Improving Laboratory Learning Through Self and Peer Assessment of Laboratory Reports

Kieran Fergus Lim (林百君)[a]

Corresponding author: kieran.lim@deakin.edu.au
[a]School of Life and Environmental Sciences, Deakin University, Burwood VIC 3125, Australia

Keywords: academic writing, laboratory learning, peer assessment, self assessment

International Journal of Innovation in Science and Mathematics Education, *23*(2), 59-73, 2015

Abstract

The laboratory provides an opportunity for students to achieve many learning outcomes including to critically evaluate information, interpret and draw conclusions from scientific data, and communicate scientific results, information, or arguments. This paper describes a laboratory-writing task that involves self and peer evaluation. After discussion of the expectations of laboratory report writing during class, students self and peer evaluate reports. In a process similar to double-blind journal refereeing, students practise critically evaluating the quality of academic writing using a rubric. The summative assessment is based on how consistent their evaluations are with the evaluations of the same reports performed by their peers. The formative assessment is that students receive peer evaluations and feedback via a rubric on reports that they have written. The skill of critically evaluating their own reports is used to improve the laboratory reports in subsequent assessment tasks.

Introduction

Written communication including the ability to write laboratory reports is an important part of a science degree (Jones, Yates, & Kelder, 2011). In addition, employers list written communication skills as a desirable attribute when hiring graduates, regardless of discipline (Fair, Kleist, & Stoy, 2014; GTI Media, 2015; The Australian Chamber of Commerce and Industry and the Business Council of Australia, 2002; Wood et al., 2010). Increasingly, there is a recognition that the development of written communication skills cannot be left to central academic skills units, but should be embedded within the curriculum (Bailey, 2001; Gordon et al., 2001; Knight & Yorke, 2003; Paulson, 2001; Tucci, O'Connor, & Bradley, 2014; Whelan & Zare, 2003; Windsor, Rutter, McKay, & Meyers 2014). Furthermore, guided writing can enhance learning about science while unguided writing assignments have little effect (Moore, 1993).

The use of computers to prepare reports has made it easy to prepare and revise more than one draft of a report, of which the final version is assessed (Chickering & Ehrmann, 1996). At Deakin University, Lim (2009) has given students the opportunity to submit draft reports; after feedback, they can "do it again, thoughtfully" and only the final report is subject to summative assessment. However, this approach would be impractical for large classes. At La Trobe University, Blanksby and Chan (2006) have used a combination of peer-review or review by demonstrators to give feedback on draft reports, with the aim of improving the final report. Russell (2005, 2013) at UCLA has a sophisticated computer-based system that enables double-blind peer evaluation of draft reports.

This paper describes a recent implementation, in which double-blind peer evaluation of a report is combined with self evaluation of subsequent laboratory reports. The explicit aim of this process is to develop the students' metacognitive skill of judging or assessing the quality of a report. While sharing some similarities with other programs (Berry & Fawkes, 2010; Blanksby & Chan, 2006; Hristova, 2014; Lim, 2009; Russell, 2005, 2013; Walker & Sampson, 2013), the current implementation assesses the quality of the self evaluation and peer evaluation by measuring the consistency of the evaluations.

Research Methodology and Study Design

This study follows the principles of Action Research (Kemmis & McTaggart, 1988; Lewin, 1946, 1948; Lytle & Cochran-Smith, 1990; Wood et al., 2010; Zuber-Skerritt, 1992), in which a teacher/researcher:

1. Plans a course of action to remedy a perceived problem;
2. Acts to implement that course of action;
3. Observes the outcomes of the course of action; and
4. Reflects on the outcomes – what worked and what did not work – and uses that to conduct further cycles of planning, acting, observing and reflecting.

The perceived problem was that laboratory reports in level 2 and level 3 units were of mixed quality, including some reports of very low quality. It was believed that this problem was due to the inability of many students to judge the quality of their own work, objectively and realistically. The hypothesis was that a focus on how to write and evaluate a simple report, without complicated scientific content, would improve metacognitive self evaluation skills.

A teaching and learning exercise was implemented first to explicitly discuss the expectations of academic report writing, and second for students to practice self evaluation of their reports. A simple kitchen chemistry exercise was used so that students could focus on writing skills without being overly concerned about scientific content. While it would be expected that students cannot assess or evaluate pieces of academic writing as well an experienced marker, nevertheless, this intervention should improve the correlation between the student's self evaluation and that of an experienced academic staff marker. The quality of writing and the correlation between self and marker evaluation in later reports were used as primary measures of the effectiveness of this intervention. A secondary measure was based on the instructor's reflections.

In this paper, the term *evaluation* (*evaluate*) will refer to the judgement of the quality of a piece of academic writing, by returning a numerical *score*. The term *assessment* (*assess*) will refer to the process of assigning a *mark* for summative assessment. In the kitchen chemistry exercise described here, the *evaluation scores* were not the direct basis of the *summative assessment marks*. In the subsequent five laboratory reports, the *evaluation scores* from academic staff were the direct basis of the *summative assessment marks*, plus a small *mark* component based on the correlation between the student's self evaluation and that of an experienced academic staff marker.

The research described here has Human Research Ethics Approvals, Deakin University DUHR-EC-29-2008 and STEC-23-2012.

Study Context and Implementation

An early version of this teaching and learning intervention to improve laboratory report writing was based on the "do-ing it again, thoughtfully" (DIAT-ing) approach described by Chickering and Ehrmann (1996; Lim, 2009). The implementation was amended through several action research cycles in two units of study spread over several years; the implementation using self and peer evaluation in the summer of 2014-2015 is described here. The unit was a level 2 chemistry unit with 68 students. Six students had completed first year in the previous teaching session (semester). There were some single- and double-degree students who were at the end of their third or fourth year, including 12 students for whom this unit was their final unit before graduating.

