the process and outcomes of these workshops are reported in section 3.3.

2.3 Evaluation of tasks ⇌ Development of evaluation framework

Independently of the workshops, the team held a series of internal meetings to develop the procedure for evaluating assessments. This evolved out of the proforma questions used in the online portal and the collaborative discussions at the meetings, this was based on the experiences and knowledge gained by assessing the items submitted and commentary from workshop trials of the process. In addition, because only a small number of assessment tasks were analysed in the workshops, evaluations of submitted tasks were carried out individually by submitters using the online portal and then by members of the project team. Following this, collaborative discussion within the project team led to a consensus evaluation of each assessment task. This extended process, together with the workshops, aimed to build a shared understanding of the CTLOs and what is required for an assessment task to confirm their attainment. Furthermore, tasks were identified as exemplars and potential exemplars, which could later be made available, with the appropriate approvals, to improve assessment quality across the sector. Finally, difficulties with the application and wording of some of the CTLOs were identified. The resolution of these is beyond the scope of this project but their identification is important for future work.

Thus, the evolving evaluation framework has formed the core of the project. The team has used an iterative process where the framework developed out of the original proforma and was further improved through trialling at workshops and by team member evaluation of tasks and of the framework itself. This iterative process using both nuanced expert use by team members and feedback from the workshops has yielded a sound outcome—the evaluation framework housed in the evaluation tool. This will be available via the project website and will be easily used as it is a spreadsheet-based tool. It will be held with explanatory notes.

The conduct of the project was approved by the University of Sydney Human Research Ethics Committee (2016/490).

Chapter 3: Outcomes

The main outcome of this project sits within the professional development of chemistry academics across the country due to the shared understandings, experiences and developed assessment practices arising from taking part in the project. This applies to the team members, those who submitted tasks, workshop participants and peers of workshop attendees. This outcome results from development of the main project components: the evaluation framework and tool, the evaluation of tasks submitted and the professional development workshops. These all progressed concurrently, each intertwined with and informing the others.

The outputs are the evaluation tool itself, the set of exemplar tasks, and an online workshop to support academics developing assessments and using the tool and the exemplars. These are housed in the project website, which is attached to the Chemistry Discipline Network website (http://www.chemnet.edu.au). Dissemination of the project has been carried out at multiple conferences and via a book chapter (Appendix I).

3.1 Task-evaluation tool

3.1.1 Development and rationale

The development of the evaluation framework and tool came about through an iterative process with inputs to and feedback from multiple team meetings and workshops (Appendices C, D and E). The first task within the evaluation procedure of the project involved two main stages: evaluators were asked whether the task assessed each of the five broad (first-tier) CTLOs, then subsequently asked whether attainment of these CTLOs was visible in samples of student work provided. Discussions following these exercises quickly revealed that what a 'pass-grade' student had actually demonstrated within their work often covered fewer CTLOs than submitters expected based on the task's design. Task submitters often claimed that CTLOs were assessed in their tasks when students had no real opportunity to demonstrate it within their work. Instead, the task often served only as a *learning experience* relevant to the CTLO, not an evaluation of student capability. A related issue was a considerable lack of detail in marking guidelines of many tasks. Consequently, evaluators could not be confident that particular CTLOs were required of students in those cases. These observations led the project team to be increasingly conscious of the subtle difference between the tasks addressing a CTLO versus those that explicitly assessed students' capability with respect to that CTLO.

Another issue, which shaped the evolving task-evaluation procedure, was the finding that many tasks would address a given CTLO only in part and this occurred for different reasons as outlined below.

- A task may address **only some subpoints** of a CTLO (e.g. CTLO 1.1, but not 1.2).
- The task may address a CTLO at an **insufficient depth**, that is, below the threshold level expectation for a graduate.

- The task may address a CTLO at an **insufficient breadth** to satisfy the full graduate-level expectation (for example, no tasks could address the full range of chemistry content knowledge expected to be covered over a whole degree).

To resolve these issues, the task-evaluation procedure shifted to focusing on the 15 second-tier CTLOs (e.g. 1.1, 1.2). The response format was also changed to allow for partial CTLO coverage for reasons of either depth, breadth or both independently. This was achieved by using a ‘four-square’ response format (Figure 1).

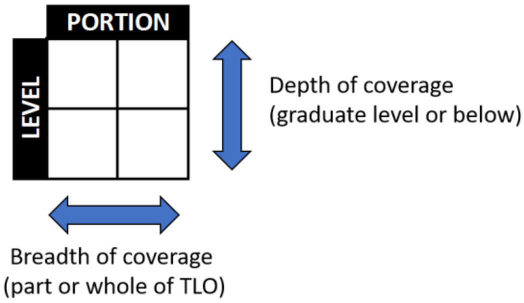


Figure 1: The ‘four-square’ response grid

This grid allowed the evaluator to report full or partial ratings for CTLO engagement, based on considerations of depth and breadth independently, by shading different quadrants of the grid. The horizontal dimension of the grid (‘portion’) allowed the evaluator to report what portion of the CTLO statement’s breadth was engaged (part or whole), while the vertical dimension of the grid allowed the evaluator to report the level at which that portion of the CTLO was engaged (graduate level or below). This provided five distinct response options for evaluators to report varying degrees of CTLO coverage within a task (Figure 2).

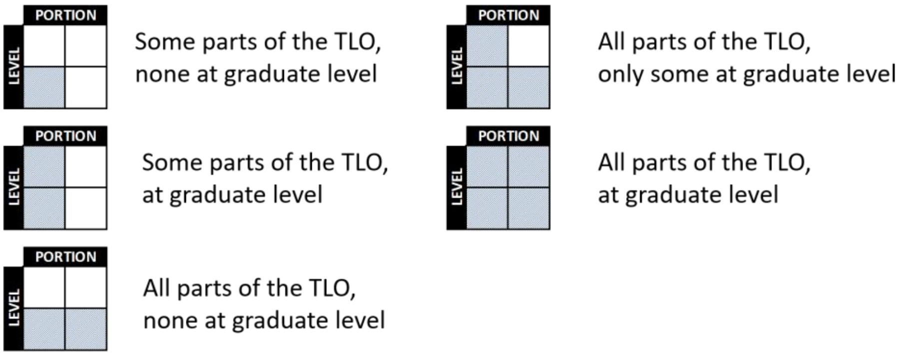


Figure 2: Possible responses using the ‘four-square’ grid

This response format was found useful, so was placed within the ‘assessment rating’ system. The assessment rating system was devised in response to the need to encapsulate a prevalent observation in the evaluation of tasks: different CTLOs were assessed at different levels of rigour, depending on task design and the requirements built into assessment criteria. Some CTLOs appeared to be core requirements of passing the task, whereas others appeared to be more incidental. The following three critical issues shaped the structure of the final rating system devised.

- CTLOs could be addressed 'incidentally', without directing students to demonstrate their attainment, for example, using an industrial or medical topic (CTLO 1.2) as context for a question that could be answered solely by using chemistry content knowledge (CTLO 2.1).
- Students could demonstrate their capabilities with respect to a CTLO, without that demonstration adding to their final mark, for example, where students execute experimental techniques (CTLO 3.3) but are marked solely on practical reports (CTLOs 3.1 and 3.4).
- CTLOs could contribute towards the marks or grade for a task, but to such a small extent that a student could pass without demonstrating them by aggregating other marks to 50 per cent.

These possibilities were used to structure a hierarchical set of ratings that are awarded to tasks evaluated for each CTLO (see Table 1). Furthermore, using this rating system, task evaluators complete the four-square response only where a CTLO is 'credited' (C), that is, if the CTLO contributes towards a student's marks or grade for the task. For CTLOs that receive the 'demonstrated' (D) or 'addressed' (A) rating, precise detail regarding CTLO coverage is unnecessary for purposes of checking HESF compliance, because these cases make no formal judgement of student capability and therefore could never confirm student CTLO attainment. However, note that demonstrated (D) or addressed (A) evaluation with four-square rating would be valuable in appraising potential opportunities to capitalise on small changes to assessment criteria to improve a task's rating and this was done when selecting potential exemplars.

Table 1: Assessment ratings in the final tool

<table>
<tr><th colspan="2">Evaluation result</th><th>Description</th></tr>
<tr><td>A</td><td></td><td>The learning outcome is addressed, but students are not required to demonstrate their capability</td></tr>
<tr><td>D</td><td></td><td>Students are required to demonstrate their capability, but are not credited based on that demonstration</td></tr>
<tr><td rowspan="2">C</td><td></td><td>Students are credited based on their demonstrated capability, but a passing grade can be achieved without that credit</td></tr>
<tr><td></td><td>Students are credited based on their demonstrated capability, and that credit is a necessary requirement for a passing grade</td></tr>
</table>

3.1.2 Designing the electronic task-evaluation tool

The final task-evaluation framework format uses a spreadsheet-based tool. The allocation of assessment ratings and completion of the four-square response is achieved through answering questions by selecting responses from drop-down menus. This removes the pressure on task evaluators to have a clear, working knowledge of all concepts present in the four-square and assessment rating systems (Appendix H).

The tool allocates an assessment rating of A, D or C (Addressed, Demonstrated and Credited, respectively) to each CTLO evaluated to be relevant to the task. The four-square response grid is also automatically shaded to reflect the depth and breadth of CTLO coverage for all CTLOs with a C rating, and this is done separately for the two different C ratings (required to pass or not; see Table 1). Cross-hatch shading is used to reflect the depth and breadth of CTLO coverage for which marks are awarded, and full colour shading is used to indicate the depth and breadth of CTLO coverage that is a mandatory requirement of earning a pass for the task—confirms attainment of the TLO. The two variants of the credited rating have been labelled 'credited *without confirmation*' and 'credited *with confirmation'* in the tool.

The array of 20 possible outcomes of the evaluation process for any individual second-tier CTLO is shown in Figure 3. The tool itself does not show this full flow chart; future questions and possible responses are selectively made available to respondents in accordance with responses already given in the tool. Evaluating any given CTLO requires answering a maximum of seven questions.

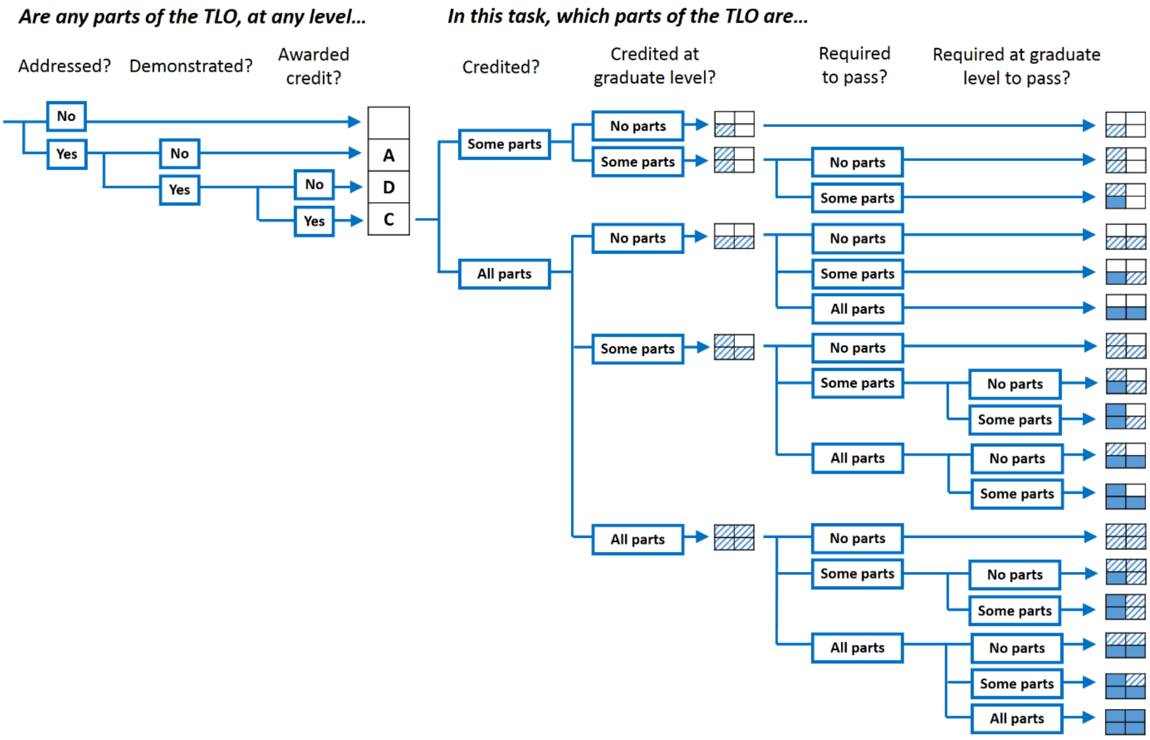


Figure 3: Flow chart map of evaluation procedure showing all possible outcomes

Thus, application of the tool to an assessment task is a stepped process through scoping against the 15 second-tier CTLOs as A, D or C; followed by analysis of breadth and depth for C-scoped CTLOs, which automatically generates the evaluation result. The evaluation summary obtained through application of the tool for a particular assignment task is shown in Figure 4 to illustrate the output.

