

## Chapter 13

# Assessing the Assessments: Development of a Tool To Evaluate Assessment Items in Chemistry According to Learning Outcomes

Siegbert Schmid,<sup>\*,1</sup> Madeleine Schultz,<sup>2</sup> Samuel J. Priest,<sup>3</sup>  
Glennys O'Brien,<sup>4</sup> Simon M. Pyke,<sup>3</sup> Adam Bridgeman,<sup>1</sup>  
Kieran F. Lim,<sup>5</sup> Daniel C. Southam,<sup>6</sup>  
Simon B. Bedford,<sup>4</sup> and Ian M. Jamie<sup>7</sup>

<sup>1</sup>School of Chemistry, The University of Sydney,  
Sydney, NSW 2006, Australia

<sup>2</sup>School of Chemistry, Physics and Mechanical Engineering,  
Queensland University of Technology, Brisbane QLD 4001, Australia

<sup>3</sup>School of Physical Sciences, The University of Adelaide,  
Adelaide, SA 5005, Australia

<sup>4</sup>School of Chemistry, University of Wollongong,  
Wollongong, NSW 2522, Australia

<sup>5</sup>School of Life and Environmental Sciences, Deakin University,  
Melbourne, VIC 3125, Australia

<sup>6</sup>Department of Chemistry, Curtin University, Perth, WA 6102, Australia

<sup>7</sup>Department of Chemistry and Biomolecular Sciences,  
Macquarie University, Sydney, NSW 2109, Australia

\*E-mail: siegbert.schmid@sydney.edu.au

Higher education in Australia is in a phase of rapid change due to significant regulatory changes, with new standards currently being implemented for registration of institutions and accreditation of degrees. Over the past five years the Australian chemistry community has come to a consensus on common Chemistry Threshold Learning Outcomes (CTLOs) that every Bachelor level chemistry graduate from an Australian university will have attained. The CTLOs will inform the standards used to accredit institutions and degrees. Building upon this, the Royal Australian Chemical Institute (RACI), the professional body for chemists in Australia, has changed its

accreditation process for chemistry degree programs and now uses these CTLOs as the basis for accreditation. Therefore, it is paramount to ensure that assessment items used allow students to demonstrate attainment of the CTLOs for a chemistry major. The “Assessing the Assessments” project has used an iterative process to develop an evaluation framework to assist academic staff at tertiary institutions to determine the alignment of their assessment items with the CTLOs. In conjunction with professional development workshops in which colleagues explore the alignment of assessment items with the CTLOs, a sophisticated tool has been developed which can be used to evaluate assessment items. The tool yields ratings for both engagement with and assessment of each CTLO within the assessment task evaluated, highlighting areas of potential improvement in current assessment practices. Comparison of self-evaluations of tasks submitted to the project by academic staff with evaluations conducted by the project team shows that in the majority of cases, faculty over-estimate the ability of their assessment items to confirm achievement of CTLOs. Recommendations to increase the coverage of CTLOs through changes to assessment procedures are presented. Through the development of the framework, difficulties with interpretation and application of some of the CTLOs have been elucidated.

## Introduction

In Australia, and increasingly worldwide, higher education institutions define attributes that graduates are expected to attain through their education (*1*, *2*). 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, including community responsibility and ethical behaviour (*3–5*). 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 (*6*, *7*). Such outcomes can be aligned to shared national (*8*) or international (*9*) normative practices and present a complex array of imposts on curriculum. A recent major project in Australia has examined methods for assuring graduate capabilities, particularly in relation to employability (*10*).

The elucidation of outcomes within a curriculum is seen as an essential cue to learner and teacher about the intention of a learning environment (*11*). Effective alignment of the objectives of the teacher to the outcomes of a learner is most influenced by what students do (*12*), and within the learning environment this is most frequently measured through assessment (*13*). Assessment of such outcomes, especially those that are shared (*14*) or are transferable across discipline boundaries (*15*) presents challenges (*16*). These challenges emerge from a shared

understanding of the outcomes, both between teacher and learner, and from those outside the immediate learning environment.

Assessment aligned to desirable outcomes can establish life-long learning (17) and skills necessary for employability (18). In the science curriculum there is frequently a gap between what is intended by teachers and what is actually achieved by learners (19), and by extension there is a gap in how the learner is assessed (20). This is especially true when intangible outcomes are reduced to facile assessment practices (21). 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 toward the attainment of a degree (22).

In this context, the corresponding regulatory requirements for Australian institutions have recently been expressed in the Higher Education Standards Framework (23) 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.*

The learning outcomes specified in the Higher Education Standards Framework thus also include both discipline-specific and generic skills, analogous to typical institutional statements of graduate attributes.

Statements 3 and 4 in the Framework indicate that for institutions to satisfy these requirements, methods of assessment must be evaluated to ensure that they allow demonstration of learning outcomes. In addition, the Framework implies that all required learning outcomes must have been demonstrated by every graduate. In order to ensure that this is the case, the degree structure must be conditional on achieving 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 (24). That project was established in 2009 by the Australian Learning and Teaching Council (ALTC) 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 (25, 26). 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, the base and stem embody a particular aspect of knowledge, skills and/or attributes that every graduate of the discipline will have explicitly demonstrated through assessment.