As a summer-session unit, the unit was taught in intensive mode with a 3-hour block of lectures/workshops on four mornings in weeks 1 and 2, and on 2 mornings in week 3. Each student attended 2 laboratory sessions in the afternoons of weeks 2 and 3, and one morning laboratory session in week 3. The teaching and learning timetable is given in Table 1. Students completed the laboratory work in weeks 2 and 3, and the write-ups were completed and submitted between weeks 4 and 12. The kitchen chemistry exercise and laboratory reports were each worth 10% of the unit assessment.

Table 1. Teaching and learning timetable for unit. The two rows of entries for weeks 1-3 represent morning and afternoon sessions.

Monday	Tuesday	Wednesday	Thursday	Friday
Week 1				
Principles of report writing	Chemistry topics	Chemistry topics	Chemistry topics	*Kitchen chemistry report submitted*
Kitchen chemistry experiment (completed outside class time)				
Week 2				
Chemistry topics	Report evaluation, including discussion of actual student reports; use of EndNote™ for referencing; marking rubric	Chemistry topics	Chemistry topics	
Group A: Laboratory 1	Group B: Laboratory 1	Group A: Laboratory 2	Group B: Laboratory 2	
Week 3				
Group B: Laboratory 3	Chemistry topics	Analysis of laboratory data; Review	Group A: Laboratory 5	*Peer and self evaluations of kitchen chemistry report submitted*
Group A: Laboratory 3	Group B: Laboratory 4	Group A: Laboratory 4	Group B: Laboratory 5	
Weeks 4-12				
Week 4	*Laboratory report 1 and self evaluation submitted*			
Week 5	*Laboratory report 2 and self evaluation submitted*			
Week 6	*Laboratory report 3 and self evaluation submitted*			
Week 7	*Laboratory report 4 and self evaluation submitted*			
Exam week (week 12)	*Laboratory report 5 and self evaluation submitted; Case study submitted*			

The unit chair was responsible for all the lecture and workshop sessions and for the logistics of peer and self evaluations. This person also marked the major case study assessment; the case study is separate from the assessments discussed in this paper. Two other instructors demonstrated in the laboratory sessions and marked laboratory reports 1-5.

In week 1, class time was devoted to the discussion of the principles of good report writing, including:

- General principles
  - A: Accuracy (scientific validity and correctness)
  - B: Brevity (conciseness and relevance)
  - C: Clarity (precise communication)
- Important parts of the report
  - C: Context (introduction, background and literature review)
  - D: Data (method of gathering data; followed by evidence, results and observations)
  - E: Explanation (discussion, significance and meaning of data)
  - F: Findings (conclusions)

In a good report, the Explanation (discussion) of Data (results) will support the Findings (conclusions).

Students then completed a kitchen chemistry experimental investigation at home, in which the effect of added salt while cooking vegetables was investigated. This is linked to the area of molecular gastronomy, which is the science of chemical changes during cooking (Blanck, 2007; Le Blanc, 1992; Lister & Blumenthal, 2005; Palermo, Pellegrini, & Fogliano, 2014; Pellegrini et al., 2010; This, 2005). Other at-home experimental investigations are also possible (Kennepohl, 2007; Kennepohl & Shaw, 2010). A report on this investigation was submitted at the end of week 1.

In earlier versions of this unit, the unit chair had written the marking rubrics for the laboratory reports, but this apparently led to a lack of student "buy-in". For example, the previous rubric allocated 10% of each report mark to the correct use of citations, but even at the end of that iteration of the unit, many students were not putting citations in the body of the reports or were omitting essential information from the list of references. Thus, in week 2, class time was devoted to a discussion of how reports should be evaluated. The class discussed and voted on the point allocations for the rubric items and the criteria. There was some disagreement about various items; for example, some (including the unit chair) believed that no points should be allocated for an informative and descriptive report title, but a majority of the class wanted that item. The final negotiated rubrics had the weightings shown in Table 2. Some members of the class gave permission for their kitchen chemistry report to be discussed in class. Hence the week 2 class discussion also applied the rubric to these student reports in order to reach a class consensus of the expectations of the writing standard and of how to apply the rubric. The rubric for the kitchen chemistry report is given in the Appendix. The details of the scientific content part of the rubric varied for each report.

Table 2. Breakdown of the rubrics for the kitchen chemistry report and the five laboratory reports.

	Technical aspects of report writing	Agreement between self evaluation and academic marker's assessment of technical aspects of report writing	Scientific content	Summative assessment (marking)
Kitchen chemistry report (40 points, peer and self evaluation)	28 points	Not applicable	12 points	Based on comparison of peer and self evaluation
Laboratory reports (60 points each, academic marker's evaluation)	28 points	4 points	28 points	Based directly on academic marker's evaluation

Evaluation and assessment of the kitchen chemistry report

The kitchen chemistry report was peer and self evaluated, but was not assessed or marked by any member of academic staff; the marks for this exercise were based on a comparison of the peer and self evaluations as follows. Each kitchen chemistry report was manually assigned by the unit chair to peer evaluators, as shown in Table 3. The shaded cells illustrate the *systematic* allocation of reports to ensure:

- Each report was self evaluated by the student author;
- Each report was circulated to three or four classmate-peer evaluators;
- No set of student evaluators for any one report had more than one classmate-peer evaluators in common with the sets of student evaluators for any other report.

The unit chair manually assigned a code to each report, based on the last five digits of each student ID number. The report submitted by student 00004 was labelled with a code like 00004-00013-00011-00007-00001-2. Each five-digit block of numbers indicates a person: the author (self evaluator) and the three or four peer evaluators. The final single-digit, 2, indicates that that version of the report was circulated to the student identified by the 2nd block of digits (00013). This last digit is needed to uniquely label the *evaluations that are returned* by different students. The use of the last five digits of the student ID numbers was sufficient to uniquely distinguish each student. The use of the last part of the 9-digit student ID numbers, not the entire number, preserves student privacy.

Table 3. Illustrative allocation of peer evaluators to each kitchen chemistry report.