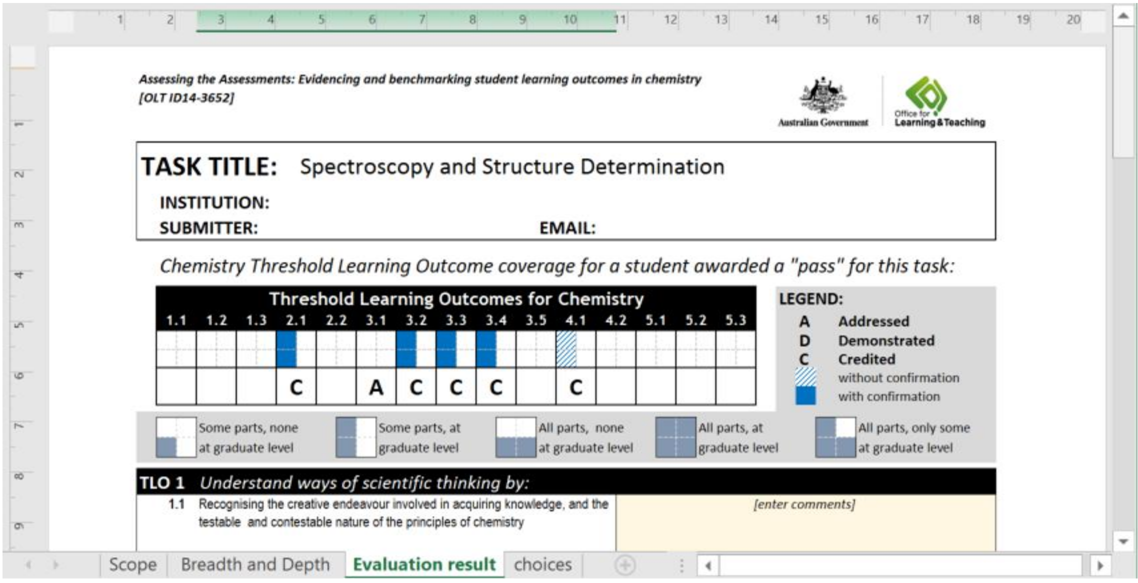


Figure 4: Evaluation summary of an assessment task, a third-year assignment

The overall structure of the task-evaluation tool, as presented here, implies a specific set of requirements of an ideal assessment for any learning outcome. Applying the questions that decide the assessment rating, **a task is capable of confirming a desired learning outcome if:**

- the desired learning outcome is ***addressed***
- students ***demonstrate*** their capability with respect to that desired outcome
- students are ***credited*** (i.e. marked or graded) based on that capability
- such credit is ***required*** to earn a pass for the task.

This inherently presents a conceptual framework for categorising assessment practice, which can be represented as a Venn diagram (Figure 5). Features of the task design (left) and features of the assessment criteria (right) intersect to determine which learning outcomes form core elements of the assessment task (centre of the diagram) and which remain incidental, though relevant (at the periphery of the diagram).

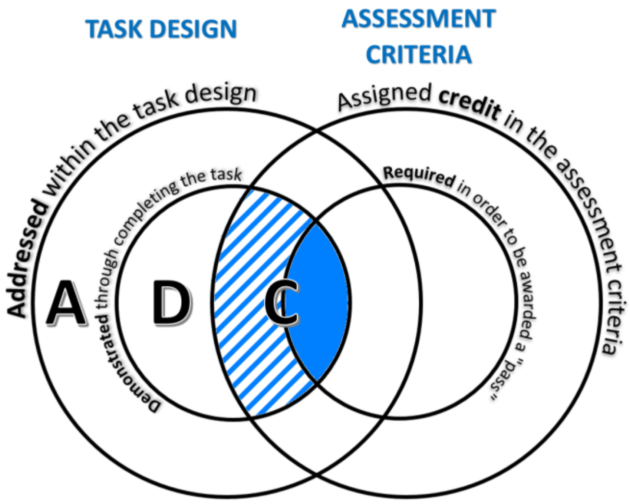


Figure 5: The ADC conceptual framework of assessment practice

In the following sections, this diagram is used to categorise the CTLOs when applied to the assessment items submitted to this project.

3.2 Assessment tasks

3.2.1 The submitted tasks

The state of existing assessment practices in chemistry in Australian universities was known to the team from the experiences of project team members in their home institutions (the team covers eight institutions) and from a ChemNet project-mapping parameters in chemistry units of study across 12 institutions (Schultz et al., 2013). Thus, most assessment types were known to be traditional: exams and laboratory/scientific reports, with some assignment-based tasks. In this project, a total of 45 tasks have been submitted by staff at 22 different universities around Australia. Most of these tasks were submitted via the project online portal before and during the project workshops.

The tasks submitted are used within the three levels of an undergraduate degree: 15 tasks at first-year level; 13 second-year tasks and 17 third-year tasks. The tasks have been categorised into broad assessment types based on the student activity, namely:

- examination (n =10)
- written assignment: set questions; problem-solving; analysis of given data (n = 17)
- lab and/or lab report (n = 8)
- project: assignment with research, analysis and various outputs (n = 2)
- project investigation: with research, lab work, analysis and various outputs (n = 5)
- presentation (n = 3).

These categories have been used to give a general overview of the types of assessment tasks submitted. Clearly, there are possible overlapping categorisations here, for example, all three presentation assessments could be regarded as assignments with a different output. In the assessments submitted, it is notable that examination assessments made up seven of the 15 first-year tasks but none of the third-year tasks submitted. Third-year tasks submitted to the project were more complex written assignments, projects or project investigations with major lab components. This does not imply that at third-year level examinations are not used (Schultz et al., 2013), rather that submitters and the project team looked for variety covering more than only CTLO 2.1 ('body of knowledge') in tasks for evaluation at the third-year level.

Many tasks include more than one type of assessment and most assess more than one CTLO. For example, in some laboratory assessment items, specific separate assessment of laboratory skills in class and a written report are included. Assignment tasks often focus on CTLO 2.1, but where assignments involve more in-depth analysis, argumentation or reasoning then CTLOs 3.4 and 4.1 are also assessed. Twelve of the submitted tasks are structured series of assessments connected to one major task (e.g. project investigation). Within these tasks, several types of assessment are performed, such as assessment of laboratory practice, presentations, various types of reporting and various types of peer assessment. These tasks are predominantly third-year tasks (8 of the 12). The five tasks classified as investigative projects are all third-year tasks.

3.2.2 Assessment of the tasks by the team

After the lengthy iterative process of developing and refining the tool, the final set of 45 assessment tasks was re-evaluated with the final version of the tool. Figure 6 shows the outcome of the project team evaluation of the 45 submitted tasks. Examination of the ADC ratings allocated to the tasks shows coverage of the CTLOs to be variable.

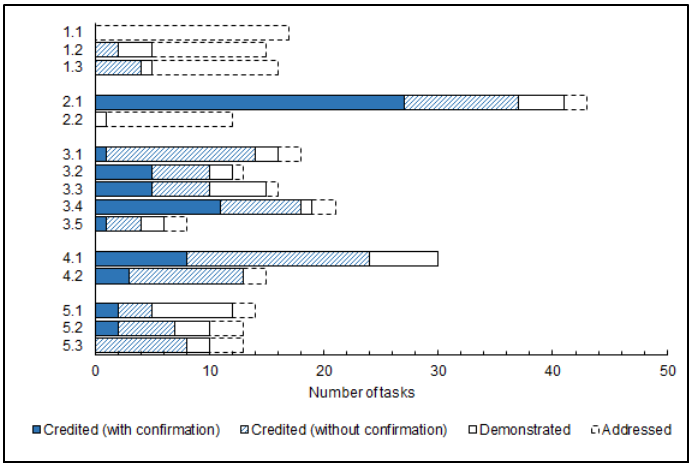


Figure 6: Summary of CTLO coverage among assessed tasks

As Figure 6 shows, all the CTLOs are addressed or demonstrated and most credited in at least some of the submitted tasks. However, fewer tasks were found to include confirmation of attainment of the CTLO. This is because aggregated marks can often carry a student over the pass line without confirming any single CTLO specifically. As would be expected, coverage of CTLO 2.1 (body of knowledge) was widely found.

Importantly, as Figure 6 shows, for all three parts of CTLO 1 as well as for CTLOs 2.2 and 5.3, no tasks were submitted that confirm student attainment. In fact, for CTLOs 1.1 and 2.2 no tasks were submitted where credit is given. The project team did find more than ten tasks that address each part of CTLO 1 and CTLO 2.2, but in very few of these tasks were students given the opportunity to demonstrate the CTLO or given credit for doing so. Based on other activities and experiences of the project team members, there was no surprise at this outcome.

It is clear from Figure 6 that the marking criteria of many assessment tasks determined that credit was given to a student for demonstrating proficiency in a particular CTLO, but the student could pass without that credit. For example, for CTLO 4.1, there were 24 tasks where aspects of 'presenting information, articulating arguments and conclusions, in a variety of modes, to diverse audiences, and for a range of purposes' were assessed with credit given. However, only one-third of those tasks were structured so that a student could pass if, and only if, the criteria for CTLO 4.1 were met, thus confirming CTLO 4.1. This applies to several assessment tasks for several CTLOs.

3.2.3 Exemplar tasks

Through the evaluation process, tasks were identified that would serve as examples of good practice in definitively confirming the attainment of a specific CTLO as ‘exemplar’ tasks. An important output of this project is this set of tasks that can be adopted by other chemistry academics. Tasks that could fulfil this purpose with minor adaptation were also identified.

Some exemplar tasks apply to a single CTLO while some longer, project-based tasks cover several CTLOs. Assessments in tasks covering multiple outcomes must be structured so that the student cannot pass the assessment without meeting the requirements of each of the nominated learning outcomes, and one important way that this can be achieved is through hurdles within the assessment.

Exemplars are characterised by being very well-defined tasks where learning outcomes are clearly stated. Furthermore, student activities, processes, timelines, outputs and other requirements of the students are made very clear in carefully structured documentation about the task. Thus, the tasks definitively allow the student to engage with each chosen learning outcome and to demonstrate their proficiency. Turning to assessment to confirm attainment of the learning outcome, the assessment criteria are clearly laid out using rubrics defining levels of achievement and associated marking criteria covering each learning outcome. Most importantly, those rubrics and assessment criteria are given to students as well as to staff. In major tasks with multiple outputs, each of those outputs necessarily has its own rubric or grading scheme, which may relate directly to individual specific learning outcomes. This is especially necessary in the case of a major task aimed at multiple CTLOs as occurs in a capstone-type unit of study. In contrast, one student output may apply to more than one learning outcome. For example, in an investigation leading to a technical report with an executive summary contextualised for an outside organisation, the report alone can cover CTLOs 3.4, 4.1 and 4.2, provided appropriately detailed marking criteria are supplied.

Ten of the tasks submitted were identified as being of exemplar status for particular CTLOs. Seven of these tasks were from third-year chemistry, being predominantly complex project tasks with multiple outputs and covering multiple CTLOs in the assessment. However, across these seven tasks, the coverage of the CTLOs was found to be narrow and confined to CTLOs 3.2, 3.3, 3.4, 3.5, 4.1 and 4.2. These particular CTLOs focus on inquiry/investigation activities with specific communications outputs. Three exemplar tasks were first-year tasks, covering writing/critique for outside audiences or multimedia production for an audience other than the academic teacher. These were imaginative tasks well removed from the usual lab/tutorial assignment/exam-type assessments. The tasks provide excellent examples of engaging in some knowledge building, argumentation and dissemination with communication via particular modes to different audiences.

3.2.4 Potential exemplar tasks

Nine further tasks were identified that are already clearly structured so that minor changes to the marking criteria can deliver a confirmed status. In general, these tasks already give credit for particular CTLOs and restructuring of assessment can require specific CTLOs to be confirmed for a student to pass. Thus, part of the assessment must be a hurdle, so that students do not pass by aggregation of marks.

In a number of cases, project team members assessed a task as crediting a CTLO at the graduate level, but confirmed the CTLO at the non-graduate level because a student could conceivably pass without providing evidence of operating at the higher level. This arose because the structure of the assessment allowed aggregation of marks, including a non-graduate-level of attainment for that part of the assessment. Minor restructure of grading schemes could prevent this occurring.

Finally, it must be noted that several of the tasks submitted were bereft of information needed to judge their potential to be used as exemplar tasks or would need substantial restructuring to be able to confirm one or more of the CTLOs.

3.2.5 Using the evaluation tool and exemplars to improve assessment practice

An important message from this evaluation of the submitted tasks is that some tasks can be easily modified to allow confirmation of student attainment of specific learning outcomes. The application of the evaluation tool will highlight this to users, as noted in section 3.1, through the questions asked in the scoping and analysis of breadth and depth. Furthermore, the tool allows notes and comments to be recorded so a user can employ the tool developmentally and keeping a record of changes and outcomes.