Within chemistry, the discipline community developed chemistry-specific TLOs as a derivative of the science outcomes, the CTLOs (27). The current two-tier set of CTLOs (28) states the following:

Upon completion of a bachelor degree with a major in chemistry, graduates will be able to:

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.
  
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.
  
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.

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.

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 within which chemistry is practised.

Following meetings organized by the Chemistry Discipline Network (ChemNet) in 2012 and 2013 to elucidate the CTLOs, further levels of detail have been expressed for CTLOs 2.1 and 3.3 as a third tier (29, 30). 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 demonstrate achievement of CTLOs was required. The design of assessment tasks is critical because excellent assessment task design can optimise student learning (22, 31–34). In contrast, poor task design may, for example:

- stop a ‘good’ student from demonstrating a high level of capability;
- prevent an ‘average’ student from meeting minimum performance requirements; or
- allow a ‘poor’ student to obtain a passing grade without having specifically met any outcomes.

While much has been written in the area of assessment design (17, 32–37), faculty who are designing assessment tasks are often not aware of the outcomes. This makes review of such tasks against shared outcomes by peers challenging, requiring extensive consultation to permit effective benchmarking (38). Two recent projects in Australia have addressed the utility of peer review of assessment

in the context of comparability and standards when implementing the threshold learning outcomes (39, 40). 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. In 2014, the Australian Government's Office for Learning and Teaching funded our project *Assessing the assessments: Evidencing and benchmarking student learning outcomes in chemistry* (OLT ID14-3652). This work outlines a process, informed by the literature and local curricular practices, to develop such a framework aligning a shared set of outcomes to a purpose-built tool. Through this process the research team have identified the critical elements necessary to prompt the identification of effective assessment of shared outcomes for the purposes of peer review. In this manuscript we describe the development of a self-evaluation tool and initial outcomes of its application. This project aimed to have an impact on all teaching staff in Australian chemistry departments and to transform assessment practice through direct and specific feedback to task designers.

## Methodology

The first stage of this project required a number of assessment items for preliminary evaluation and to obtain an overview of what would be involved in the evaluation process with a diverse set of assessment items. It was critical to not only have access to the assessment item as given to students, but also associated documentation including the criteria for assessment as well as samples of student work to see how the criteria were applied. The first assessment items evaluated were provided by members of the project team from their own teaching practice. Over a series of meetings a small selection of those assessment items was discussed within the project team and a preliminary template was developed as a pro forma for item submission by the wider community. The pro forma was refined upon further discussion and as shortcomings were identified after application to a wider variety of assessment tasks.

Using the refined pro forma as a guide, an on-line submission portal was developed to allow assessment items to be submitted to the project team and the associated documentation to be uploaded. This tool first collected information about the item, including how it fitted into the assessment pattern of the unit or subject (equivalent to an American "course") 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 next section allowed the submitter to nominate which CTLOs they thought were demonstrated (fully or partially) by successful completion of the assessment item. Finally, submitters were asked whether quality assurance processes were used in developing the item and how they ensured that the assessment item was valid and reliable.

The evaluation of whether graduates from a particular university degree program have achieved all CLTOs would normally be evidenced through examination of assessment tasks in upper year levels. Nevertheless, the project team invited the submission of assessment tasks from all year levels to build up

a varied collection of assessment tasks. Thus, for some tasks it was not expected that they required performance at a graduate level.

In parallel with collecting assessment items through the pro forma, a series of professional development workshops was held around Australia to trial and disseminate the evaluation process. Professional development workshops have been used in a related project and were shown to reduce variability in marking of threshold learning outcomes in accounting (38). That project specifically attempted benchmarking of final semester assessment items and so differs in aim and scope from the work described here. For the workshops conducted in this project, two assessment items were discussed on each occasion: one exemplar item supplied by the project team and one item submitted by workshop delegates. Discussions at these workshops were guided by the project team's suggested evaluation process and served to increase awareness within the chemistry education community of the importance of careful assessment design and the application of the CTLOs. In particular, 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 will be reported separately.

Independently of the workshops, which only discussed a small number of assessment items to stimulate discussion about the process, evaluations of assessment items 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 final evaluation of each assessment item. This extended process together with the workshops aimed to build a mutual understanding of the CTLOs and what is required for a task to confirm their achievement. Over the course of the project, the procedure used to evaluate tasks at workshops and amongst members of the project team has continually evolved to accommodate issues evident from feedback received.

## Results and Discussion

### Development of the Evaluation Tool

The initial evaluation procedure was split into two stages. First, a judgement was made as to whether the task's design allows for the assessment of each base (first tier) CTLO claimed, based solely on the marking scheme and instructional material for the task as provided to students. Subsequently, examples of student work associated with the task were used to confirm whether student outputs actually did exemplify each CTLO claimed, in line with the design of the task. Evaluators provided a yes/no or met/not met rating to each of the base CTLOs at the two stages, respectively.