Report author and self evaluator	Peer evaluator 1	Peer evaluator 2	Peer evaluator 3	Peer evaluator 4
00001	Did not submit report			
00002	Did not submit report			
00003	Did not submit report			
00004	00014	00012	00008	00001
00005	00004	00013	00009	00002
00006	00005	00014	00010	00003
00007	00006	00004	00011	00001
00008	00007	00005	00012	00002
00009	00008	00006	00013	00003
00010	00009	00007	00014	00001
00011	00010	00008	00004	00002
00012	00011	00009	00005	00003
00013	00012	00010	00006	
00014	00013	00011	00007	

Journals have strict rules about how to prepare manuscripts for double blind refereeing. However, students did not follow the instructions to prepare their reports in a suitable format. Due to the tight timelines for circulating the reports to students for peer evaluation, the unit chair manually edited each report to remove all identifying names and student numbers, and renaming the report using the codes described above. This process took several hours, but ensured a faster turn-around than the estimated days that would have been involved if reports had been returned to students for editing into the correct format.

The marks for the kitchen chemistry task were based on the agreement between peer evaluation and self evaluation of the kitchen chemistry reports, rather than the quality of the reports. Each report was self evaluated and also had at least three peer evaluations. Comments made to the instructor indicated that the 13 students who chose not to submit a kitchen chemistry report did this to guarantee perfect agreement between peer evaluation and self evaluation scores of zero.

The scores for the kitchen chemistry were given out of 40 points (the rubric is provided in the Appendix). Following this process, the student evaluations were awarded a mark as follows:

- Full marks (4 out of 4), if the student's evaluation was within 5% (2 points) of the *range* of scores spanned by the other evaluations;
- Three marks (3 out of 4), if the student's evaluation was within 10% (4 points) of the *range* of scores spanned by the other evaluations;
- Two marks (2 out of 4), if the student's evaluation was within 15% (6 points) of the *range* of scores spanned by the other evaluations;
- One mark (1 out of 4), if the student's evaluation was within 20% (8 points) of the *range* of scores spanned by the other evaluations;
- Zero marks, if the student's evaluation was outside 20% (8 points) of the *range* of scores spanned by the other evaluations.

Tables 4 and 5 show illustrative evaluation scores from seven students for two separate reports. The evaluation scores in the Tables are not real class data.

Table 4. Example of marks awarded to four students who evaluated report 00013-00012-00010-00006 with scores of 24, 32, 37, and 36 out of 40.

Student	Student's score	Evaluation scores from other students	Mark	Comment
00006	24	32, 36, 37	1 out of 4	Peer evaluation is within 20% (8 points) of the range of other scores
00010	32	24, 36, 37	4 out of 4	Peer evaluation is within 5% (2 points) of the range of other scores
00012	37	24, 32, 36	4 out of 4	Peer evaluation is within 5% (2 points) of the range of other scores
00013 (author)	36	24, 32, 37	4 out of 4	Self evaluation is within 5% (2 points) of the range of other scores

It can be seen that Students 00010 and 00013 (Table 4) and Students, 00007, 00011 and 00013 (Table 5) had evaluations that were consistent with the *range* of scores spanned by the other students' evaluations, and therefore gained full marks. In Table 4, Student 00006 was a 'rogue' evaluator who returned an evaluation that was very inconsistent with those of the other students. Without Student 00006's 'rogue' evaluation, Student 00010 would have probably scored a lower mark. The presence of a rogue evaluator in the group has made it easier for the other students in the group to gain higher marks.

Table 5 shows the marking for a report that was evaluated to be of low quality by all four student evaluators. It is clear that the marks do *not* depend on the *quality* of the report, but only on the *consistency* of each student's evaluation with the other students' evaluations. In Table 5, Student 00014 has optimistically self evaluated the report, and received a lower mark than the three peers.

Table 5. Example of marks awarded to four students who evaluated report 00014-00013-00011-00007 with scores of 7, 4, 6 and 11 out of 40.

Student	Student's score	Evaluation scores from other students	Mark	Comment
00007	7	4, 6, 11	4 out of 4	Peer evaluation is within 5% (2 points) of the range of other scores
00011	4	6, 7, 11	4 out of 4	Peer evaluation is within 5% (2 points) of the range of other scores
00013	6	4, 7, 11	4 out of 4	Peer evaluation is within 5% (2 points) of the range of other scores
00014 (author)	11	4, 6, 7	3 out of 4	Self evaluation is within 10% (4 points) the range of other scores

The completed rubrics for each report were shared with all students who evaluated that report, so students could learn from the consistency (or lack thereof) of their evaluations (formative assessment). Significantly, it also means that each student can compare his/her self evaluation with peer evaluations.

In addition to his/her own self evaluation, each student was required to peer evaluate three other reports, so the final mark for this exercise (out of 16) is a combination of four evaluations like those shown in Tables 4 and 5. Students who chose not to submit a report are guaranteed a high mark (summative assessment) for the self evaluation of *their own reports*, but they lost the opportunity for feedback via peer evaluations of their reports (formative assessment).

Evaluation and assessment of the five subsequent laboratory reports

Students submitted both a laboratory report and a self evaluation of their technical writing for each of the five subsequent laboratory exercises. Academic staff marked these laboratory reports based on a 60-point rubric (see Table 2) including:

- evaluation of the technical writing using the same first part (first 28 points) of the rubric used for the kitchen chemistry report (see Appendix),
- comparison between the staff and student self evaluation of the technical writing, and
- evaluation of the scientific content.

Some students did not submit a self evaluation for some laboratory reports and lost the small mark component based on the agreement between self evaluation and staff assessment of technical aspects of report writing. Any student who did not submit a laboratory report received zero marks for that particular laboratory exercise. Many subsequent laboratory reports were submitted as team reports, resulting in less data for each subsequent laboratory exercise than the first kitchen chemistry report, which was submitted individually.

Results and discussion

This self and peer evaluation exercise was first implemented in the summer of 2012-2013. After some adjustment, it was implemented again in the summer of 2014-2015 and the results from that latter cohort are reported here. Fifty-five reports were written and underwent self evaluation. Of the 55 reports, 26 reports had three peer evaluations each, while 29 reports had four peer evaluations each. The additional peer evaluations ensured redundancy in case some students withdrew from the unit.