Indeed, one of the project members has gone down this development pathway with the intention of generating a capstone-type task that confirms several CTLOs from an existing task. Starting from an existing third-year project-based assessment worth 40 per cent of the total unit mark, the project was redesigned with multiple assessed outputs to evidence attainment of several CTLOs and the project now constitutes 100 per cent of a 6-credit point unit. For a CTLO to be confirmed, that CTLO must be worth more than half the marks or must be a hurdle. The assessments are tabulated below with two specified hurdles highlighted (Table 2).

Table 2: Complex project-type task: assessment summary

#	Assess	%	Description
1	Laboratory performance	20	Individual: In-class observation, assessment of laboratory practice and laboratory notebook
2	Individual contribution to group work Hurdle component: minimum 10/20	20	Individual: In four parts: Collaborative activities and practices, contributions to online discussion, peer and self-assessment, self-reflection writing
3	Week 4 Project Proposal	5	Group: Project proposal. Submitted via Moodle
4	Week 9 Results Summary	5	Individual: Table of results to date
5	Major Report Hurdle component: minimum 20/40	40	Individual: Written as a technical report, 30-40 pages
6	Group Presentation	10	Group: Presentation to class

The result of this assessment development is one of the project exemplars. The evaluation summary of the new assessment task by other project members is shown in Figure 7.

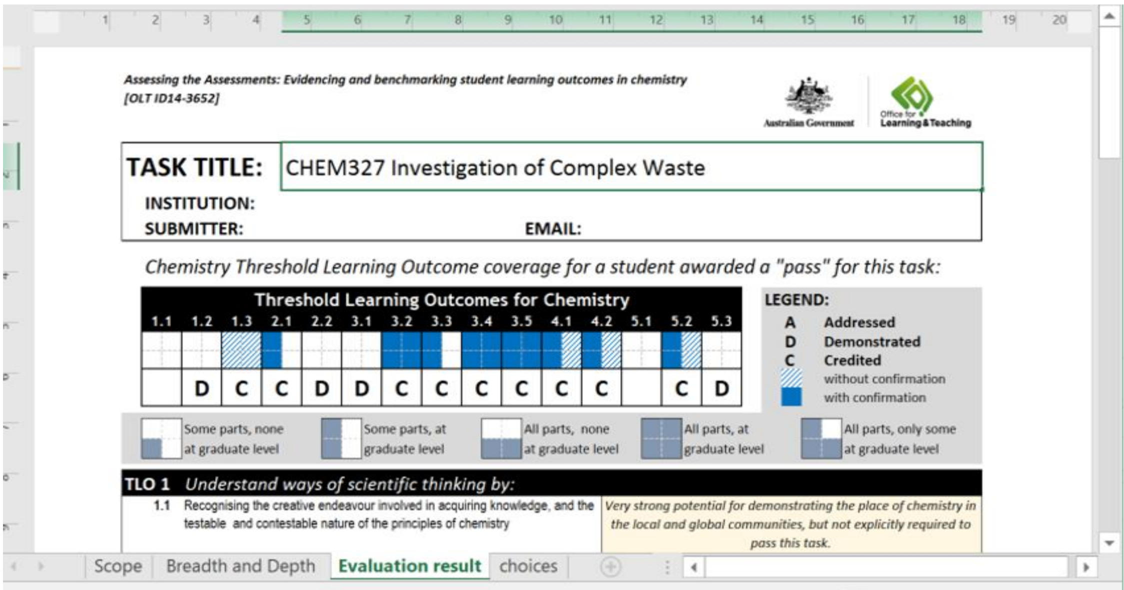


Figure 7: Evaluation summary of the assessment of a complex project-type task

The exemplars can be used by academics in various ways. The assessment practices used in the tasks can be adopted in part or entirely, the structuring of assessment to confirm attainment of a specific CTLO can be adopted, or an assessment task can be adapted. As an example, parts of the peer assessment could be used from the above item for a different type of student activity. Exemplar items will be made available on the project website with associated evaluation and commentary.

3.3 Professional development

3.3.1 Project workshops

The workshop program aimed at supporting professional development of chemistry academics by enhancing their assessment practices. This was focused specifically on assessment to confirm the attainment of CTLOs and the new Australian regulatory framework. While it is acknowledged that there are other purposes for assessment, those are outside the context of this project.

The workshops were designed to provide a professional development sequence. Within the workshop itself (section 2.4 and Appendix F and G) the program included (i) the learning and teaching landscape and HESF, (ii) the purpose of assessment, (iii) Bloom's revised taxonomy, (iv) task collection, and (v) evaluation of tasks. Participants discussed the learning and teaching environment within which the changed practices were needed and developed shared understandings from the consensus moderation of tasks. On returning to their home institution, these 'newly calibrated' academics have their own assessment task evaluated by themselves and others, and one or two other example tasks. Their task may or may not have been fit for purpose, but importantly, they know where that task needs changing and what designs are possible to bring about required improvements. The academics can share this practice within their own institutions.

3.3.2 Purpose of assessment

During task-evaluation sessions in workshops and from portal-submitted tasks, it became evident that traditional assessment practices influenced judgement in task evaluation. In considering retrofitting established assessment tasks to confirm a new set of learning outcomes (the CTLOs) many evaluators searched for any relevance of the task to each CTLO, whether student capability was assessable or not. Consequently, task evaluations resulted in claims that a task 'ticked the box' for assessing a particular CTLO simply because the task gave students an opportunity to engage with it, not because the task assessed student capability. This is unsatisfactory under the new paradigm where attainment of learning outcomes must be confirmed. A key factor shaping the task-evaluation procedure was the recognition that two very different interpretations of assessing a CTLO existed, and were frequently conflated, namely:

1. providing students with a task to complete which relates to the CTLO, *vs*
2. making judgement of student capability with respect to the CTLO.

Historically, assessment practices have allowed that the first may be fulfilled in the absence of the second. Tasks that serve as learning experiences will inherently provide students with the opportunity to develop skills (provide the relevant *inputs*), but may not necessarily demand any specific demonstration of CTLO attainment (require sufficient *outputs*). Only the second definition fulfils the new HESF requirement that methods of assessment must be capable of confirming CTLO attainment.

Introductory activities in the workshops included a 'warm up' discussion where participants were asked to rank various descriptors of the purpose of assessment from most to least important, and whether they agreed with each of a set of statements (listed in Figure 8). The responses indicated understanding of the multiple purposes of assessment, but favoured statements relating to definition (1) above (Figure 8). The two statements involving 'learning' and 'feedback' were overwhelmingly reported as being the two most important features of assessment. Documentation of skill development was most often ranked third, while the remaining statements about benchmarking, standards, grading or certification for professional practice all received the lowest rated importance (Figure 8). Agreement with the statements followed the same pattern in each workshop. The focus for the project, and thus the workshops, is those assessment methods that provide confirmation of student capability with respect to CTLOs; in other words, those validating certification. The evaluation activities following this survey in the workshop were designed to highlight the requirements of assessments fit for this purpose.

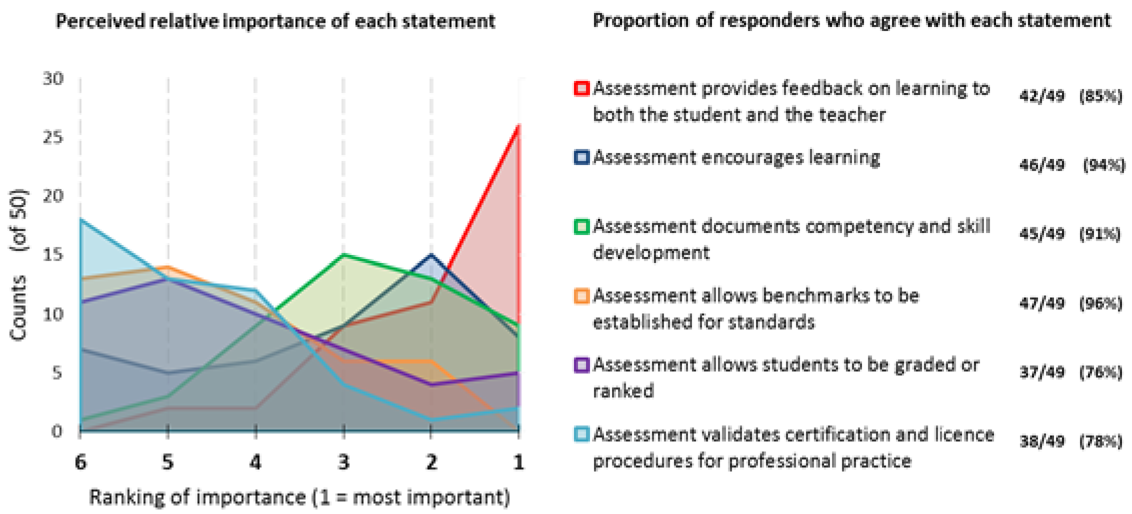


Figure 8: Perceived importance of various purposes of assessment

3.3.3 Levelling: using Bloom's taxonomy

The workshop exercise involving the revised edition of Bloom's taxonomy (Anderson, Krathwohl et al., 2001) was presented as a tool to assist in deciding if a task was pitched at a graduate level, where evidence of higher order thinking is usually sought. Workshop participants completed an introductory exercise where they were asked to classify a series of statements (Anderson, Krathwohl et al., 2001) using the taxonomy. Results for classifications of both the cognitive domain (Figure 9) and the knowledge domain (Figure 10) of the taxonomy show a diverse range of understandings for each of these common classifications. Although 'correct' answers suggested in Anderson et al. (2001) exist for all of these tasks, the range of responses found indicates that appropriate classification of scope and complexity depends on many contextual factors. The cognitive dimension of the taxonomy particularly shows a broad range of classifications received for most statements posed. The imperative for discussion to develop shared understandings of graduate level to calibrate assessors is obvious.

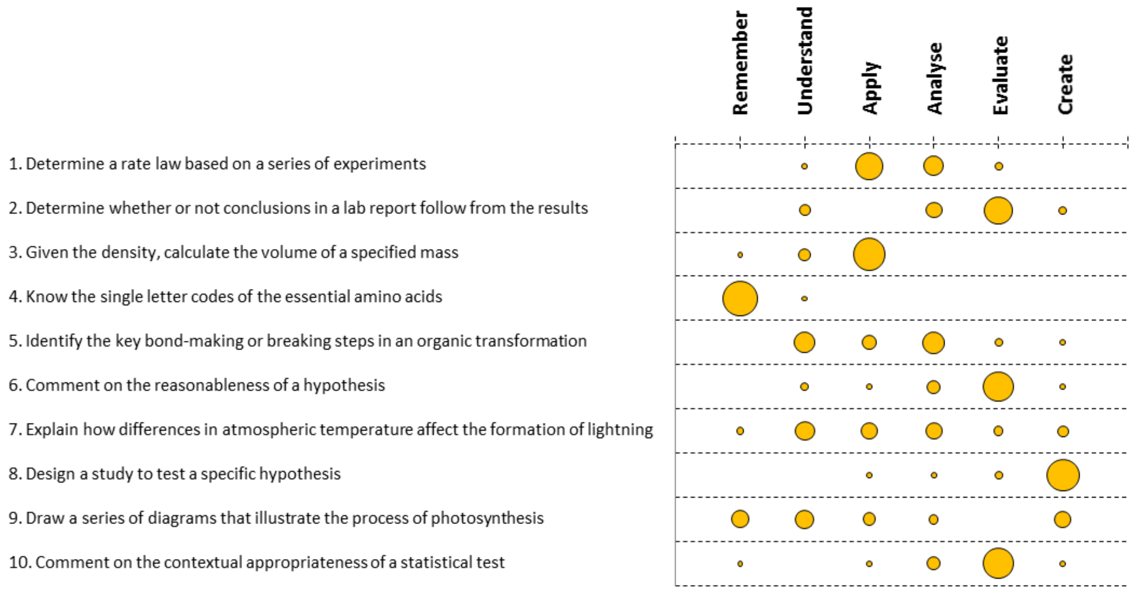


Figure 9: Classifications assigned for the cognitive dimension of the revised Bloom's taxonomy (*Circle sizes are proportional to the frequency of response. Some respondents classified a single statement into more than one category. Responses shown were gathered from 42 participants).*