The straightforward yes/no rating was immediately found to be too simple, as evaluators wanted the ability to rate partial engagement with a CTLO. In addition, evaluation needed to distinguish whether the task merely allowed potential engagement with the CTLO or whether it was actually required to be demonstrated. A further issue noted in the workshops was that some evaluators

had the tendency to re-mark student work rather than evaluating the design of the assessment item itself. It was also recognised that the potential for activities to engage students with a CTLO (task design) was often conflated with whether or not their achievement of the CTLOs was actually evaluated (assessment design).

To resolve these issues, the task evaluation procedure was then revised into three stages which concerned the task design, student work and assessment strategy respectively. Here the “task design” refers to the structure of the activities conducted by the students in completion of the assessment item, judged based on the instructional material provided to students. The “student work” comprises the material a student may submit towards marking and assessment, potentially evidencing their attainment or otherwise of the task’s learning objectives. Finally, the “assessment strategy” refers to the means by which attainment or otherwise of the task’s learning objectives is judged: typically a rubric or marking scheme. Thus, the revised evaluation procedure included the following three parts for each first tier CTLO:

1. Task design: Does the task design *allow students to engage with* the CTLO? Does the task design *require the demonstration of* the CTLO?
2. Student work: Does the student work provide evidence that the student has achieved the CTLO?
3. Assessment design: For each CTLO identified in Stage 1, does the assessment strategy require the student to evidence achievement of the CTLO in order to gain credit?

The first and second parts of this evaluation process were similar to the previous format, with the exception that it was explicitly recognised that a task could potentially allow engagement with a CTLO without requiring that engagement to complete the task. The final part of this evaluation format was intended to be informed by the first two stages and explicitly relate to the assessment strategy only, keeping judgement of CTLO attainment (assessment) distinct from the activities enabling CTLO engagement (design of the task). The provision of evidence by students was emphasised, because workshop discussion had revealed that many tasks allowed students to engage with a CTLO through the activities conducted, but evidence of this was not contained within the submitted student work samples. Thus, the final part of the evaluation allows comment on the extent to which the assessment strategy (particularly the marking scheme) draws on student evidence to demonstrate that the CTLO has been met. This is crucial and is the origin of many differences in evaluation outcomes from submitters compared with the project team.

Following the completion of a number of evaluations in this format, some further issues were noted that made the process unsatisfactory to users. Evaluations using this format almost universally resulted in different comments about different sub-tiers of the first tier CTLOs, making it necessary to allow separate feedback for each. A modification was therefore made to allow evaluation of each CTLO at the second tier level (e.g. 2.1, 2.2, 2.3 etc.) rather than at the first level (e.g. 1, 2, 3).

The evaluation form also did not allow for partial ratings, which were recognised as necessary in many cases even when evaluating the CTLOs at the second tier. The evaluation forms were initially amended to include partial ratings, but confusion arose as to when a partial rating was appropriate. Through discussion of specific assessment items, two independent reasons were identified for partial ratings at the second tier level:

- addressing only part of the (second tier) CTLO statement, some of which are expressed as several parts (13); and
- addressing the second tier CTLO to a level insufficient for graduate standard.

These two dimensions are separate and therefore a response grid was designed with axes labeled “portion” and “level” to reflect these respective aspects of engaging with or demonstrating the second tier CTLOs. Thus, a ‘four-square’ decision matrix was developed using a 2 x 2 grid to evaluate both the level and portion of a CTLO addressed by either the task design, submitted student work or the assessment design (as defined previously). The “portion” dimension of the grid is shaded based on whether all features contained within the statement of the second tier CTLO are addressed. The “level” dimension of the grid is shaded based on whether the relevant portion of the CTLO is addressed at a level of scope and complexity suitable for graduate level attainment. Anderson and Krathwohl’s revision to Bloom’s taxonomy (28) was suggested as a tool to assist in classifying the “level” dimension: broader scope (a range of types from the knowledge dimension of the taxonomy) and greater complexity (higher stages of the cognitive domain of the taxonomy) were suggested to be associated with “graduate” level work, whilst narrow scope and lesser complexity were suggested to be more appropriate to a “developing” level. Figure 1 shows the matrix used and Figure 2 shows the five possible outcomes of the evaluation.

		PORTION	
		Part of TLO	Whole TLO
LEVEL	Graduate		
	Developing		

Figure 1. ‘Four-square’ decision matrix.

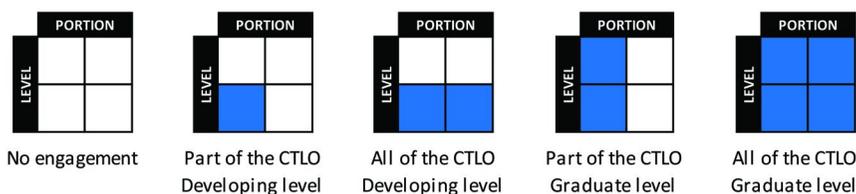


Figure 2. Five possible outcomes of the ‘four-square’ classification

This matrix allowed evaluators to simply and quickly shade the corresponding quadrants to illustrate their judgements for the different parts of the task evaluation procedure, for each CTLO. It also allowed rapid visual comparison of matrices completed by different members of the project team and workshop participants, highlighting components of the CTLOs found not to be addressed within the task evaluated.