Figure 1 shows the raw evaluation scores. The technical aspects of laboratory report writing (Findlay, 1996; Lim, 2003; Silyn-Roberts, 1996; Tilstra, 2001) for each report were evaluated out of a maximum of 28 points (see Table 2 and the Appendix). These evaluations also had a small component on the scientific content of each report, which is not discussed here. Each report is associated with 3-4 data points in Figure 1, representing one self evaluation score plotted against 3 or 4 peer evaluation scores. For example, the three data points with a self evaluation score of 10 represent peer scores for a single report. All self evaluation scores with more than 4 data points represent several reports and there are some coinciding data points. The quality of the reports has wide variation, with some data points in the lower-left quadrant, associated with three reports that were both self and peer evaluated to be of fail standard.

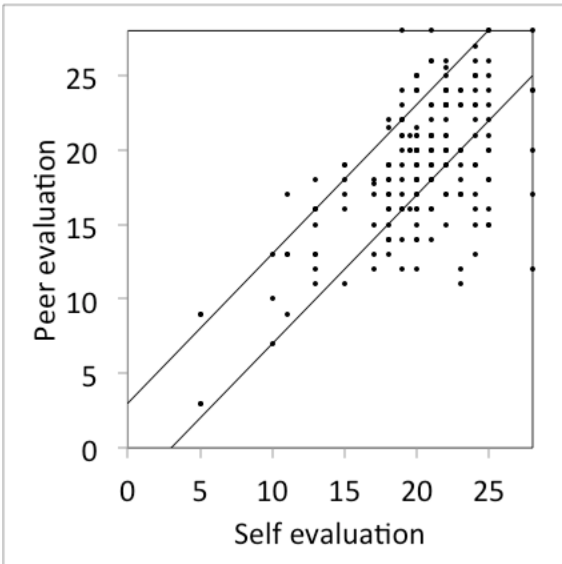

Figure 1. Comparison of peer evaluations with self evaluations of the kitchen chemistry report.

A small number of reports were either peer or self evaluated as perfect (28 out of 28 points). The diagonal lines in Figure 1 indicate peer and self evaluations that are consistent to within a score of ±3, or approximately 10% or one grade. A significant number of reports fall outside the lines, with a greater number below the lines. Reports in the lower-right triangle have self evaluations significantly higher than peer evaluations.

Using the statistic y = (*self evaluation score*) – (*peer evaluation score*), the Student's paired t-test ($\bar{y}$=1.1029, stdev=4.0097, n=192, df=191, t=3.8112) indicates that at a confidence level in excess of 99.9%, self evaluations have higher scores than peer evaluations. However while the difference is statistically significant, it is also sufficiently small that if these peer evaluations were real marks from academic staff, most (87%) of the self evaluations would be considered consistent with the peer evaluations under the University's assessment policy.

Figure 2 shows the academic staff markers' evaluations and students' self evaluations for the first and fifth laboratory reports after the kitchen chemistry exercise. There are fewer data points than in Figure 1 because some students had withdrawn from the unit during the summer while other students exercised the option of submitting team reports. (This unit had several non-traditional characteristics: for example, students were given the option of submitting individual or team laboratory reports, but that aspect is not discussed here.) As with Figure 1, there are some coinciding data points. Overall, there are less data for laboratory report 1 (left-hand panel of Figure 2) than laboratory 5 (right-hand panel), because many students did not submit a self evaluation for the first laboratory report.

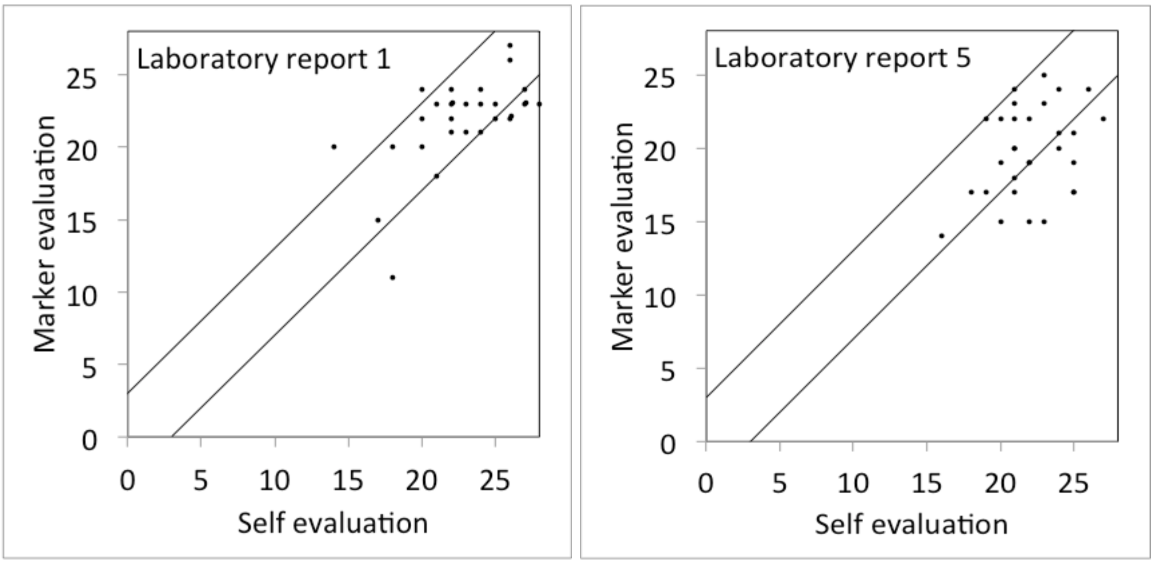

Figure 2. Comparison of academic staff evaluations with self evaluations of the first and fifth laboratory report.