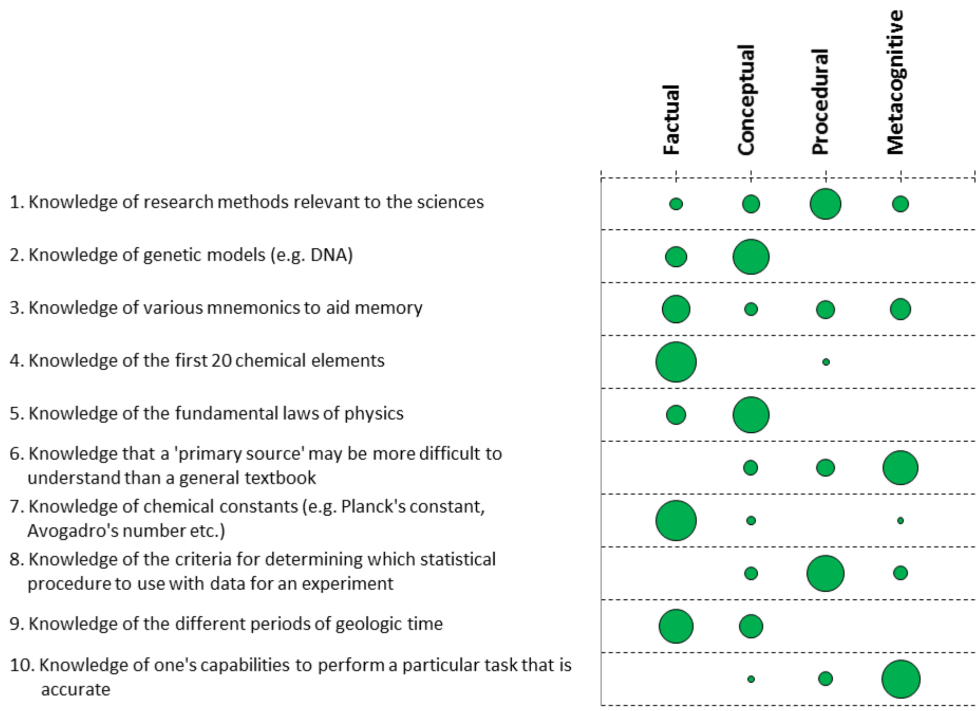


Figure 10: Classifications assigned for the knowledge dimension of the revised Bloom's taxonomy *(Circle sizes are proportional to the frequency of response. Some respondents classified a single statement into more than one category. Responses shown were gathered from 44 participants).*

3.3.5 Outcomes of workshops

The workshops were received very positively. A key reason for this was the surprising results of the task evaluation. After having applied the evaluation process, workshop participants identified that many tasks did not live up to the expectations of their creators. Surveys gathered at each workshop revealed comments such as: *'Very interesting to see how tasks that seem satisfactory on the "surface" were not when looked at using the tools'; 'It was interesting that students are often given the opportunity to gain experience in the TLOs however, they are not necessarily assessed, and therefore we do not have evidence that they meet the outcomes we aim for'; and 'Surprised by mismatch between TLO and assessment task.'*

This was a notable success for the project, and led to interest in running similar workshops in new locations (see Appendix J). Workshop participants also reflected on their own assessment practices, saying that the experience was an '*eye opener*', and that it made them '*v.* [sic] *worried about my current assessment*'. This point of success for the project, however, highlights that existing knowledge among academics about good assessment practices is relatively poor. Comments received on workshop surveys such as: '*Eye opener. I did not know this level of detail existed*'; '*I have not before thought in depth about the topic, so this has been a real eye opener';* and '*I don't think as academics we think about this sufficiently*' reveal a potential disconnect between most academics' understanding of valid assessment and the new HESF requirements. Furthermore, the academics who attended workshops are already engaged in teaching and are therefore likely more receptive to new and relevant assessment practices. The broader academic community, however, typically specialises in discipline knowledge rather than education practice, and many research-focused chemistry academics are not interested in spending time to modify traditional assessment practices.

During one of the last major workshops associated with the project, participants were presented with a list of the major components of the workshop, then asked to choose which three they felt were the most valuable (Figure 11). The most frequently valued feature, with votes from all but two participants, was the information about the structure and rationale of the task-evaluation tool itself.

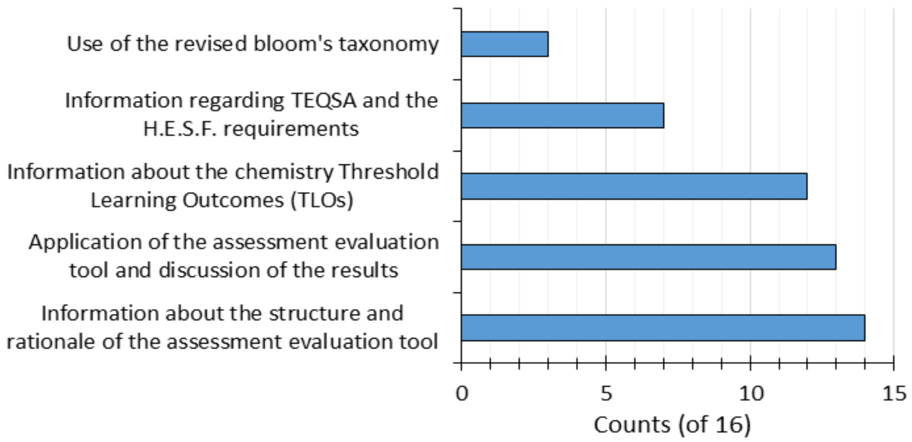


Figure 11: Perceived value of components within the workshops conducted

As a result of the response to workshops and the perceived need for spreading this changed practice through the sector, an online self-paced form of the workshop materials is being developed and will be available through the project website.

Chapter 4: Impact

4.1 Synergy of this project and the ChemNet

The *Assessing the assessments project* was proposed and funded at an important time in the development of ChemNet as a community of practice (CoP). The funding period for ChemNet had ended several years earlier and the series of meetings to discuss the details of several of the CTLOs had been concluded. Without a common goal, activity of the CoP was at an ebb. The initiation of writing the grant proposal brought renewed energy to the group of eight individuals involved in drafting the application, most of whom had previously been actively engaged in the core group of the CoP. Starting to think deeply about how to be sure that CTLOs are adequately assessed within individual assessment items restarted the conversation about the importance and value of the CTLOs from a different perspective.

Once the funding was announced, activity grew as assessment items were initially sought for testing and refinement of the evaluation tool under development. These were invited through the ChemNet email list, which remains up-to-date as the set of tertiary chemistry educators gradually changes over time. In parallel with collecting and collating these items, the series of workshops (Appendix E) lent a focus to the efforts and discussions about use and assessment of CTLOs. The principal mechanism for advertising these workshops was also the ChemNet email list.

The impact of the project achievements and mechanisms that enabled this impact can be evaluated through the application of the Impact Management Planning and Evaluation Ladder (IMPEL) model (Appendices J and K). It is clearly evident that each stage of the project has achieved increasingly broader impact (designated as 'rungs' in the model) and that these indicators confirmed a climate of readiness for change in assessment practice (Hinton, 2014).

We have observed that any face-to-face interactions have significantly more impact than virtual (whether synchronous or asynchronous) on interpersonal relationships, sense of ownership of the project and moving forward with project goals. The workshops also gave new impetus to ChemNet members who were not members of the *Assessing the assessments* project team. These members were engaged through workshops held on their university campuses, at national meetings (Appendix E) and through being invited to submit assessment items to the project. Thus, starting in late 2014 with the first large workshop held at the RACI Congress there was good awareness of the project among the whole ChemNet community, which has been sustained and grown throughout the duration of the project, particularly as dissemination activities at national meetings (Appendix E) have been conducted. In the final stages of the project, the main activities that involved ChemNet were dissemination of findings at national meetings and via email updates to the CoP. This is not an ending, however, because a significant finding is that the wording of some of the CTLOs makes their evaluation within assessment items unwieldy. Thus, the process of re-examining the wording of the CTLOs and potentially rewriting some of them again gives a boost to engagement with the CoP. This is particularly significant given that the CTLOs must be demonstrated for institutions to satisfy national regulations, and for RACI accreditation of degrees.

It is important to note that as new people join the community of tertiary chemistry educators and others resign, the population of ChemNet gradually and continually changes. The process of discussing and agreeing on the wording of the CTLOs is now seven years in the past and a new generation of lecturers are arriving on the scene without much background in how the consensus process was conducted and how the wording was agreed. Thus, an important function of the *Assessing the assessments project* has been to engage newcomers to tertiary chemistry education in Australia with the work that has gone before around the CTLOs, and to give them the opportunity to engage with and take ownership of a new round of discussions that is beginning. This is critical to their movement from the periphery towards the core of the CoP and to the ongoing existence of ChemNet.

The website for this project exists within the umbrella ChemNet website (www.chemnet.edu.au) and so as users search for exemplars and learn about the outcomes of this project, they will also be reminded of and likely revisit the existing ChemNet website.

Thus, the *Assessing the assessments project* has given renewed life to ChemNet through many face-to-face activities and engagement with an important project. Conversely, ChemNet formed an important foundation for the *Assessing the assessments project* to communicate with the CoP of tertiary chemistry educators in Australia, not only to obtain submitted assessment items but to also provide an audience for feedback within workshops and in the dissemination stage. As the project draws to an end, activity within ChemNet will move to ensuring that the tool is widely adopted within our CoP and becomes the standard for evaluating the fitness for purpose of assessment items to allow students to demonstrate CTLOs. ChemNet is also tasked with coordinating discussions and consensus activities around reframing the wording of the CTLOs.

4.2 Chemistry accreditation: RACI

The RACI is the professional and accrediting body for chemical scientists in Australia. The RACI accredits bachelor-level courses across Australia to ensure that graduate chemists have the necessary skills and knowledge to practice and, after a requisite period of relevant work, become a Chartered Chemist. Until very recently, however, the accreditation of a course was driven by inputs (laboratory hours principally) rather than any direct assessment of knowledge and skills.

In parallel with changes to the higher education landscape, the RACI has embarked on a process to realign its accreditation process to modern curricular practice by focusing on outcomes and assessment of student learning. The RACI was a sponsor for the project that derived the CTLOs from the broader science TLOs. The institute wanted a consensus view of the Australian chemical sciences community, and then sought to harness this consensus to reshape its accreditation processes.

4.2.1 Approach

To evidence the attainment of the CTLOs the notion of curriculum described by Keeves (1972), and extended by Friedel and Treagust (2005), was adapted. The accreditation process has four stages of curriculum review:

- *intended*: CTLOs need to be clearly expressed and assessed throughout the curriculum
- *implemented*: each institution interprets each of the CTLOs, and reports their self-assessment of student attainment in each unit of study, which is then aggregated in a curriculum map
- *perceived*: an accreditation panel validates the alignment between the intended and implemented curriculum before recommending the award of accredited status for a given program or programs
- *achieved*: this accreditation panel summarises the student achievement through the lens of the CTLOs and reports to the RACI Board, who endorses the accreditation.

4.2.2 Outcomes

The pilot and first year of the new process saw 11 undergraduate degree courses at seven institutions around Australia nominated for accreditation. This represented a reasonable cross-section of metropolitan, regional and remote campuses admitting students from a range of backgrounds and abilities, and offering instruction to them in a number of modes. These degrees had not been designed with the CTLOs in mind, thus without explicit intent this presents a unique a posteriori perspective on chemistry curriculum to test whether the outcomes and/or curriculum are fit for purpose.

At the implemented phase of accreditation, the teaching staff responsible for each unit of study in the degree complete a survey about their curriculum for the purposes of mapping. This survey collects data twice about outcomes; first, whether each CTLO is taught, practiced or assessed in the unit; and then whether the CTLO is explicitly evaluated within each type of assessment task (Sumsion & Goodfellow, 2004). These data for the unit level and then explicit evaluation in assessment tasks are aggregated in Figure 12.

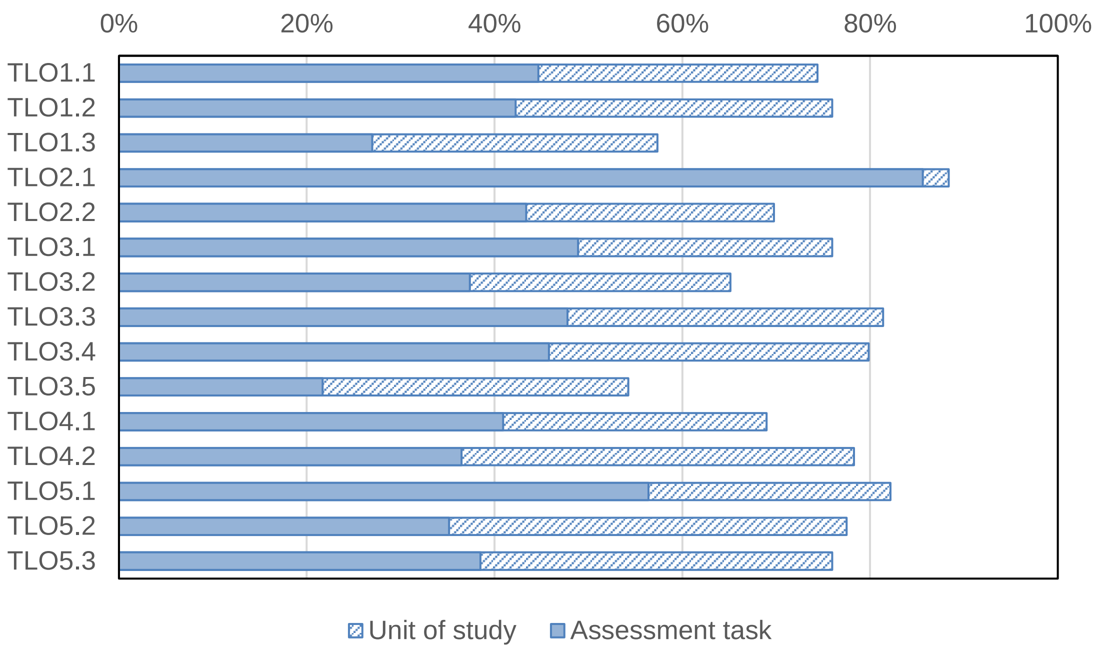


Figure 12: Distribution of coverage of CTLOs by units of study (*n* = 129) and assessment tasks (*n* = 452) self-reported by institutions seeking accreditation (*n* = 7) in the first year of the accreditation

Figure 12 illustrates that staff recognise and acknowledge there are opportunities for assessment of the CTLOs, but only between one- and two-thirds of tasks in each unit explicitly evaluate and presumably evidence attainment of outcomes other than content knowledge (CTLO2.1). This provides insight into the capacity of any chemistry curriculum to evidence attainment of the CTLOs, and thus this gap suggests that curriculum design will more probably adapt to the outcomes rather than *vice versa*. The key emergent need for accreditation is a process for the evidencing of attainment of outcomes through assessment that allows a shared understanding, and to share best practice in assessment of outcomes.