In its first iteration, this ‘four square’ shading process was used in response to the three questions below, for each second tier CTLO. As previously, “task design”, “student work” and “assessment strategy” here refer to the structure of activities conducted, the work submitted for assessment and the explicit judgement of learning outcome attainment, respectively.

1. To what extent does the task design allow students to engage with the CTLO?
2. To what extent does the student work evidence attainment of the CTLO?
3. To what extent does the assessment strategy (particularly the marking scheme) require students to evidence the CTLO in order to gain credit?

However, it was found that this was time consuming and responses engaged far too heavily at the task design and student work sections (questions 1 and 2 above) with little engagement with the assessment component of the evaluation (question 3). Focus on exemplars of student work frequently resulted in re-marking rather than informing the final decision on the rating for assessment. Thus, in the final iteration, no component was included for the evaluation of student work. Instead student work exemplars were only used to illustrate any differences between the intended assessment strategy (as described in stated marking criteria) and the enacted assessment strategy (as observed in the allocation of marks in practice). The ‘four square’ format was used solely for evaluating engagement with the CTLO through the design of activities conducted by students (task design), whilst judgement of student attainment of the CTLO (assessment) was evaluated separately using a newly devised rating system.

Evaluation of the assessment aspect of a task consisted of determining whether marks or credit for the task are explicitly conditional on students demonstrating the evaluated component of the CTLO. One of four possible “assessment ratings” could be assigned: if there was no engagement with the task in the first component, a rating of zero was recorded. If any level of engagement had been determined, but no marks were conditional on demonstrating the CTLO, then a rating of 1

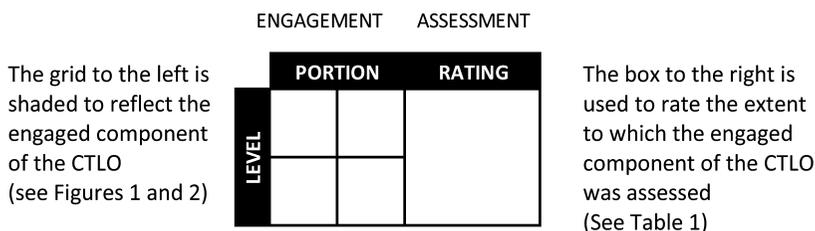
was given. If marks were conditional on demonstrating the engaged component of the CTLO, the assessment rating was decided based on whether a student could feasibly obtain a “pass” mark for the task without demonstrating the engaged component of the CTLO. If so, then a rating of 2 was recorded. If not, an optimal rating of 3 was assigned. Table 1 summarises the rating method for assessment that was developed.

**Table 1. Assessment ratings for the engaged component of CTLOs**

0	There is no engagement with the CTLO.
1	The task is designed to allow <i>engagement</i> with the engaged component of the CTLO, but <i>no marks or credit</i> are conditional on demonstrating it.
2	<i>Marks or credit</i> are explicitly conditional on demonstrating the engaged component of the CTLO, but <i>students could feasibly pass</i> the task without doing so.
3	Students <i>cannot pass</i> the task unless the engaged component of the CTLO is demonstrated.

The numerical ratings were designed to specifically highlight the difference between allowing engagement with a CTLO versus actually assessing its attainment (the difference between 1 and 2), and additionally to directly focus on whether passing the assessment could be said to confirm attainment of the CTLO (the difference between 2 and 3).

The full tool for evaluation of assessment items is shown in Figure 3. Although it was developed in an organic way based on the CTLOs, it can be applied to evaluate assessment items against any desired set of learning outcomes to be achieved. The ‘four-square’ grid to the left reflects student engagement with the CTLO through the design of activities students are required to complete (task design) whilst the square to the right is used to report an assessment rating, judged using stated marking criteria.



*Figure 3. Final tool used for evaluation of each CTLO.*

This tool allowed for a multidimensional final scoring for each CTLO for a particular assessment item, allowing for several levels of differentiation between the multiplicity of different ways that assessment tasks are implemented within degree programs. Together, the engagement and assessment components gave the final rating for how capable each given assessment item was at confirming

a particular CTLO had been achieved, as required within the Higher Education Standards Framework (23). This response format resolved many ambiguities present in the previous task evaluations. A typical example of the outcome of evaluation of an assessment item using the tool is given in Figure 4.

		Chemistry Threshold Learning Outcomes														
		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
Engagement classification																
Assessment rating		0	1	0	1	0	2	0	0	3	0	2	2	1	0	1

Figure 4. Example of an evaluation of an assessment item using the tool developed in this project.

In the example above, the shaded ‘four-square’ component of the evaluation coupled with the assessment rating give a quick and simple visualisation of the degree to which each CTLO is assessed within the task.

For CLTO 1.2: *Recognising that chemistry plays an essential role in society and underpins many industrial, technological and medical advances*, the particular task evaluated above concerned the control and monitoring of atmospheric pollution, but not other features of the CTLO statement (part of the CTLO), at a high level of complexity in a rich, broad task pitched at final year undergraduate students (graduate level). However, no marks were seen to be allocated to students for directly evidencing this understanding in their work (Assessment rating 1). These ideas are all encompassed in the simple visual presented above of a two-vertical square engagement classification with an assessment rating of 1.