As with the kitchen chemistry exercise, there is a tendency towards overly optimistic self evaluations. Using the statistic y = (*self evaluation score*) – (*academic staff evaluation score*), the Student's paired t-test for laboratory exercise 1 ($\bar{y}$=0.93103, stdev=2.8900, n=24, df=23, t=1.5782) indicates that self evaluations have higher scores than academic staff markers' evaluations at a confidence level of 85%, but there is no statistically significant difference at a confidence level of 90%. For laboratory exercise 5 ($\bar{y}$=2.4634, stdev=2.9419, n=41, df=40, t=5.3617), self evaluations have higher scores than academic staff markers' evaluations, at a confidence level in excess of 99.9%.

It is also clear that the overall distribution of scores moved to higher values: in Figure 1 there were a significant number of reports that were either self or peer evaluated as of poor quality. The data shown in Figure 2 for reports later in the unit indicates that there were *no* reports of poor technical quality. Significantly, Figure 2 has a greater proportion data points that are within the diagonal lines than Figure 1, indicating that students' self evaluations are fairly close to those of the markers. This indicates that the students have improved their metacognitive skill of evaluating the quality of their own reports.

Academic marker evaluations of the self and peer evaluation exercise are not available because the kitchen chemistry report was not marked or evaluated by an academic staff member. Hence there is no direct objective measure of whether the quality of technical writing changed. Instead, Figure 3 explores the students' self evaluations of their technical writing. Each data point represents one of the 37 students who completed both the kitchen chemistry report (Figure 1), and laboratory 5 (right-hand panel of Figure 2). Students' self evaluated technical writing skills generally improved during the summer as shown by the predominance of points towards the top of the graph. Two students who gave themselves perfect (28 out of 28) self evaluations at the start of the summer returned more realistic self evaluations at the end of the summer. Overall, the data in Figure 3 suggest that through a process of self evaluations of reports and feedback on those self evaluations, students improved their metacognitive self evaluation skills.

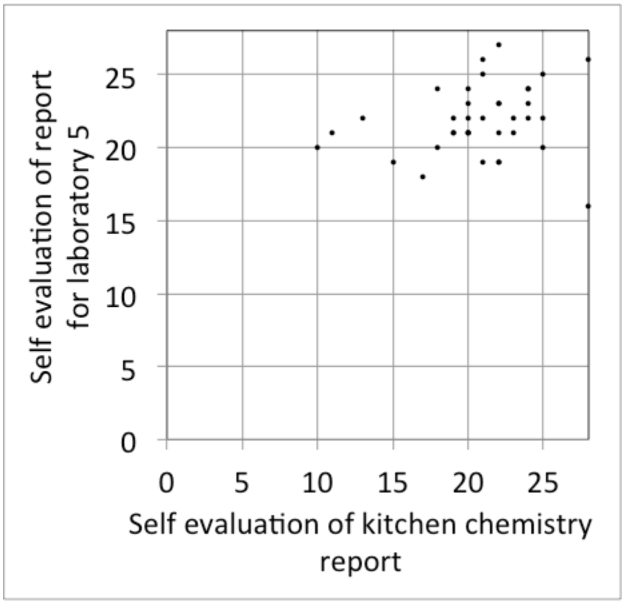

Figure 3. Comparison of self evaluations of the kitchen chemistry task (Figure 1) with self evaluations of the 5th laboratory report (Figure 2). Each data point represents one student who completed both the assignment task and the 5th subsequent laboratory report.

Comparing reports submitted in this level 2 summer unit with other level 2 units without a similar exercise gave the instructor the impression that this exercise improved the writing skills of students. However, no other evidence of the effectiveness of this self and peer evaluation exercise is available.

The university seeks feedback from students about every unit that is offered. The feedback for this unit contained very diverse views, with comments from some students in direct opposition to those from other students. Samples of feedback from the 2012-2013 cohort and from the 2014-2015 cohort are provided in Tables 6 and 7, respectively. There was some feedback that was very enthusiastic about the kitchen chemistry exercise, with students stating that they learnt a lot about report writing in this unit. There was also some extremely negative feedback that the exercise was too much work with insufficient guidance and that it was generally a waste of time. The instructor believes that the latter negative feedback was mainly from graduating students, who were seeking an easy summer unit to complete their degree. This belief is based on informal comments made in class and on the general attitudes of different groups of students. On the other hand, the positive feedback about the exercise suggests that it is worthwhile, albeit with further improvements and refinement.

Table 6. Typical free-text feedback to the questions "What were the best aspects of your unit?" and "What aspects of your unit were most in need of improvement?" from the 2012-2013 cohort. Responses were received from 18 students (45%), out of a total cohort of 40 students.

- It was good to learn how to actually write a good lab report. Helpful that the chemistry content wasn't too challenging, since the emphasis was on using this knowledge to write effective reports.
- The workload was manageable and there was always help given in the discussion section of [the learning management system].
- ... Due to the extreme range of abilities in our group, the lectures had to jump between extremely complex and very basic, but I found this really helpful.
- The unit was good in that it concentrated and taught how to write a proper lab report.
- ... The workload was very manageable - which i think is especially important, being a T3 (summer) subject.
- feedback about lab reports [needs improvement].

Table 7. Typical free-text feedback to the questions "What were the best aspects of your unit?" and "What aspects of your unit were most in need of improvement?" from the 2014-2015 cohort. Responses were received from 22 students (32%), out of a total cohort of 68 students.

- Marking each other's work and self evaluation was very helpful
- The most helpful part was lab report writing, with templates and writing manuals available on [the learning management system].
- There was no helpful aspect to this unit…
- I don't think there is anything helpful of the unit …
- No aspects of this unit were helpful. This unit suffers from a major design flaw in that being offered only every 2 years it tries to cater for too broad a level of abilities from 1st years through to 4th year.
- This unit is almost perfect…

The feedback in Tables 6 and 7 and the instructor's impressions indicate that this self and peer evaluation exercise needs further improvements and refinement. The data in Figures 1-3 is incomplete because some students did not do all parts of the exercise, or did not submit one or more of the subsequent laboratory reports. The data (especially Figure 3) could provide motivation for future students to fully engage with this exercise, and strongly suggest that the self and peer evaluation exercise has promising learning outcomes and is worth pursuing. However, the School has discontinued the kitchen chemistry experiment, because it and the subsequent five laboratory reports are considered to be too much work for students, especially in a summer unit. This is disappointing for the instructor, but will lead to a significant reduction in academic workload.