The aggregation of individual assessment tasks puts pressure on these tasks to be fit for purpose and the mechanism to demonstrate these outcomes should withstand robust external scrutiny. Therefore, the objectives of this project are central to building an evidence base for the validity of the RACI's accreditation processes and to lead the development of chemistry curriculum design to be well-aligned with these threshold learning outcomes.

Chapter 5: Conclusions

We have developed a sophisticated tool for determining the ability of assessment items to demonstrate attainment of stated learning outcomes. This tool can be applied to any assessment item and any set of learning outcomes, although is most practical for learning outcomes at an intermediate breadth. For the purposes of this project, the Australian CTLOs were evaluated at the second-tier level (i.e. CTLOs 1.1, 1.2, 1.3 etc., not CTLOs 1, 2, 3, etc.). The top tier was trialled in the initial stages of the evaluation process, but was considered too coarse for useful feedback on the assessment items, while using the third tier (CTLOs 2.1.1, 2.1.2 etc.) would have required significantly more time and was not expected to lead to more useful information.

Using this tool, the combination of a set of assessment items in a unit of study or degree program can be easily evaluated to determine whether all desired learning outcomes are required to be demonstrated by students. It is clear from evaluation of the 45 assessment items so far submitted that not all CTLOs are assessed equally well, and some do not seem to be assessed at all. Thus, use of the tool can inform design and modification of assessment tasks to ensure that students are given the opportunity to achieve all required learning outcomes during their degree programs. Moreover, application of the tool requires deep reflection on the set of learning outcomes in use and may lead to revision or modification of the wording of these to allow their practical application.

In addition to the tool and website as the major outputs of the project, significant outcomes are an increased readiness of the chemistry community at large for the changed assessment paradigm, a requirement of which is implicit in the HESF and renewed engagement with the Chemistry Discipline Network as a Community of Practice. Professional development workshops run as part of this project have helped to prepare agents of change, but it must be recognised that the workload associated with wholesale changing of assessment regimes is enormous.

References

American Chemical Society Committee on Professional Training *Undergraduate professional education in chemistry: ACS guidelines and evaluation procedures for Bachelor's degree programs*, American Chemical Society, Washington, DC, 2015.

Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., Wittrock, M.C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition). New York: Longman.

Barrie, S. C. A research-based approach to generic graduate attributes policy. *High. Ed. Res. & Dev.* **2004**, *23*, 261-275. doi: 10.1080/0729436042000235391

Barrie, S. C. A conceptual framework for the teaching and learning of generic graduate attributes. *Stud. High. Educ.* **2007**, *32*, 439-458. doi: 10.1080/03075070701476100

Biggs, J. Enhancing teaching through constructive alignment. *High. Educ.* **1996**, *32*, 347-364. doi: 10.1007/BF00138871

Biggs, J. What the student does: Teaching for enhanced learning. *High. Ed. Res. & Dev.* **1999**, *18*, 57-75. doi: 10.1080/0729436990180105

Booth, S.; Beckett, J.; Saunders, C. Peer review of assessment network: Supporting comparability of standards. *Qual. Assur. Ed.* **2016**, *24*, 194-210. doi: 10.1108/QAE-01-2015-0003

Boud, D.; Falchikov, N. Aligning assessment with long-term learning. *Assess. Eval. High. Ed.* **2006**, *31*, 399-413. doi: 10.1080/02602930600679050

Boud, D., and Associates. *Assessment 2020: Seven propositions for assessment reform in higher education*, Australian Learning and Teaching Council: Sydney, 2010. http://www.olt.gov.au/system/files/resources/Assessment%202020_final.pdf (accessed July 28, 2017).

Bowman, N. A. Understanding and addressing the challenges of assessing college student growth in student affairs. *Res. Pract. Assess.* **2013**, *8*, 5-14.

Bryce, T. G. K.; Robertson, I. J. What can they do? A review of practical assessment in science. *Stud. Sci. Ed.* **1985**, *12*, 1-24. doi: 10.1080/03057268508559921

Buntine, M.; Price, W.; Separovic, F.; Brown, T.; Thwaites, R. *Learning and teaching academic standards: Chemistry Academic Standards Statement*, Australian Learning and Teaching Council, 2011. http://www.olt.gov.au/system/files/altc_standards_SCIENCE_240811_v3.pdf (Appendix 3) (accessed July 28, 2017).

De la Harpe, B.; David, C. Major influences on the teaching and assessment of graduate attributes. *High. Ed. Res. & Dev.* **2012**, *31*, 493-510. doi: 10.1080/07294360.2011.629361

Deane, E.; Krause, K.-L. Towards a learning standards framework. Learning and teaching standards (LaTS) project: peer review and moderation in the disciplines. (2013).

http://www.westernsydney.edu.au/__data/assets/pdf_file/0010/398620/Learning_Stds_Framewk_Final_Dec_2012.pdf (accessed July 28, 2017).

Ewan, C. Disciplines setting standards: The learning and teaching academic standards (LTAS) project. Paper presented at Australian Quality Forum, Gold Coast, 2010.

Friedel J.M., Treagust D.F. Learning bioscience in nursing education: perceptions of the intended and the prescribed curriculum. Learning in Health & Social Care. 2005;4(4):203–16.

Gibbs, G.; Simpson, C. Conditions under which assessment supports students' learning. *Learn. Teach. High. Ed.* **2004**, *2004-05*, 3-31.

Gibbs, G.; Dunbar-Goddet, H. Characterising programme-level assessment environments that support learning. *Assess. Eval. High. Ed.* **2009**, *34*, 481-489. doi: 10.1080/02602930802071114

Green, W.; Hammer, S.; Star, C. Facing up to the challenge: Why is it so hard to develop graduate attributes? *High. Ed. Res. & Dev.* **2009**, 28, 17 - 29.

Hager, P. Nature and development of generic attributes. In *Graduate attributes, learning and employability*; Hager, P., Holland, S., Eds.; Springer: Dordrecht, Netherlands, 2006.

Hager, P.; Holland, S.; Beckett, D. *Enhancing the learning and employability of graduates: The role of generic skills*, Business/Higher Education Round Table (Australia) (B-HERT), Melbourne, Australia, 2002.

Higher Education Standards Framework (Threshold Standards) paragraph 1.4. Australian Government: Canberra, 2015. *F2015L01639* Made under subsection 58(1) of the Tertiary Education Quality and Standards Agency Act 2011. https://www.legislation.gov.au/Details/F2015L01639 (accessed July 28, 2017).

Hinton, T. (2014). https://docs.education.gov.au/documents/impact-management-planning-and-evaluation-ladder-impel (accessed July 28, 2017).

Hughes, C.; Barrie, S. Influences on the assessment of graduate attributes in higher education. *Assess. Eval. High. Ed.* **2010**, *35*, 325-334. doi: 10.1080/02602930903221485

Jackson, D. Employability skill development in work-integrated learning: Barriers and best practice. *Stud. High. Educ.* **2015**, *40*, 350-367. doi: 10.1080/03075079.2013.842221

Jones, S. M.; Yates, B. F.; Kelder, J.-A. *Science Learning and Teaching Academic Standards Statement*, Australian Learning and Teaching Council, 2011. http://www.olt.gov.au/system/files/altc_standards_SCIENCE_240811_v3.pdf (accessed July 28, 2017).

Keeves, J. P. The IEA science project: science achievement in three countries—Australia, the Federal Republic of Germany, and the United States. University of Kiel, Kiel: German Commission for UNESCO; 1972.

Kelder, J.-A.; Jones, S. M. *The Science Learning and Teaching Academic Standards project: A discipline community's response to regulatory change in Australian higher education* Paper presented at Higher Education Research and Development Society of Australasia, Melbourne, 2015.

Kemp, I. J.; Seagraves, L. Transferable skills—can higher education deliver? *Stud. High. Educ.* **1995**, *20*, 315-328. doi: 10.1080/03075079512331381585

Krathwohl, D. R. A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice,* **2002***,* *41*, 212-218.

Krause, K.-L.; Scott, G.; Aubin, K.; Alexander, H.; Angelo, T.; Campbell, S.; Carroll, M.; Deane, E.; Nulty, D. D.; Pattison, P.; Probert, B.; Sachs, J.; Solomonides, I.; Vaughan, S. *Assuring learning and teaching standards through inter-institutional peer review and moderation*, Office for Learning and Teaching, 2014. www.olt.gov.au/system/files/resources/SP10_1843_Krause_report_2014.pdf (accessed July 28, 2017).

Macquarie University *Learning and teaching - Sustainability - Macquarie University*. http://www.mq.edu.au/about_us/strategy_and_initiatives/sustainability/ (accessed July 28, 2017).

Meyers, N. M.; Nulty, D. D. How to use (five) curriculum design principles to align authentic learning environments, assessment, students’ approaches to thinking and learning outcomes. *Assess. Eval. High. Ed.* **2009**, *34*, 565-577. doi: 10.1080/02602930802226502

Nicol, D. Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assess. Eval. High. Ed.* **2009**, *34*, 334-352. doi: 10.1080/02602930802255139

O’Connell, B.; De Lange, P.; Freeman, M.; Hancock, P.; Abraham, A.; Howieson, B.; Watty, K. Does calibration reduce variability in the assessment of accounting learning outcomes? *Assess. Eval. High. Ed.* **2016**, *41*, 331-349. doi: 10.1080/02602938.2015.1008398

Pinto, G. The Bologna process and its impact on university-level chemical education in Europe. *J. Chem. Educ.* **2010**, *87*, 1176-1182. doi: 10.1021/ed1004257

Pyke, S.; O’Brien, G.; Yates, B.; Buntine, M. *Chemistry Academic Standards Statement - revised*, 2014. http://chemnet.edu.au/sites/default/files/u39/CHEMISTRY_Academic_Standards_Accreditation_Trial.pdf (accessed July 28, 2017).

Rice, J. *Good practice report: Assessment of science, technology, engineering and mathematics (STEM) students*, Australian Learning and Teaching Council, 2011. http://www.olt.gov.au/system/files/resources/GPR_Assessment_STEM_Students_Rice_2011.pdf (accessed July 28, 2017).

Sadler, D. R. Three in-course assessment reforms to improve higher education learning outcomes. *Assess. Eval. High. Ed.* **2015**, 1-19. doi: 10.1080/02602938.2015.1064858

Sadler, D. R. Formative assessment and the design of instructional systems. *Instr. Sci.* **1989**, *18*, 119-144. doi: 10.1007/BF00117714

Schultz, M.; Mitchell Crow, J.; O’Brien, G. Outcomes of the chemistry discipline network mapping exercises: Are the threshold learning outcomes met? *Int. J. Innov. Sc. Math. Ed.* **2013**, *21*, 81-91.

Schultz, M.; O’Brien, G. The Australian Chemistry Discipline Network - a supportive community of practice in a hard science. In *Implementing Communities of Practice in higher education - dreamers and schemers*; McDonald, J., Cater-Steel, A., Eds.; Springer: Singapore, Singapore, 2017, pp 501-530.

Scouller, K. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *High. Educ.* **1998**, *35*, 453-472. doi: 10.1023/A:1003196224280

Sumsion J, Goodfellow J. Identifying generic skills through curriculum mapping: a critical evaluation. Higher Education Research & Development. **2004**, 23(3):329–46.

The University of Sydney Institute for Teaching & Learning *The Sydney Graduate*. http://www.itl.usyd.edu.au/graduateAttributes/ (accessed July 28, 2017).

van den Akker, J. The science curriculum: Between ideals and outcomes. In *International handbook of science education*; Fraser, B. J., Tobin, K. G., Eds.; Kluwer: Dordrecht, Netherlands, 1998, pp 421-448.

Appendix A: Certification by Deputy Vice-Chancellor (or equivalent)

I certify that all parts of the final report for this OLT grant/fellowship (remove as appropriate) provide an accurate representation of the implementation, impact and findings of the project, and that the report is of publishable quality.