Similarly, for CTLO 3.4: *Collecting, recording and interpreting data and incorporating qualitative and quantitative evidence into scientifically defensible arguments*, the task above was judged to involve some components of the CTLO statement: namely the incorporation of qualitative and quantitative evidence into defensible arguments, but not others: there was no clear direct application of experimental techniques to collect the data (part of the CTLO). The parts of the CTLO covered were, however, engaged at a deep level of scope and complexity (graduate level). Evaluators also judged that no student could possibly obtain a “pass” for the task without demonstrating attainment of these components of the CTLO at graduate level, justifying an assessment rating of 3. Again, this information is all encapsulated within the simple and efficient format above.

### Preliminary Observations about Assessment Tasks Evaluated

With this tool, based on consideration of a variety of assessment items and feedback from the professional development workshops, the project team evaluated a further 17 different assessment items that had been submitted using this format. During this process our personal understanding of what was sufficient to demonstrate achievement of a CTLO was shared and a consensus was usually

reached. The availability of two-dimensional partial ratings largely solved previously reported problems with evaluating engagement with many CTLOs, particularly the multi-part CTLOs 3.2 “formulating hypotheses, proposals and predictions and designing and undertaking experiments” and 3.3 “applying recognised methods and appropriate practical techniques and tools, and being able to adapt these techniques when necessary” (29). Each of these CTLOs encompasses quite distinct sets of skills, which are at different levels of cognitive process within the revised Bloom’s taxonomy (41). Only tasks that require the higher order process are likely to address the full CTLO at the graduate level.

However, some ambiguity still existed regarding what it meant to have engaged with some CTLOs, due to issues of interpretation. For example, the meaning of “self-directed” learning within the statement of CTLO 5.1 is open to interpretation at present. The question exists as to how a task can be designed to explicitly engage students with this CTLO, given that it would seem a student cannot be said to be “self-directed” if the teacher directed them to complete the task. Ambiguity also exists for all TLOs regarding what it means for a CTLO to be attained at graduate level. Because a definition of the graduate level threshold itself was beyond the scope of the project, this issue has been flagged as a tension and remains unresolved.

Related to these issues are ambiguities regarding what constitutes engagement with CTLOs relevant to laboratory work, which is a fundamental part of the study of chemistry (42–44). Some assessment tasks asked students to design or describe experimental work, but did not actually involve laboratory work, raising questions as to whether tasks could be said to involve engagement with CTLOs 3.3, 3.4, 4.2 and 5.2. Tasks in which this issue arose involved laboratory preparation work including safety analyses, virtual laboratory work and descriptive tasks about laboratory work. Some members of the team considered that these tasks could achieve partial engagement and an assessment score of 2 in the tool (Figure 3). Other team members strongly felt that without physically entering a laboratory, there is no engagement with these CTLOs and they must have an assessment rating of zero, regardless of how marks are allocated for the task. This was a disagreement that may be resolved through further consultation within the chemistry community, because it is finally the decision of the chemistry community as to how the CTLOs should be interpreted.

If a student can obtain a pass mark by partially completing many assessment items in a one semester unit or a whole degree program of study, but without demonstrating any CTLO completely, this is an indication that the assessment format is inadequate and does not satisfy the requirements of the regulatory framework (23). Assessment items with an assessment rating of 3 within our tool prevent this; a pass mark can then only be obtained if the CTLO is demonstrated. However, the tool presented here applies to singular tasks only and does not in itself account for cumulative demonstration of CTLOs across many tasks. The evaluation of assessment practice only at the level of isolated assessment tasks thus presents a limitation in the tool’s capabilities, most evident for CTLOs 2.1 and 4.1. CTLO 2.1 refers to the principles and concepts of chemistry, which are expected to be taught and assessed throughout a three year degree. (Australia does not have a liberal arts tradition and has 3-year bachelor degrees in science and

related disciplines.) A single assessment task will likely only require engagement with and deliver assessment of a fraction of the full list of principles and concepts described in the third tier of this CTLO (2.1(i), 2.1(ii) etc), meaning that for this CTLO, partial engagement ratings were inevitable. A similar issue exists for CTLO 4.1, which addressed communication "...in a variety of modes, to diverse audiences, and for a range of purposes". Typical assessment tasks only covered one mode, one audience and a single purpose for the task, again meaning partial engagement with the CTLO was common within singular assessment tasks. The diversity required by these CTLOs is intended to be reached across a degree program, meaning that results of single task evaluations, performed using the tool presented here, must be viewed in the context of the wider program in which the single assessment task is a small part. In order to demonstrate that a student has achieved these CTLOs completely, all assessment tasks of their degree must be considered together. Such mapping of degrees has been undertaken at some institutions for specific programs of study (29, 45, 46) although without the detailed analysis of assessment items that our tool allows. Evaluation of an entire degree program including analysis of student work on all assessment items entails an enormous amount of work, and given the continual changes to assessment strategies may not be feasible or worthwhile.