Conclusions

A double-blind peer evaluation has been combined with self evaluation of laboratory reports, in order to improve the metacognitive skill of evaluating the quality of their own reports, and thus to improve the writing skills of students. The summative assessment is that students are assessed on how closely their evaluations are consistent with peer evaluations of the same reports. The formative assessment is that students receive peer evaluations and feedback via a rubric on reports that they have written. Students then wrote further laboratory reports in the same unit of study. Self evaluation of laboratory reports tended to return higher scores than both peer evaluation and academic staff marker evaluations, indicating that students are not totally objective in self evaluation.

Comparisons of the peer and self evaluations at the start of the summer, with the academic staff marker and self evaluations at the end of the summer, indicate that students with poor writing skills had improved their technical writing skills, and that all students had improved their metacognitive self evaluation skills.

Acknowledgements

The work reported here has human research ethics approval (DUHR-EC-29-2008, and STEC-23-2012-LIM). The author acknowledges the contributions of the reviewers, whose critical comments have helped to make this a better paper.

References

Bailey, P. D. (2001). Teaching chemists to communicate? Not my job! *University Chemistry Education, 5*(2), 80-86.

Berry, D. E., & Fawkes, K. L. (2010). Constructing the components of a lab report using peer review. *Journal of Chemical Education, 87*(1), 57-61. doi: 10.1021/ed8000107

Blanck, J. F. (2007). Molecular gastronomy: Overview of a controversial food science discipline. *Journal of Agricultural & Food Information, 8*(3), 77-85. doi: 10.1300/J108v08n03_07

Blanksby, T., & Chan, C. K. (2006). Feedback assessment of Science report writing for first year Genetics students. In *Proceedings of the Uniserve Assessment in Science Teaching and Learning Symposium* (pp. 161-164). Sydney, NSW: University of Sydney.

Chickering, A. W., & Ehrmann, S. C. (1996). Implementing the seven principles: Technology as lever. *AAHE Bulletin, 49*(2), 3-6. Retrieved from http://www.case.edu/ucite/media/caseedu/ucite/documents/seven_principles.pdf

Fair, J. D., Kleist, E. M., & Stoy, D. M. (2014). A survey of industrial organic chemists: Understanding the chemical industry's needs of current bachelor-level graduates. *Journal of Chemical Education, 91*(12), 2084-2092. doi: 10.1021/ed400570f

Findlay, B. (1996). *How to write psychology laboratory reports and essays*. Sydney: Prentice Hall.

Gordon, N. R., Newton, T. A., Rhodes, G., Ricci, J. S., Stebbins, R. G., & Tracy, H. J. (2001). Writing and computing across the USM chemistry curriculum. *Journal of Chemical Education, 78*(1), 53-55. doi: 10.1021/ed078p53

GTI Media. (2015, May 8). The top 10 skills that'll get you a job when you graduate. *TargetJobs.* Retrieved from targetjobs.co.uk/careers-advice/career-planning/273051-the-top-10-skills-thatll-get-you-a-job-when-you-graduate

Hristova, Z. (2014). Using peers to assess oral presentations to foster learning. *International Journal of Innovation in Science and Mathematics Education, 22*(3), 74-80.

Jones, S., Yates, B., & Kelder, J.-A. (2011). *Science: Learning and Teaching Academic Standards Statement*. Strawberry Hills, NSW: Australian Learning and Teaching Council. Retrieved from http://www.olt.gov.au/system/files/altc_standards_SCIENCE_240811_v3.pdf

Kemmis, S., & McTaggart, R. (Eds.). (1988). *The action research planner* (3rd ed.). Geelong, Vic: Deakin University Press.

Kennepohl, D. (2007). Using home-laboratory kits to teach general chemistry. *Chemistry Education Research and Practice, 8*(3), 337-346. doi: 10.1039/B7RP90008B

Kennepohl, D., & Shaw, L. (Eds.). (2010). *Accessible elements: Teaching science online and at a distance*. Edmonton, Canada: Athabasca University Press.

Knight, P., & Yorke, M. (2003). *Learning, curriculum and employability in higher education*. London and New York: Routledge Falmer.

Le Blanc, B. (1992). *Chemistry in the kitchen*. Bath, UK: Cherrytree Press.

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues, 2*(4), 34–46. doi: 10.1111/j.1540-4560.1946.tb02295.x

Lewin, K. (1948). *Resolving social conflicts: Selected papers on group dynamics*. New York: Harper and Row.

Lim, K. F. (2003). *The Chemistry Style Manual* (2nd ed.). Geelong, Vic: Deakin University.