Name: Professor Pip Pattison

Date: 15 September 2017

Professor Pip Pattison AO
Deputy Vice-Chancellor (Educatior

Appendix B: External Evaluator's Report

ID14-3652: Assessing the Assessments:

Evidencing and Benchmarking Student Learning Outcomes in Chemistry

The need for a focus on assessment in chemistry

This project, funded by the Office for Learning and Teaching (OLT), addresses a fundamental issue at the centre of challenges to, not only chemistry, but also to the whole raft of disciplines in science, technology, engineering, and mathematics (STEM) education in universities worldwide. The question, in the context of chemistry, is: How can tertiary chemical educators ensure the community that graduate chemists have the requisite knowledge, skills and values that university chemistry departments claim are covered in chemistry major programs, and are articulated in defined program learning outcomes? Put in common parlance: Is chemistry, as a discipline in higher education, delivering on the promises made about the value that students will gain by studying chemistry, and about the benefits that society can gain from taxpayer investments in chemistry departments? Does the rhetoric match the reality?

These are serious questions, made even more so by decreasing government funding to higher education through what appears to be an increasing user-pays government policy for higher education.

So, what evidence can be produced and will be convincing to reassure community and government stakeholders about the quality of chemistry graduates? Clearly, the answer lies in having robust assessment strategies that are aligned with stated learning outcomes, and that require students to demonstrate that they have attained those learning outcomes. Of course, as we all know, this is far from easy. The Assessing the Assessments project is an amazingly timely project. It has been well-designed to produce clear evidence about the level of assessment being conducted in Australian chemistry departments. The project team has tackled this complex and multi-faceted task in a logical and progressive fashion, and made some significant advances to validating and enhancing assessment in Australian chemistry departments.

Overall success of the Assessing the Assessments project

In the project application, the four main anticipated outcomes of this project were stated as:

1. "A framework for objective determination as to whether a particular CTLO has been met by a specific assessment item that can be used by all staff teaching chemistry;
2. A catalogue of validated, reliable exemplar assessment items for demonstrating evidence of achievement of each of the CTLOs;
3. Informed contribution from the chemistry community to STLOs (1, 3, 4, and 5); and

4. Strong links to similar activities in chemistry teaching communities overseas for benchmarking assessment of learning outcomes in order to enhance the international standing of Australian programs."

I can confidently say that all four outcomes have been achieved, and this can be clearly seen in the main report. There is (matching the outcomes stated above):

1. A robust and workable task-evaluation tool for chemistry assessment tasks;
2. A catalogue of 45 chemistry assessment tasks that have been exhaustively analysed;
3. A number of successful workshops and a strategy for continuing interactions within the Chemistry Discipline Network (ChemNet), and with the Royal Australian Chemical Institute (RACI); and
4. Engagement with the international community through publications, conference presentations and other academic linkages that members of the highly skilled project team have formed.

There have been challenges – and I will comment in more details on at least some of them later in this report – but the careful design of the various iterations of the task-evaluation tool for the assessment tasks collected by the project team, combined with an ongoing reflective process guided by an evaluation plan, ensured valuable project outcomes.

The project was carried out during a difficult time in higher education in Australia. Funding cuts have been mentioned above. In addition, the demise of the OLT was a blow to academics who were working hard on a project, not only for the nature of the project itself, but also in the hope that future endeavours would be funded. In all projects, initial enthusiasms can easily be eroded under the sheer weight of local responsibilities. In OLT projects commitment to the profession is a strong driver as adequate time-release is a pipedream; so, with the ongoing stream of OLT funding and recognition disappearing, the particular pressures and competing priorities that each team member has loom larger. In my view, this made it difficult for everyone to commit easily to project activities, maintain a detailed level of understanding of what was happening and what decisions needed to be made.

I have commented that the project team was comprised of highly skilled and experienced chemistry educators. However, the team was quite large (ten persons), and some moved to different roles/jobs during the project, resulting in further pressures on the project timeline.

Given these overarching factors, the project team can be congratulated on the success of the project.

The evaluation plan

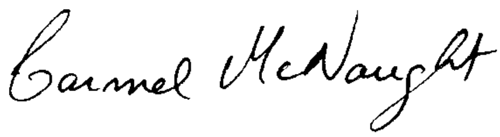
I have been involved in five OLT projects as External Evaluator. In the other four projects, my report included an annotated copy of the evaluation plan. However, in the Assessing the Assessments project, the design of the project, rightly, was an iterative process of collecting assessments, working with chemistry academics to analyse the levels of design and marking schemes, refining the task-evaluation tool, followed by further cycles of collection, analysis, development and reflection. So, the design of the evaluation was one where each event (workshop, meeting, conference, etc.) and phase in the project was the focus of evaluation.

For each event and phase, reflection took place over a number of aspects which were: Process, Outcome and Outputs, Learnings, Investment, Relevance, Efficiency, Effectiveness, Sustainability, and What next. Rather than comment on the minutiae of the records kept excellently by Samuel Priest, the Project Officer (for which he deserves commendation), I will use the evaluation aspect headings to comment on the project as a whole.

Focus of evaluation	Comments on the project as a whole
Process	This was not a strong feature of the project. The project had some periods of less-than-optimal activity. Some of the reasons for this have been described above, and alarm bells sound with me when external pressures appear to grind processes to a halt. In OLT projects, leaders are usually academics who have significant administrative/ leadership roles in their own institutions. If funding resumes for projects in higher education, time and space for project leaders needs to be assured if project processes are to be optimal. From my point of view, the project had quite poor communication strategies, though I acknowledge that within-team dynamics may well have been much better, and indeed seem to have been so. I sent many emails that were not adequately responded to and I felt that my expertise was under-utilized. I attended one workshop and had several detailed Skype conversations with the Project Leader. However, these were isolated events and I was not interacted with for months at a time and, then, usually after my outreach. There was some initial use of the project-management software Basecamp but this lapsed and some project documents could later be found on a Google Drive site. However, it was hard to get a sense of the ongoing processes. I have commented above on the wonderfully detailed notes that were kept on the evaluation plan. However, I was not kept in the loop; I had a version dated 15 April 2015 and, despite numerous requests, only received the next (and final) version in April 2017. I first received a draft final report on 22 August 2017 and so could then provide detailed feedback which I have done. My report was completed on 25 August 2017, and hence is not based on the final version of the final report.
Outcome and Outputs	As noted above, I am satisfied that the project outcomes have been achieved. However, one aspect deserves strong comment. Despite

	rigorous searching by the project team, only 45 assessment tasks exist in the project database. These come from 22 Australian universities. Only 45 assessment tasks! What is the rest of the assessment like? What does the project tell us about the level of assessment in chemistry in Australia? The data in the main report clearly indicate that the Chemistry Threshold Learning Outcomes (CTLOs) are not assessed adequately and that students can get a chemistry degree without demonstrating adequate attainment of CLTOs. This is a very important piece of evidence that should be considered and acted upon in the highest levels of government and senior university management. I am pleased that there have been several dissemination activities, including publications and conference presentations. The project team is exhorted to map out a plan for more dissemination activity. There is much that needs to be said, strongly and widely.
Learnings	A strong feature of the project has been its iterative design and the way in which the project team has used feedback data – workshop activities, surveys, interviews, reflective discussions, etc. – to refine the task-evaluation tool.
Investment	One reason why assessment in higher education is in such a parlous state is that academics are now time-poor and good assessment is time-consuming and, in that sense, expensive. This project has clearly evidenced an emerging crisis in chemistry assessment. I believe that the sound evidence that the project has garnered is extremely valuable and worth the investment provided by the OLT. Of course, the government and individual universities have to respond to this evidence and make a plan to enhance students' assessment experiences. As noted above, the data from this project are of value to disciplines apart from chemistry.
Relevance	Absolutely! This is a timely and important input to discussions about assessment in chemistry, in science and, more generally, in higher education.
Efficiency	My comments about the project's communication strategies as they pertain to me indicate some inefficiencies but, as noted earlier, within-team communications appear to have been better. I have heard nothing about the Reference Group and am not sure what role these skilled and experienced academics played.

Effectiveness	In my opinion, the project team has produced a report that effectively describes, with evidence, a serious problem in chemistry assessment, and also provides an effective task-evaluation tool and a strategy for moving forward in concert with ChemNet and the RACI. Evidence + Strategy = Effectiveness.
Sustainability	This gets a heads-up, mostly because of the strong links with ChemNet and the RACI. The combination of a strong professional network and an esteemed professional organization is a powerful force; and government should recognize and support these existing positive aspects of tertiary chemistry education. Both ChemNet and the RACI have been integral to the outcomes of the Assessing the Assessments project, and offer an avenue for sustainability of the work.
What next?	This is an open question! I am delighted that the project has the potential to be sustainable and that the excellent work done in this project should continue to grow and inform Australian and international best practice in assessment in chemistry, in science more generally, and indeed in many other discipline areas. However, it depends on whether those in senior government posts and those in the senior management of universities consider and act upon this report. We will see …



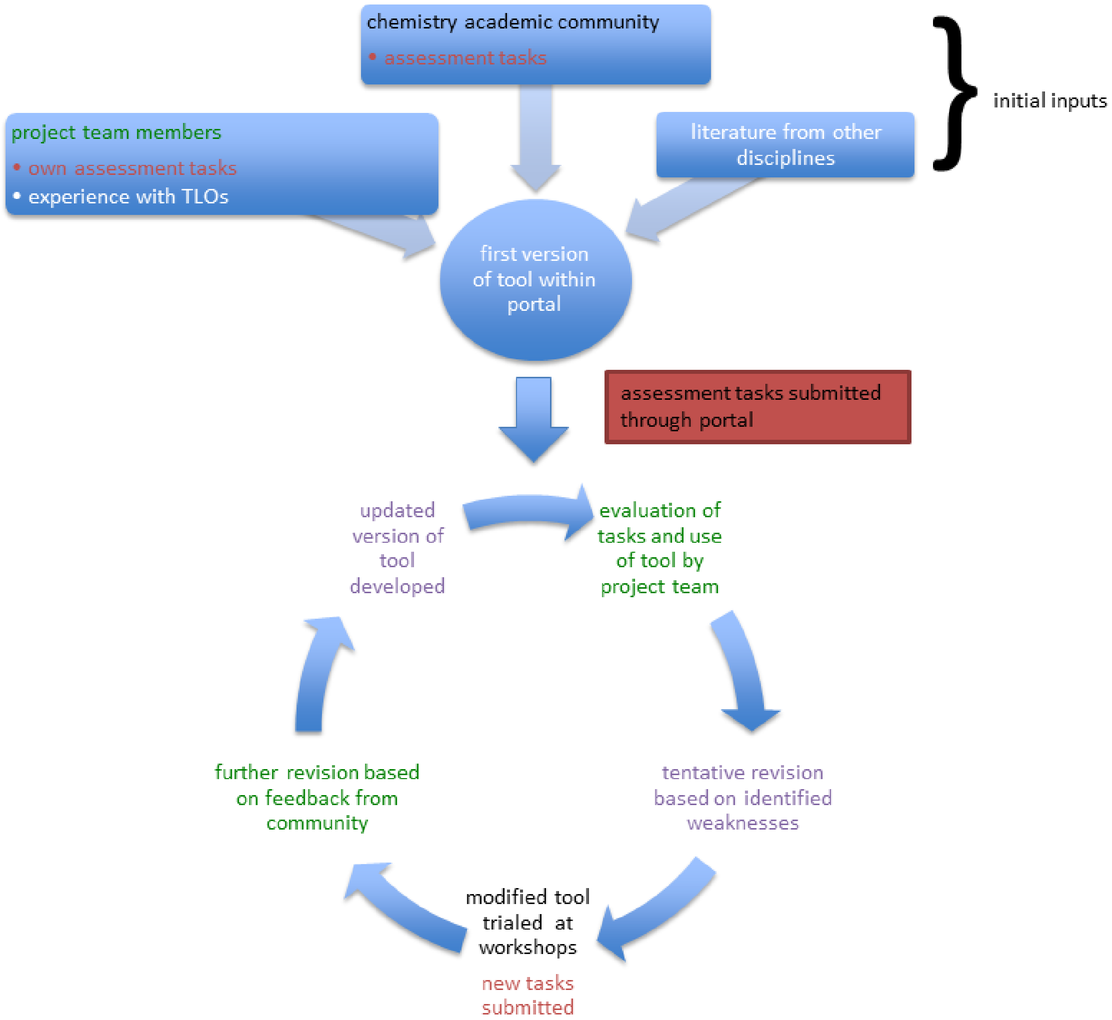
Carmel McNaught, PhD, FAACE

Emeritus Professor of Learning Enhancement, The Chinese University of Hong Kong

Distinguished Visiting Professor, University of Johannesburg

25 August 2017

Appendix C: Flowchart of action research sequence



Appendix D: Questions in the original online task submission portal

1. **In which part of the degree program is the assessment item used and how does it fit with assessment items earlier and later in the degree?**

   1st year
   2nd year
   3rd year
   FIT:_____________________________

2. **Do all students that major in Chemistry complete this assessment item?**

   YES
   NO

3. **Which types of knowledge (Krathwohl, 2002) are required to successfully complete the assessment item?**

   Factual
   Conceptual
   Procedural
   Metacognitive

4. **What cognitive processes (Bloom's taxonomy (Krathwohl, 2002)) are required to successfully complete the assessment item?**