Important insights into optimal assessment design together with characteristics of well-designed assessment tasks have been obtained through the process of evaluating assessment items using the two-stage engagement and assessment approach within our tool. The first significant finding is that having a well-defined marking rubric is essential. This is because otherwise it is impossible to determine whether a student could pass without demonstrating the CTLO, and whether marks are allocated for demonstrating the CTLO at all. Achieving an assessment rating of 2 (reflecting judgement of CTLO attainment) rather than a rating of 1 (reflecting engagement with the CTLO, without judgement of student attainment) could often be easily rectified by stated allocation of marks to various aspects of the task which were present, but not awarded credit explicitly. Second, we have found simple ways to improve existing assessment tasks without major changes or increases in workload, in particular with regard to all parts of CTLOs 1 and 4, as follows. Requiring commentary on the role of chemistry in society is easily added to many assessment items that at the moment do not assess any aspect of CTLO 1. Requiring oral presentations or other forms of reports rather than the typical laboratory or research reports can satisfy CTLO 4. Third, we found many tasks that involved group work, which were interpreted by submitters to therefore demonstrate achievement of CTLO 3.5. However, working in a group is not in itself teamwork. This observation suggests a lack of understanding among Australian chemistry faculty about how to teach and assess group work to encourage interdependence and cooperativity as required by the CTLO (47–49), as opposed to sharing a task, experiment, report or piece of equipment. This implies an urgent need for professional development on this topic.

The process that was undertaken by the project team also leads to further questions, including:

- How many CTLOs can reasonably be met in one assessment task, or what is the ideal number of CTLOs to attempt to achieve per assessment item?
- How do we know when a student has ‘achieved’ a CTLO?
- How many times does a CTLO need to be ‘achieved’ so that we have confidence in the student’s capabilities?

Having evaluated numerous assessment items within small groups of project team members, it was observed that almost all submissions, even from members of the project team, were overly optimistic in their ratings. That is, the self-evaluations claimed many more CTLOs and a more complete coverage of these compared with the evaluations conducted by the project team. A similar finding was reported regarding inquiry in laboratory experiments, with many faculty characterizing their experiments as involving inquiry although this was not supported by application of an appropriate rubric (50). Figure 5 illustrates this finding by showing the task submitter’s self-evaluation data for the same assessment item shown in Figure 4 compared with the outcome of the project team’s evaluation.

		Chemistry Threshold Learning Outcomes														
		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
Submitter rating		●	●	●	●	●	◐	◐	◐	◐			●	●	◐	◐
Engagement classification			■		■		■	■		■		■	■	■	■	■
Assessment rating		0	1	0	1	0	2	0	0	3	0	2	2	1	0	1

Figure 5. Comparison of submitter’s evaluation with project team evaluation of CTLOs for a single assessment item.

It can be seen that while the submitter considered that this item fully achieved seven of the CTLOs and partially achieved another six, the evaluation of the project team only found engagement with eight CTLOs in total. Of those, four did not have marks attached to the CTLO (rating of 1) and of those which were allocated marks, only one CLTO was necessarily demonstrated in order to obtain a “pass” (rating 3). In addition, it is worth noting that the project team found better engagement with two of the CTLOs than the submitter suggested.

To illustrate the extent of the issue with self-evaluation, Figure 6 shows the component of each CTLO judged by the project team to be assessed within the task, for all cases where the submitter of the task claimed the full CTLO was addressed. The number under each CTLO shows how many tasks were thought by submitters to address that CTLO in full, whilst darkness of the shading in each square reflects the proportion of these tasks in which that quadrant of the engagement classification was shaded by the project team (as per Figure 2) and had marks associated (assessment rating 2 or 3). Assessment of the whole TLO, as claimed by submitters in this data, would be observed as a fully shaded grid.

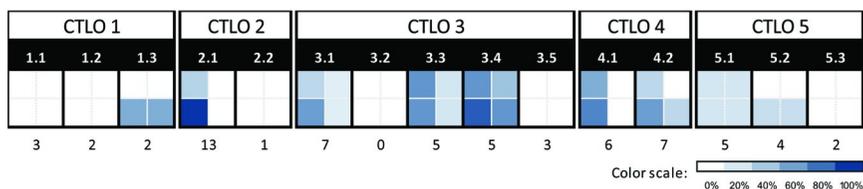


Figure 6. Components of the CTLOs assessed as judged by the project team for cases where the submitter claimed that the full CTLO was addressed.

There are several important pieces of information within Figure 6. First, the numbers of assessment items thought by submitters to engage students with each CTLO in full was highly variable. In fact CTLO 3.2 “formulating hypotheses, proposals and predictions and designing and undertaking experiments” has never been suggested by submitters to be addressed in full by any task in this sample of 17 assessment tasks. Some tasks within the sample are claimed by submitters to address this CTLO in part (not contributing to the data presented in Figure 6), but not in full. Much like CTLOs 2.1 and 4.1 discussed previously, this CTLO may also only be addressed fully using multiple tasks across an entire program of study.

Second, other than CTLOs 2.1 and 3.4, fewer than half of the assessment items thought by submitters to assess CTLOs in full were found by the project team to even assess them to the smallest extent (bottom left quadrant). That is, most items did not assess the CTLO even at a developmental level, but were considered by their submitters to be capable of confirming achievement of the full CTLO. Based on these preliminary observations, there is a gap here which may need to be remedied through extensive training and professional development.