Lim, K. F. (2009). Doing it again, thoughtfully: Using feedback on draft reports to improve learning outcomes. *Australian Journal of Education in Chemistry, 70*, 11-16.
Lister, T., & Blumenthal, H. (2005). *Kitchen chemistry*. London, UK: Royal Society of Chemistry.
Lytle, S. L., & Cochran-Smith, M. (1990). Learning from teacher research: A working typology. *Teachers College Record, 92*(1), 83-104.
Moore, R. (1993). Does writing about science improve learning about science? *Journal of College Science Teaching, 22*, 212-217.
Palermo, M., Pellegrini, N., & Fogliano, V. (2014). The effect of cooking on phytochemical content in vegetables: A review. *Journal of the Science of Food and Agriculture, 94*(6), 1057-1070. doi: 10.1002/jsfa.6478
Paulson, D. R. (2001). Writing for chemists: Satisfying the CSU upper-division writing requirement. *Journal of Chemical Education, 78*(8), 1047-1049. doi: 10.1021/ed078p1047
Pellegrini, N., Chiavaro, E., Gardana, C., Mazzeo, T., Contino, D., Gallo, M., . . . Porrini, M. (2010). Effect of different cooking methods on color, phytochemical concentration, and antioxidant capacity of raw and frozen brassica vegetables. *Journal of Agricultural and Food Chemistry, 58*(7), 4310-4321. doi: 10.1021/jf904306r
Russell, A. A. (2005). Calibrated Peer Review™ A writing and critical-thinking instructional tool. In *Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics (STEM) Education*. Washington, DC: American Association for the Advancement of Science. Retrieved from http://www.aaas.org/sites/default/files/03_Suc_Peds_Russell.pdf
Russell, A. A. (2013). The Evolution of Calibrated Peer Review™ (Chapter 9). In T. Holme, M. M. Cooper & P. Varma-Nelson (Eds.), *Trajectories of Chemistry Education Innovation and Reform* (ACS Symposium Series 1145, pp. 129–143). Washington, DC: American Chemical Society. doi: 10.1021/bk-2013-1145.ch009
Silyn-Roberts, H. (1996). *Writing for Science: A practical handbook for science, engineering and technology students*. Auckland, New Zealand: Addison Wesley Longman.
The Australian Chamber of Commerce and Industry and the Business Council of Australia. (2002). *Employability Skills for the Future*. Canberra, ACT: Department of Education, Science and Training (Australian Commonwealth Government).
This, H. (2005). *Molecular gastronomy: Exploring the science of flavor (Arts & Traditions of the Table: Perspectives on Culinary History)* (M. Debevoise, Trans.). New York: Columbia University Press.
Tilstra, L. (2001). *Laboratory report guidelines*. Retrieved from http://www.rose-hulman.edu/~tilstra/laboratory_report_guidelines.htm
Tucci, V. K., O'Connor, A. R., & Bradley, L. M. (2014). A three-year chemistry seminar program focusing on career development skills. *Journal of Chemical Education, 91*(12), 2071-2077. doi: 10.1021/ed400667q
Walker, J. P., & Sampson, V. (2013). Argument-driven inquiry: Using the laboratory to improve undergraduates' science writing skills through meaningful science writing, peer-review, and revision. *Journal of Chemical Education, 90*(10), 1269-1274. doi: 10.1021/ed300656p
Whelan, R. J., & Zare, R. N. (2003). Teaching effective communication in a writing-intensive analytical chemistry course. *Journal of Chemical Education, 80*(8), 904-906. doi: 10.1021/ed080p904
Windsor, S. A. M., Rutter, K., McKay, D. B., & Meyers , N. (2014). Embedding graduate attributes at the inception of a chemistry major in a Bachelor of Science. *Journal of Chemical Education, 91*(12), 2078-2083. doi: 10.1021/ed5001526
Wood, L., Daly, A., Clark-Murphy, M., Kavanagh, M., Leveson, L., Dixon, P., . . . Rigby, B. (2010). *Graduate Skills Project.* Retrieved from http://graduateskills.edu.au/
Zuber-Skerritt, O. (1992). *Professional development in higher education: A theoretical framework for action research*. London, UK: Kogan Page, Ltd.

Appendix 1. Assessment rubric

My student number: [ ] Report code: [ ] - [ ] - [ ] - [ ] - [ ]

SLE235 Report Evaluation Rubric 2014
Assignment 1 only

The Title.

The Title is not informative and/or descriptive about the content of the report. 0	The Title is informative and descriptive about the content of the report. 1	

The Introduction section is a concise introductory statement of the novel features of the work; the object of the investigation with any essential historical background; and a justification for publication followed, if necessary, by a brief account of preliminary experimental work with relevant references. In a university laboratory or assignment report, do not state the teaching-and-learning aims: only include the scientific aims of the investigation.

- ☐ Accuracy
- ☐ Brevity
- ☐ Clarity
- ☐ Is there sufficient background to the problem or investigation
- ☐ Are there sufficient citations to scientific journals and/or books

The Introduction section is either lacking or so deficient that it is essentially lacking. 0	The Introduction section is severely deficient. 1	The Introduction section requires more information to explain background to the problem or investigation. 2	The Introduction section sets out the background to the problem or investigation. It has an appropriate length and style for a general reader. 3	The Introduction section is too long and/or too technical. Some of the could be moved to the Discussion section 2	

The Method section is a description of the experimental (or computational) procedures. Working details must be given concisely, with sufficient detail for an informed reader to reproduce the experiment or computation: well-known operations should not be described in detail. Suppliers of equipment and materials, and their locations, should be mentioned.

- ☐ Accuracy
- ☐ Brevity
- ☐ Clarity
- ☐ Is the Method section an account of what has been done (good), or is it a set of instructions for someone else to do (poor)?
- ☐ Is the Method section written in sentences and paragraphs (good), or is it a set of dot points (poor)?
- ☐ Have all essential masses, volumes and other details been incorporated into the body of the Method section?
- ☐ Has the make and model of specialist equipment been incorporated into the body of the Method section?

The Method section is either lacking or so deficient that it is essentially lacking. 0	The Method section is severely deficient with regard to the details of the procedure. 2	The Method section requires more details of the procedures. 3	The Method section is very good in terms of completeness, preciseness and conciseness: it includes the details of the procedures. 5	The Method section is too long. Description of the procedures should more concise by omitting details that a specialist reader would know, or by summarising the procedures to a greater extent. 3	

The Results section would include information such as what you observed and what you achieved. Numerical results are often (not always!) best presented in tabular or diagrammatic form (but not both for the same results).

- ☐ Accuracy
- ☐ Brevity
- ☐ Clarity
- ☐ Has there been appropriate use of tables and graphs to summarise and present results?
- ☐ Is the text of the Results section written in sentences and paragraphs?
- ☐ Are all relevant data, measurements and observations in the Results section (good), or are some relevant results in the appendices (poor)?
- ☐ Does the text of the Results section inform the reader what key results are in the tables and graphs?