   Remember
   Understand
   Apply
   Analyze
   Evaluate
   Create

5. **Which Chemistry TLOs does the submitter think are demonstrated by successful completion of the assessment item?**

   Understand ways of scientific thinking by:
   - recognising the creative endeavour involved in acquiring knowledge, and the testable and contestable nature of the principles of chemistry
   - recognising that chemistry plays an essential role in society and underpins many industrial, technological and medical advances
   - understanding and being able to articulate aspects of the place and importance of chemistry in the local and global community

   Exhibit depth and breadth of chemistry knowledge by:
   - demonstrating a knowledge of, and applying the principles and concepts of chemistry
   - recognising that chemistry is a broad discipline that impacts on, and is influenced by, other scientific fields

   Investigate and solve qualitative and quantitative problems in the chemical sciences by:
   - synthesising and evaluating information from a range of sources, including traditional and emerging information technologies and methods
   - formulating hypotheses, proposals and predictions and designing and undertaking experiments in a safe and responsible manner

- applying recognised methods and appropriate practical techniques and tools, and being able to adapt these techniques when necessary
- collecting, recording and interpreting data and incorporating qualitative and quantitative evidence into scientifically defensible arguments
- demonstrating the cooperativity and effectiveness of working in a team environment

Communicate chemical knowledge by:

- presenting information, articulating arguments and conclusions, in a variety of modes, to diverse audiences, and for a range of purposes
- appropriately documenting the essential details of procedures undertaken, key observations, results and conclusions

Take personal, professional and social responsibility by:

- demonstrating a capacity for self-directed learning
- demonstrating a capacity for working responsibly and safely
- recognising the relevant and required ethical conduct and behaviour within which chemistry is practised

6. Which of the discipline knowledge categories listed below does this assessment item contribute to?

- Stoichiometry, structure and characteristic properties of chemical substances
- Methods of structure determination
- Properties of matter in relation to structure
- Chemical thermodynamics, equilibrium and kinetics
- Reaction processes can transform substances into very different products
- Reactions of metal and non-metal compounds including carbon compounds
- Quantifying concentrations and amounts of elements and compounds in simple and complex mixtures
- Other (please specify)

7. Which, if any, of the practical (laboratory) skills listed below are required to successfully complete the assessment item?

Generic laboratory skills

Synthesis skills

Qualitative and quantitative analysis skills

Modelling and/or simulation skills

8. Are communication skills assessed in this assessment item?

YES; Explicitly

YES; Implicitly

NO; Not assessed

9. Is achievement of any other Chemistry TLO potentially demonstrated by the assessment item?

10. How many markers assess this item? If there is more than one, what moderation processes (if any) are used to ensure that marking criteria are applied fairly and consistently?

11. What steps are taken to ensure that the assessment item is both valid & reliable?

Valid assessment will measure student learning (not something else) and the extent of that learning.
Reliable assessment will yield consistent results regardless of whom it is marked by or when it is marked.

Appendix E: Listing of workshops and team meetings

Event	Details
Adelaide workshop (6/12/2014)	Public project launch; gathering feedback on initial project design. **25 participants**.
Melbourne workshop (6/6/2015)	Initial tool – yes/no to broad TLOs. Two parts: task design, student work. Key workshop format delivered (persists for future workshops): intro about assessment+ learning outcomes (data gathered), revised bloom's taxonomy (data gathered), info about TLOs, upload of tasks to portal (tasks gathered), evaluation of provided tasks, evaluation of own tasks. **14 participants.**
Brisbane workshop (15/7/2015)	Revised tool – yes/no/partial, broken into parts: task design, student work, assessment. Same workshop format. Low attendance. **4 participants**.
Project team task evaluations **(Round 1)**	Results illustrate nuance needed: "partial" ratings given for multiple reasons. Two key reasons identified: depth of TLO coverage (grad level or not?) and breadth of coverage (part of the TLO statement or all of it?).
Curtin workshop (29/9/2015)	New tool – "four-square" model. Three parts: task design, student work, assessment. Same workshop format, plus some introduction to four-square model. High attendance, mostly from Curtin. **10 participants.**
Curtin team meeting	Discussion of issues regarding reaching consensus in task evaluations. Issues regarding "what is graduate level" recognized to be necessary for task evaluation, but inappropriate for project team to define this. Issues regarding aggregating multiple assessment tasks at course/ program level highlighted.
Project team task evaluations **(Round 2)**	Results illustrate lack of consensus. Evaluation method needs revision to promote more consistency in response and reduce the laborious procedure. A hierarchy of "assessment ratings" devised following this, to better categorise outcomes. "decision tree" devised for the four-square format to promote more consistency in response.
Sydney workshop (4/11/2015)	New tool – "assessment-rating + four-square" model. Same workshop format, plus introduction to assessment rating/ four square model. High attendance. **14 participants.**
Sydney team meeting (5/11/2015)	Discussion of new tool structure and efficacy. Evaluation of tasks by project team to check level of consensus and utility of tool outcomes.
Sydney team meeting (18-19/7/2016)	Intended to discuss exemplars. Lack of consensus in task evaluations prevented this – extensive discussion about tool design/ question phrasing. Alternative versions of the

	tool developed for testing at ACSME.
ACSME Brisbane (28-29/8/2016)	Presentation about the project (Schmid). Presentation about related content (Southam).
ACSME discipline day workshop (30/9/2016)	Trialing different versions of the tool devised during previous team meeting, now in spreadsheet format (revisions include question phrasing and order). Interviews recorded and survey responses gathered, both concerning the tool. Different workshop format due to time restrictions: no bloom's or learning outcomes/ assessment intro. No tool introduction (purpose is to gauge reception and accessibility). Original portal now not used – new web page and upload system exists. Initial portal questions no longer used. **9 participants interviewed for feedback.** Greater numbers attended presentation, workshop survey feedback from **12 participants.**
Small meeting in Wollongong (16-17/9/2016)	Aspects about exemplar items – generating items and developing existing items.
Sydney team meeting (12-13/12/2016)	Clarification of "framework" underpinning the tool process. Discussion of identifying exemplar items. Possible aspects of web presentation of the tool and presentation of exemplar items. Discussions about papers being written (workshops paper, "soft skills" paper, tool/framework paper). Discussing content for final report. Data analysed for papers. Feedback from survey online.
Adelaide workshop (15/2/2017)	Running final version of tool (with success), gathering data about workshops for new project output: online workshop module. **17 participants.**
Adelaide team meeting (16/2/2017)	Decision that issues of disagreement are no longer about the tool structure – now about TLO phrasing/ interpretation. Discussion about TLOs 1, 4 and aspects of 3 and 5 in terms of clarifying intent and possible rephrasing (TLO elucidation was a proposed output). Further discussion about final reporting.
CSU workshop (10/4/2017)	Workshop run at CSU in Wagga Wagga. Final version of the tool. Main purpose for dissemination of information and resources output by the project. **10 participants.**
Project team task evaluations **(Round 3)**	Evaluation of all **45 tasks submitted to the project** completed using the final framework/tool for valid comparisons and outcomes reporting. Identification of limited number of clear exemplars of good assessment, standardization into an exemplars "template" for website presentation.

Appendix F: Workshop timetable

10:00-10:10	Opening
10:10-10:15	Introduction
10:15-11:30	**Session #1 – The L&T landscape** The regulatory framework for Learning & Teaching Assessment, achievement and outcomes
11:30-12:00	**Session #2 – Uploading assessment tasks** Participants upload assessment items to the submission portal
12:00-12:30	LUNCH
12:30-14:00	**Session #3 – 'Calibration' Session A** [using previously contributed items] Two stage process: (1) look at the task alone – does it allow the students to meet the specified outcome; (2) look at the student work – does the student work demonstrate the stated outcome.
14:00-14:15	AFTERNOON TEA
14:15-15:45	**Session #4 – 'Calibration' Session B** [using participant submitted items] Two stage process: (1) look at the task alone – does it allow the students to meet the specified outcome; (2) look at the student work – does the student work demonstrate the stated outcome.
15:45-16:00	Wrap-up, feed-forward & close.
16:00-17:00	Team debrief

Appendix G: Introductory workshop exercises

Which of the following statements best describe 'learning outcomes'?

		Best?
1.	Learning outcomes are usually expressed as knowledge, skills, or attitudes.	
2.	Learning outcomes are statements which describe a desired condition	
3.	Learning outcomes are statements that specify what learners will know or be able to do as a result of a learning activity.	
4.	Learning outcomes must be observable, measurable and be demonstrated by the learner	
5.	Learning outcomes should be assessable to determine the gap between an existing and a desired condition.	

Which of the following statements do you agree with? Which is the most important?

		Agree? (Yes or No)	**Importance? (Rank 1st to 6th)**
1.	Assessment encourages learning		
2.	Assessment provides feedback on learning to both the student and the teacher		
3.	Assessment documents competency and skill development		
4.	Assessment allows students to be graded or ranked		
5.	Assessment validates certification and licence procedures for professional practice		
6.	Assessment allows benchmarks to be established for standards		

Appendix H: Pages of the assessment evaluation tool

Assessing the Assessments: Evidencing and benchmarking student learning outcomes in chemistry
[OLT ID14-3652]

Australian Government | Office for Learning & Teaching

Addressed?	A "yes" should be given if students are ***exposed to any parts of the TLO in any way***, either implicitly or explicitly, to any level of achievement through completion of the task as instructed.
Demonstrated?	A "yes" should be given if students ***are directly instructed to demonstrate or evidence their attainment*** of any parts of the TLO, to any level of achievement, through completion of the task as instructed.
Awarded credit?	A "yes" should be given if ***"marks", "grades", compulsory "hurdles" or other affirmations of attainment*** are dependent on students providing ***clear evidence of achievement*** with respect to the TLO, within their submitted work.

ARE ANY PARTS OF THE TLO AT ANY LEVEL...

		Addressed?	Demonstrated?	Awarded credit?	
TLO 1	*Understand ways of scientific thinking by:*				
1.1	Recognising the creative endeavour involved in acquiring knowledge, and the testable and contestable nature of the principles of chemistry	yes	yes	yes	C
1.2	Recognising that chemistry plays an essential role in society and underpins many industrial, technological and medical advances	yes	yes	-- (yes / no / --)	D
1.3	Understanding and being able to articulate aspects of the place and importance of chemistry in the local and global community	no			
TLO 2	*Exhibit depth and breadth of chemistry knowledge by:*				
2.1	Demonstrating a knowledge of, and applying the principles and concepts of chemistry	no			

Scope | Breadth and Depth | Evaluation result | choices

Assessing the Assessments: Evidencing and benchmarking student learning outcomes in chemistry
[OLT ID14-3652]

Australian Government | Office for Learning & Teaching

Credited?	A part of the TLO is credited if there is a ***measurement or judgement of student capability*** with respect to the TLO, ***explicitly associated with cedit*** (marks/ hurdles etc) towards "passing" the task.
At Graduate level?	Graduate level attainment of any part(s) of a TLO statement is the ***minimum standard of achievement sufficent for a student to be awarded a major in chemistry*** (At AQF Level 7). The *revised Bloom's Taxonomy* may be used as a tool to aid this judgement
Required to pass?	Any part (or parts) of a TLO which is(/are) ***necessary to be awarded a passing grade***. That is, no student could "pass" without completing them satisfactorily. Examples may include compulsory hurdle components of the task, or components of the task worth a substantial enough proportion of the marks that the student would necessarily fail if those components were not successfully completed.

"All parts" of the TLO statement refers to the full ***breadth*** of content implied by the TLO statement, ***not the depth*** at which it is covered.

IN THIS TASK, WHICH PARTS OF THE TLO STATEMENT ARE...

		Credited?	Credited at graduate level?	Required to pass?	Required at graduate level to pass?	without confirmation	with confirmation	Overall
TLO 1	*Understand ways of scientific thinking by:*							
1.1	Recognising the creative endeavour involved in acquiring knowledge, and the testable and contestable nature of the principles of chemistry	all parts	some parts	all parts	-- (no parts / some parts / --)			
1.2	Recognising that chemistry plays an essential role in society and underpins many industrial, technological and medical advances							
1.3	Understanding and being able to articulate aspects of the place and importance of chemistry in the local and global community							
TLO 2	*Exhibit depth and breadth of chemistry knowledge by:*							
2.1	Demonstrating a knowledge of, and applying the principles and concepts of chemistry							
2.2	Recognising that chemistry is a broad discipline that impacts on, and is influenced by, other scientific fields							
TLO 3	*Investigate and solve qualitative and quantitative problems in the chemical sciences by:*							
3.1	Synthesising and evaluating information from a range of sources, including traditional and emerging information technologies and methods							

Scope | Breadth and Depth | Evaluation result | choices

TASK TITLE:

INSTITUTION:

SUBMITTER: **EMAIL:**

Chemistry Threshold Learning Outcome coverage for a student awarded a "pass" for this task.