Finally, it can be seen by the prevalence of white squares in the upper right hand quadrant that very few assessment items were found to assess the full CTLO at a graduate level. While many more assessment items are needed to obtain a full picture, this may already point to some problems for institutions in satisfying the new regulatory requirements.

## Conclusions

We have developed a sophisticated tool for determining the ability of assessment items to demonstrate attainment of stated learning outcomes. The tool leads to an engagement rating expressed as a 2 x 2 matrix examining the portion and level of engagement with the learning outcome, and a numerical assessment rating from 0 - 3 for actual marks awarded. Together these ratings give a quick visual overview of the coverage by an assessment item of a specified outcome. 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. CTLO 1.1, 1.2, 1.3 etc., not CTLO 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

(CTLO 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 stated learning outcomes are required to be demonstrated by students. It is clear from evaluation of the approximately 40 assessment items so far submitted that not all CTLOs are assessed equally well, and some do not seem to be assessed at all (29). Thus, use of the tool can inform design and modification of assessment tasks to ensure that students are given the opportunity to demonstrate 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.

## Acknowledgments

We are grateful to all of the participants in our workshops as well as to all of the people who submitted assessment items to this project. We thank Gwen Lawrie and Carmel McNaught, both of whom contributed extensively to the project. We acknowledge funding from the Australian Government's Office for Learning and Teaching (OLT ID14-3652).

## References

1. Barrie, S. C. A research-based approach to generic graduate attributes policy. *Higher Educ. Res. Dev.* **2004**, *23*, 261–275, DOI: 10.1080/0729436042000235391.
2. Barrie, S. C. A conceptual framework for the teaching and learning of generic graduate attributes. *Studies Higher Educ.* **2007**, *32*, 439–458, DOI: 10.1080/03075070701476100.
3. 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.
4. Macquarie University. *Learning and Teaching for Sustainability*. [https://www.mq.edu.au/about\\_us/strategy\\_and\\_initiatives/sustainability/areas\\_of\\_focus/education\\_for\\_sustainability/learning\\_and\\_teaching\\_for\\_sustainability/](https://www.mq.edu.au/about_us/strategy_and_initiatives/sustainability/areas_of_focus/education_for_sustainability/learning_and_teaching_for_sustainability/) (accessed June 14, 2016).
5. Institute for Teaching & Learning, University of Sydney. *The Sydney Graduate*. <http://www.itl.usyd.edu.au/graduateAttributes/> (accessed June 14, 2016).
6. 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.
7. Hughes, C.; Barrie, S. Influences on the assessment of graduate attributes in higher education. *Assess. Eval. Higher Educ* **2010**, *35*, 325–334, DOI: 10.1080/02602930903221485.

8. 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.
9. 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.
10. Oliver, B. *Assuring Graduate Capabilities*. <http://www.assuringgraduatecapabilities.com> (accessed June 14, 2016).
11. Biggs, J. Enhancing teaching through constructive alignment. *Higher Educ.* **1996**, *32*, 347–364, DOI: 10.1007/BF00138871.
12. Biggs, J. What the student does: Teaching for enhanced learning. *Higher Educ. Res. Dev.* **1999**, *18*, 57–75, DOI: 10.1080/0729436990180105.
13. 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. Higher Educ* **2009**, *34*, 565–577, DOI: 10.1080/02602930802226502.
14. Hager, P. Nature and development of generic attributes. In *Graduate Attributes, Learning and Employability*; Hager, P., Holland, S., Eds.; Springer: Dordrecht, The Netherlands, 2006.
15. Kemp, I. J.; Seagraves, L. Transferable skills—can higher education deliver? *Studies Higher Educ.* **1995**, *20*, 315–328, DOI: 10.1080/03075079512331381585.
16. Green, W.; Hammer, S.; Star, C. Facing up to the challenge: Why is it so hard to develop graduate attributes? *Higher Educ. Res. Dev.* **2009**, *28*, 17–29, DOI: 10.1080/07294360802444339.
17. Boud, D.; Falchikov, N. Aligning assessment with long-term learning. *Assess. Eval. Higher Educ.* **2006**, *31*, 399–413, DOI: 10.1080/02602930600679050.
18. Jackson, D. Employability skill development in work-integrated learning: Barriers and best practice. *Studies Higher Educ.* **2015**, *40*, 350–367, DOI: 10.1080/03075079.2013.842221.
19. 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, The Netherlands, 1998, pp 421–448.
20. Bryce, T. G. K.; Robertson, I. J. What can they do? A review of practical assessment in science. *Studies Sci. Educ.* **1985**, *12*, 1–24, DOI: 10.1080/03057268508559921.
21. Bowman, N. A. Understanding and addressing the challenges of assessing college student growth in student affairs. *Res. Pract. Assess.* **2013**, *8*, 5–14.
22. Sadler, D. R. Three in-course assessment reforms to improve higher education learning outcomes. *Assess. Eval. Higher Educ.* **2015**, 1–19, DOI: 10.1080/02602938.2015.1064858.
23. 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 June 14, 2016).
24. Ewan, C. Disciplines setting standards: The learning and teaching academic standards (LTAS) project. Paper presented at *Australian Quality Forum*, Gold Coast, 2010.
  25. 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](http://www.olt.gov.au/system/files/altc_standards_SCIENCE_240811_v3.pdf) (accessed June 14, 2016).
  26. 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.
  27. Buntine, M.; Price, W.; Separovic, F.; Brown, T.; Thwaites, R. *Learning and Teaching Academic Standards: Chemistry Academic Standards Statement* (Appendix 3), Australian Learning and Teaching Council, 2011. [http://www.olt.gov.au/system/files/altc\\_standards\\_SCIENCE\\_240811\\_v3.pdf](http://www.olt.gov.au/system/files/altc_standards_SCIENCE_240811_v3.pdf) (accessed June 14, 2016).
  28. 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](http://chemnet.edu.au/sites/default/files/u39/CHEMISTRY_Academic_Standards_Accreditation_Trial.pdf) (accessed June 14, 2016).
  29. 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. Sci. Math. Educ.* **2013**, *21*, 81–91.
  30. 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.
  31. 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](http://www.olt.gov.au/system/files/resources/GPR_Assessment_STEM_Students_Rice_2011.pdf) (accessed June 14, 2016).
  32. Gibbs, G.; Simpson, C. Conditions under which assessment supports students' learning. *Learn. Teach. High. Educ.* **2004**, *2004-2005*, 3–31.
  33. Gibbs, G.; Dunbar-Goddet, H. Characterising programme-level assessment environments that support learning. *Assess. Eval. Higher Educ.* **2009**, *34*, 481–489, DOI: 10.1080/02602930802071114.
  34. Boud, D., et al. *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](http://www.olt.gov.au/system/files/resources/Assessment%202020_final.pdf) (accessed June 14, 2016).
  35. Nicol, D. Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assess. Eval. Higher Educ.* **2009**, *34*, 334–352, DOI: 10.1080/02602930802255139.