The Results section is either lacking or so deficient that it is essentially lacking 0	The Results section is severely deficient with regard to the observations, measurements and/or results. 2	The Results section is clear and (too) concise. However, it lacks completeness and/or precision. It requires more details of the observations and results. 3	The Results section is complete and precise. However, it lacks clarity and/or conciseness. The data should be summarised more, perhaps in tabular or diagrammatic form. 3	The Results section sets out the observations and results of the various procedures in a clear, complete, precise and concise manner. 5	The Results section is too long. It should be more concise by better use of appropriate use of tables and graphs. 3	

The Discussion section should comment on the scope of the method and its validity, or the generality of the synthesis or computation. Where appropriate, there should be a discussion of the similarities and differences between the results presented in the report and any similar results previously published in the literature (textbooks and journals). The Discussion section should include a statement of any conclusions drawn from the work.

- ☐ Accuracy
- ☐ Brevity
- ☐ Clarity
- ☐ Is the Discussion section written in sentences and paragraphs?
- ☐ Is the scientific explanation clear and easy to understand?
- ☐ Are there sufficient citations to scientific journals and/or books

The Discussion section is either lacking or so deficient that it is essentially lacking 0	The Discussion section is severely deficient with regard to the meaning and significance of the results. 2	The Discussion section is clear and (too) concise. However, it lacks completeness and/or precision. It requires more commentary on the scientific meaning and significance of the results. 4	The Discussion section is complete and precise. However, it lacks clarity and/or conciseness. Possibly, a greater focus on similarities, patterns and differences will make the Discussion section clearer and/or more concise. 4	The Discussion section sets out the scientific meaning and significance of the results in a clear, complete, precise and concise manner. 6	The Discussion section is too long. It should be more concise by omitting irrelevant or unimportant discussion. 4	

The (summary or) Conclusion section should not simply duplicate statements in the discussion, but present an overview of the whole study, with a summary of the main findings and conclusions of the study.

☐ Accuracy
☐ Brevity
☐ Clarity

☐ Is the Conclusion section written in sentences and paragraphs?
☐ If you only read the Conclusion section, is it a clear and complete summary of all the main points in the entire report?

The Conclusion section is either lacking or so deficient that it is essentially lacking. 0	The Conclusion section is severely deficient. 1	The Conclusion section requires more information on the main findings and conclusions. 2	The Conclusion section sets out summary of the main findings and conclusions of the study. It has an appropriate length and style for a general reader. 3	The Conclusion section is too long and/or too technical. The main findings and conclusions should be summarised more and/or the less important details omitted. 2	

The References section should be a full list of the books and journal papers that you have cited in your report. References should not include the laboratory manual or lecture notes – you should refer to the books and papers on which the laboratory manual or lecture notes are based.

☐ Accuracy
☐ Brevity
☐ Clarity

☐ Do the in-text citations follow a recognised and approved style?
☐ Do the references list all essential details?
☐ Do the references consistently follow a recognised and approved style, which is consistent with the in-text citations?

The citations in the body of the report and/or the References section are either lacking or so deficient that it is essentially lacking. 0	The citations and/or the References follow an appropriate and/or consistent style, but are missing many essential details. 1	The citations and/or the References are complete in the bibliographic details, but do not follow an appropriate and/or consistent style. 1	The citations and the References are complete in the bibliographic details, and follow an appropriate and consistent style. 2	

The report should use appropriate language and conventions to communicate with the reader.

☐ Is there correct and consistent use of scientific names and units (eg mass, not weight)?
☐ Is there correct and consistent use of SI units?
☐ Is there correct and consistent use of superscripts and subscripts?
☐ Is there correct and consistent use of multiplication signs and other mathematical and scientific symbols?
☐ Is there consistent use of correct grammar?
☐ Is there consistent use of correct spelling?

There is little or no use of scientific language and conventions 0	There is some (but insufficient and/or inconsistent) use of scientific language and conventions. 1	There is mostly good use of scientific language and conventions. 2	There is appropriate and consistent use of scientific language and conventions. 3	
			Technical report writing sub-total (out of 28)	

The report, especially the Results and Discussion sections, should demonstrate the writer's knowledge, skills and understanding.

☐ Is there use of a control experiment, against which the other test results can be compared? The control should give a negative result.

There is no or inadequate use of a control experiment. 0	There is adequate use of a control experiment. 1	

☐ Can the chosen quantities or properties be reliably and objectively observed or measured?

The chosen quantities or properties cannot be reliably and objectively observed or measured. 0	It is possible, but difficult, to reliably and objectively observe or measure the chosen quantities or properties. 1	The chosen quantities or properties can be reliably and objectively observed or measured. 2	

☐ Is the experiment designed so that someone one can relatively easily repeat the experiment with the same or very similar samples?
☐ Does the experiment have statistical validity? Either there are multiple similar items (eg peas) or multiple repeats of the experiment if each trial has a small number of items.

The experiment would be difficult or impossible to repeat. The experiment lacks statistical validity. 0	The experiment is repeatable, but lacks statistical validity. 1	The experiment is repeatable, and has some (but inadequate) statistical validity. 2	The experiment is repeatable, and has statistical validity. 3	

☐ Does the report, as a whole, have correct and relevant scientific concepts?
☐ Does the report, as a whole, apply correct and relevant scientific concepts to explain the actual results (not ideal results)?
☐ Does the report, as a whole, give reasons/evidence to link the results to the claims, findings or conclusions?
☐ Does the discussion resolve any apparent contradictions or problems?
☐ Is the report, as a whole, free of scientific inconsistencies and scientific contradictions?
☐ Is the scientific reasoning valid and free of error?

The report demonstrates little or no knowledge, and/or has little or no application of knowledge to the interpretation of the data. 0	The report has some use of scientific concepts as reasons/evidence to link the results to the claims, findings or conclusions, but there are major inadequacies either in the scientific concepts or in the scientific reasoning. 2	The report, as a whole, has good use of scientific concepts as reasons/evidence to link the results to the claims, findings or conclusions, but there are minor inadequacies either in the scientific concepts or in the scientific reasoning. 4	The report, as a whole, uses correct and relevant scientific concepts as reasons/evidence to link the results to the claims, findings or conclusions. 6	
			Report total (out of 40)	