Threshold Learning Outcomes for Chemistry														
1.1	1.2	1.3	2.1	2.2	3.1	3.2	3.3	3.4	3.5	4.1	4.2	5.1	5.2	5.3
C	D													

LEGEND:

- **A** Addressed
- **D** Demonstrated
- **C** Credited
- without confirmation
- with confirmation

Some parts, none at graduate level | Some parts, at graduate level | All parts, none at graduate level | All parts, at graduate level | All parts, only some at graduate level

TLO 1 *Understand ways of scientific thinking by:*	
1.1 Recognising the creative endeavour involved in acquiring knowledge, and the testable and contestable nature of the principles of chemistry 1.2 Recognising that chemistry plays an essential role in society and underpins many industrial, technological and medical advances 1.3 Understanding and being able to articulate aspects of the place and importance of chemistry in the local and global community	*[enter comments]*
TLO 2 *Exhibit depth and breadth of chemistry knowledge by:*	
2.1 Demonstrating a knowledge of, and applying the principles and concepts of chemistry 2.2 Recognising that chemistry is a broad discipline that impacts on, and is influenced by, other scientific fields	*[enter comments]*
TLO 3 *Investigate and solve qualitative and quantitative problems in the chemical sciences by:*	
3.1 Synthesising and evaluating information from a range of sources, including traditional and emerging information technologies and methods 3.2 Formulating hypotheses, proposals and predictions and designing and undertaking experiments 3.3 Applying recognised methods and appropriate practical techniques and tools, and being able to adapt these techniques when necessary 3.4 Collecting, recording and interpreting data and incorporating qualitative and quantitative evidence into scientifically defensible arguments 3.5 Demonstrating the cooperativity and effectiveness of working in a team environment	*[enter comments]*
TLO 4 *Communicate chemical knowledge by:*	
4.1 Presenting information, articulating arguments and conclusions, in a variety of modes, to diverse audiences, and for a range of purposes 4.2 Appropriately documenting the essential details of procedures undertaken, key observations, results and conclusions	*[enter comments]*
TLO 5 *Take personal, professional and social responsibility by:*	
5.1 Demonstrating a capacity for self-directed learning 5.2 Demonstrating a capacity for working responsibly and safely 5.3 Recognising the relevant and required ethical conduct and behaviour with which chemistry is practised	*[enter comments]*

Appendix I: Outputs

Books:

M. Schultz, S. Schmid & T. Holme, (Eds., 2016), Technology and Assessment Strategies for Improving Student Learning in Chemistry, *Proceedings of the Technology and Assessment Strategies for Improving Student Learning in Chemistry Symposium, Pacifichem Conference 2015*. Washington: ACS Books. DOI: 10.1021/bk-2016-1235

Book chapters:

S. Schmid, M. Schultz, S. J. Priest, G. O'Brien, S. M. Pyke, A. Bridgeman, K. F. Lim, D. C. Southam, S. B. Bedford, I. M. Jamie (2016). Assessing the Assessments: Development of a tool to evaluate assessment items in chemistry according to learning outcomes. In M. Schultz, S. Schmid, T. Holme (Eds.), Technology and Assessment Strategies for Improving Student Learning in Chemistry, *Proceedings of the Technology and Assessment Strategies for Improving Student Learning in Chemistry Symposium, Pacifichem Conference 2015*, Washington: ACS Books. DOI: pubs.acs.org/doi/abs/10.1021/bk-2016-1235.ch013

Journal publications:

M. Elmgren, F. Ho, E. Åkesson, S. Schmid & M. Towns, *J. Chem. Educ.* **92**, 427 - 432 (2015). Comparison and evaluation of learning outcomes from an international perspective: Development of a best-practice process. DOI: 10.1021/ed500542b.

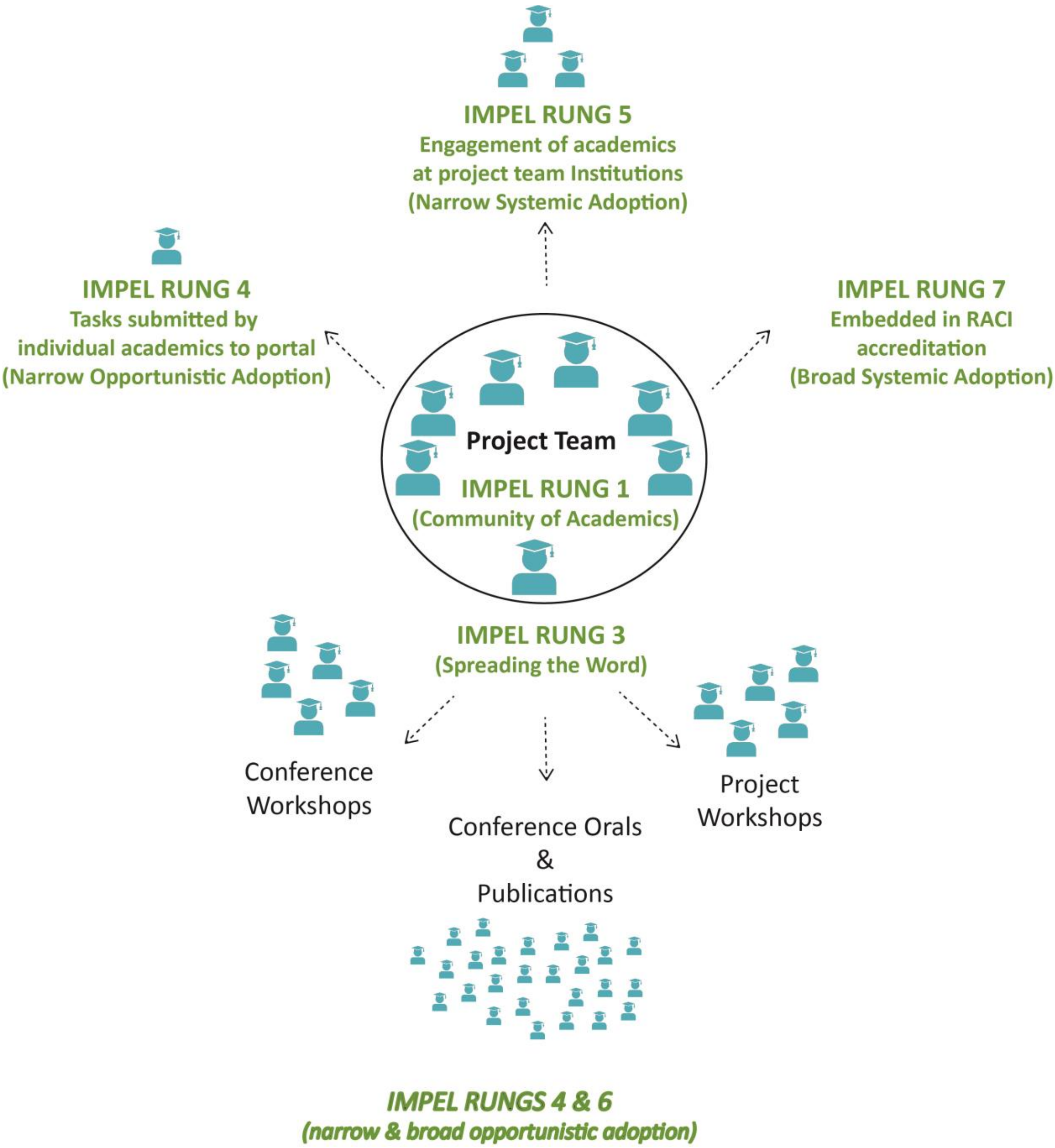
Conference symposia:

December 2015	Technology and Assessment Strategies for Improving Student Learning in Chemistry (Schultz and Schmid with Holme)

Oral presentations:

September 2014	ACSME Discipline Day, University of Technology Sydney (Schmid)
September 2015	ACSME Conference, Curtin University, Perth (Schmid)
December 2015	Pacifichem Conference, Honolulu, USA (Schmid)
July 2016	BCCE Conference, Colorado, USA (Schultz)
September 2016	ACSME Conference, University of Queensland, Brisbane (Schmid)
July 2017	RACI Congress Melbourne (Schultz)
July 2017	RACI Congress Melbourne (Schmid)
September 2017	ACSME Conference, Monash University, Melbourne (Schmid)

Appendix J: IMPEL framework



Appendix K: Contacts/requests

E-mail from **Damian Laird**, Murdoch University, 07/08/2017:

I'm just following up on our brief conversation regarding the Excel spreadsheet you have developed for the Assessing the Assessments project. As Leonie and I mentioned, we would be interested in getting a copy of that as the teaching group here at Murdoch is wanting to look more closely at how our assessments are actually working (or not). And we're planning on going through the RACI course accreditation process in the second half of this year, so tools that give us information we can feed into that would be most useful.

Would it be possible for you to email us that working spreadsheet you have?

E-mail from **Peter McCallum, Director, Education Strategy**, The University of Sydney, 15/06/2017:

Just a quick note of thanks for your contribution to the Academic Board forum on Tuesday. I thought it was an engaged session and was glad we had you experience and examples to give a concrete aspect to the theoretical discussion. You example about lab skills was a very good one in terms of getting feedback on learning and the need to be sure that the assessment task is carefully designed to test the skill it is supposed.

E-mail from **Ron Oliver, Deputy Vice-Chancellor** (Teaching & Learning), Edith Cowan University, 18/07/2016:

I am doing some workshops soon in Malaysia for a group of Saudi academics from Jazan University. My workshops are all about program learning outcomes, threshold standards and benchmarking. I have been watching Daniel Southam's work at Curtin so I have seen quite a bit of the good work being done in this space in Chemistry across Australia.

I have read your document Assessing the Assessments: Evidencing and Benchmarking Student Learning Outcomes in Chemistry which was used by the HES and Sara Booth team in their recent Peer Review workshops across the country. I worked with Sara on the PRAN project.

I was wondering if you would give me permission to use this document as a resource with the Saudi academics as an example of best practice in this space? I'm sure many will be impressed with the science behind the approach and it may be that some might want to follow up with you. Not sure what your thoughts might be about this.

Anyway, I think the document provides a wealth of good information which could be shared as I have suggested and was wondering if you were amenable to this.

E-mail from **Ron Oliver, Deputy Vice-Chancellor** (Teaching & Learning), Edith Cowan University, 24/03/2017:

I did use your materials in my workshop as I requested. I used the Powerpoint presentation as a resource for the participants (academics from Saudi Arabia) to peruse as an example of a programmatic benchmarking activity.

E-mail from **Sara Booth, Principal Fellow, Strategic Advisor-Quality External,** University of Tasmania and Jacqui Elson-Green, Convenor, Higher Education Compliance and Quality Network, H Ed Services, 19/07/2016:

Thank you for your generous support of the peer review of assessment workshops and for attending the Sydney events. Your presentation helped to ensure the success of the workshop and, as you will have realised from the questions you received, participants appreciated being able to interact with you and learn more about your work.

The workshops attracted just 600 people from 81 institutions which reflects the interest within the higher education sector about peer review of assessment and recognition of its importance in enhancing learning and teaching. The attendance figures also suggest that the workshops have played a role in capacity building within the sector and your contribution has enabled this to occur.

Presentations from all the workshops have been uploaded to the Higher Ed Services and University of Tasmania websites as part of our strategy to provide the sector with resources for peer review of assessment.

We received very positive feedback about the value of the workshops and many participants expressed a strong desire for further events. We hope that you will consider participating in future workshops as we value your expertise and knowledge.

Appendix L: Chemistry threshold learning outcomes (CTLOs)

TLO 1. Understanding the culture of chemistry	***Understand ways of scientific thinking by:***	1.1 recognising the creative endeavour involved in acquiring knowledge, and the testable and contestable nature of the
		1.2 recognising that chemistry plays an essential role in society and underpins many industrial, technological and medical advances
		1.3 understanding and being able to articulate aspects of the place and importance of chemistry in the local and global community
TLO 2. Scientific knowledge	***Exhibit depth and breadth of chemistry knowledge by:***	2.1 demonstrating a knowledge of, and applying the principles and concepts of chemistry
		2.2 recognising that chemistry is a broad discipline that impacts on, and is influenced by, other scientific fields
TLO 3. Inquiry, problem solving and critical thinking	***Investigate and solve qualitative and quantitative problems in the chemical sciences by:***	3.1 synthesising and evaluating information from a range of sources, including traditional and emerging information technologies and
		3.2 formulating hypotheses, proposals and predictions and designing and undertaking
		3.3 applying recognised methods and appropriate practical techniques and tools, and being able to adapt these techniques
		3.4 collecting, recording and interpreting data and incorporating qualitative and quantitative evidence into scientifically defensible
		3.5 demonstrating the cooperativity and effectiveness of working in a team environment
TLO 4. Communication	***Communicate chemical knowledge by:***	4.1 presenting information, articulating arguments and conclusions, in a variety of modes, to diverse audiences, and for a range
		4.2 appropriately documenting the essential details of procedures taken, key observations, results and conclusions
TLO 5. Personal and social responsibility	***Take personal, professional and social responsibility by:***	5.1 demonstrating a capacity for self---directed
		5.2 demonstrating a capacity for working responsibly and safely
		5.3 recognising the relevant and required ethical conduct and behaviour within which chemistry is practised