36. Sadler, D. R. Formative assessment and the design of instructional systems. *Instr. Sci.* **1989**, *18*, 119–144, DOI: 10.1007/BF00117714.
37. Scouller, K. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Educ.* **1998**, *35*, 453–472, DOI: 10.1023/A:1003196224280.
38. 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. Higher Educ.* **2016**, *41*, 331–349, DOI: 10.1080/02602938.2015.1008398.
39. Booth, S.; Beckett, J.; Saunders, C. Peer review of assessment network: Supporting comparability of standards. *Qual. Assur. Educ.* **2016**, *24*, 194–210doi:10.1108/QAE-01-2015-0003.
40. 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](http://www.olt.gov.au/system/files/resources/SP10_1843_Krause_report_2014.pdf) (accessed June 14, 2016).
41. Krathwohl, D. R. A revision of Bloom's taxonomy: An overview. *Theory Pract.* **2002**, *41*, 212–218, DOI: 10.1207/s15430421tip4104\_2.
42. Bruck, L. B.; Towns, M. H. Faculty perspectives of undergraduate chemistry laboratory: Goals and obstacles to success. *J. Chem. Educ.* **2010**, *87*, 1416–1424, DOI: 10.1021/ed900002d.
43. Hofstein, A. The laboratory in chemistry education: Thirty years of experience with developments, implementation, and research. *Chem. Educ. Res. Pract.* **2004**, *5*, 247–264, DOI: 10.1039/B4RP90027H.
44. Reid, N.; Shah, I. The role of laboratory work in university chemistry. *Chem. Educ. Res. Pract.* **2007**, *8*, 172–185, DOI: 10.1039/B5RP90026C.
45. Watts, L.; Hodgson, D. Whole curriculum mapping of assessment: Cartographies of assessment and learning. *Soc. Work Educ.* **2015**, *34*, 682–699, DOI: 10.1080/02615479.2015.1048217.
46. Plaza, C. M.; Draugalis, J. R.; Slack, M. K.; Skrepnek, G. H.; Sauer, K. A. Curriculum mapping in program assessment and evaluation. *Am. J. Pharm. Educ.* **2007**, *71*, 1–8, Article 20.
47. Johnson, D. W.; Johnson, R. T. An educational psychology success story: Social interdependence theory and cooperative learning. *Educ. Res.* **2009**, *38*, 365–379, DOI: 10.3102/0013189X09339057.
48. Johnson, D. W.; Johnson, R. T. Making cooperative learning work. *Theory Pract.* **1999**, *38*, 67–73, DOI: 10.1080/00405849909543834.
49. Dunne, E.; Rawlins, M. Bridging the gap between industry and higher education: Training academics to promote student teamwork. *Innov. Educ. Train. Int.* **2000**, *37*, 361–371v10.1080/135580000750052973.
50. Fay, M. E.; Grove, N. P.; Towns, M. H.; Bretz, S. L. A rubric to characterize inquiry in the undergraduate chemistry laboratory. *Chem. Educ. Res. Pract.* **2007**, *8*, 212–219, DOI: 10.1039/B6RP90031C.