Investigating the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments in undergraduate chemistry laboratories.

Stephen Robert Daniel George-Williams

Bachelor of Advanced Science (Honours)
Doctor of Philosophy (Chemistry)

A thesis submitted for the degree of *Doctor of Philosophy (Chemistry Education)* at
Monash University in 2019
School of Chemistry, Faculty of Science

Copyright notice

Abstract

The use of inquiry-based learning in the laboratory to develop higher order thinking skills (such as critical thinking) is well reported in the literature. Context-based learning is increasingly popular due to the notable increases in student engagement and performance. Problem-based learning combines the best of each of these pedagogies and can be utilised in the laboratory to enhance students' ability to deal with uncertainty in both their results and in choosing appropriate methodologies from a range of possibilities. Utilisation of an industry or workforce context have been shown to be particularly impactful. Whilst many individual examples of these approaches have been published, few studies have investigated their use on a larger scale than a single unit or a longer timeline than one or two semesters.

Transforming Laboratory Learning (TLL) was a programme at Monash University that ran from early 2016 to early 2019. This programme sought to modernise the undergraduate laboratory activities delivered in the School of Chemistry over all year levels. The desired outcomes matched those delivered by inquiry-, context-, and problem-based learning. This entailed moving away from expository activities and increasing the contextualisation of the laboratory activities and the subsequent assessment tasks where appropriate. Furthermore, efforts were undertaken to tackle issues in marking variation and feedback processes.

The impacts of these changes on the perceptions of students was monitored using surveys after the completion of individual laboratories or on an annual basis alongside focus groups at the completion of units significantly redesigned by the TLL programme. Staff perceptions of the new activities and assessment tasks were also investigated.

The results indicated that students were highly engaged with the new laboratory activities and were able to articulate a wider range of developed skills, particularly those associated with scientific methodology. These findings were evident immediately after the completion of individual activities but also after the completion of units. However, when considered on an annual scale, it became apparent that the continued presence of traditional laboratory activities running alongside the new activities undermined the potential benefits of the large-scale inclusion of inquiry/problem/context/industry-based experiments. Very few changes were noted in the survey responses of students when they were invited to consider the entirety of their experiences in the 2nd or 3rd year of the programme. Preliminary results suggested that the use of authentic assessment aided in student recognition of skill development. The use of electronic marking criteria helped reduce marking variation between assessors due to the removal of academic judgment from the marking process. Overall, the large-scale use of these new activities was considered impactful but further changes, particularly in the 3rd year of the program, would be required to effect meaningful large-scale changes to student perception.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature [redacted]

Print Name: Stephen Robert Daniel George-Williams

Date: 15/2/19

Publications during enrolment

George-Williams, S. R., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2019). Inquiry-, problem-, context- and industry-based laboratories: An investigation into the impact of large-scale, longitudinal redevelopment on student perceptions of teaching laboratories. International Journal of Science Education, Awaiting review

George-Williams, Caroll, M.-R., Ziebell, Thompson, C. D. & Overton, T. L. (2018). Curtailing marking variation and enhancing feedback in large-scale undergraduate chemistry courses through reducing academic judgement: A case study. Assessment and Evaluation in Higher Education, Awaiting page numbers. doi: 10.1080/02602938.2018.1545897

George-Williams, S. R., Karis, D. Ziebell, A. L., Kitson, R. R. A., Coppo, P., Thompson, C. D. & Overton, T. L. (2018). Investigating student and staff perceptions of students' experiences in teaching laboratories through the lens of meaningful learning. Chemistry Education Research and Practice, 20, 187-196. doi: 10.1039/C8RP00188J

George-Williams, S. R.; Ziebell, A. L.; Thompson, C. D. & Overton, T. L. (2018). Electronic Waste – A case study in attempting to reduce cognitive overload whilst increasing context and student control in an undergraduate laboratory. Monash Education Academy Digest, Awaiting review.

George-Williams, S. R.; Ziebell, A. L.; Thompson, C. D. & Overton, T. L. (2018). Enhancing inquiry in the teaching laboratory; a consecutive expository/inquiry-based laboratory exercise. Monash Education Academy Digest, Accepted, awaiting publication.

George-Williams, S. R., Soo, J. T., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2018). Inquiry and industry inspired laboratories: the impact on students' perceptions of skill development and engagements. Chemistry Education Research and Practice, 19(2), 583-596. doi: 10.1039/C7RP00233E

George-Williams, S. R., Ziebell, A. L., Kitson, R. R. A., Coppo, P., Thompson, C. D., & Overton, T. L. (2018). 'What do you think the aims of doing a practical chemistry course are?' A comparison of the views of students and teaching staff across three universities. Chemistry Education Research and Practice, 19(2), 463-473. doi: 10.1039/C7RP00177K

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes (*5)* original papers published in peer reviewed journals and (*2)* submitted publications. The core theme of the thesis is the impact of the large-scale inclusion of context-, inquiry-, industry- and problem-based teaching laboratories on the perceptions of students. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the School of Chemistry under the supervision of Professor Tina Overton.

(The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.)

In the case of chapters three to seven my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status (published, in press, accepted or returned for revision, submitted)	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution*	Co-author(s), Monash student Y/N*
2	'What do you think the aims of doing a practical chemistry course are?' A comparison of the views of students and teaching staff across three universities.	Published	65%. Data collection, analysis and primary author.	Angela Ziebell, 10% input into analysis and manuscript.	N
				Russel R A Kitson, 2.5%, grant acquisition	N
				Paolo Coppo, 2.5%, grant acquisition	N
				Christopher Thompson, 5%, edit of manuscript.	N
				Tina Overton, 15%, input into analysis and manuscript.	N
3	Investigating student and staff perceptions of students' experiences in teaching laboratories through the lens of meaningful learning.	Published	60%. Data collection, analysis and primary author.	Angela Ziebell, 5%, input into manuscript.	N
				Russel R A Kitson, 2.5%, grant acquisition.	N
				Paolo Coppo, 2.5%, grant acquisition.	N
				Dimitri Karis, 7.5%, data collection.	N
				Siegbert Schmidt, 5%, edit of manuscript.	N
				Christopher Thompson, 5%, edit of manuscript.	N
				Tina Overton, 12.5%, input into analysis and manuscript.	N
4	Enhancing inquiry in the teaching laboratory; a consecutive expository/inquiry-based laboratory exercise.	Published	60%. Data collection, analysis and primary author.	Angela Ziebell, 20%, generated the new exercise.	N
				Chris Thompson, 5%, edit of manuscript.	N
				Tina Overton, 15%, input into analysis and manuscript.	N

4	Electronic Waste – A case study in attempting to reduce cognitive overload whilst increasing context and student control in an undergraduate laboratory.	Submitted	60%. Data collection, analysis and primary author.	Angela Ziebell, 20%, generated the new exercise.	N
				Chris Thompson, 5%, edit of manuscript.	N
				Tina Overton, 15%, input into analysis and manuscript	N
5	Inquiry and industry inspired laboratories: The impact on students' perceptions of skill development and engagement.	Published	60%. Data collection, analysis and primary author.	Angela Ziebell, 15%, generated most of the new exercises.	N
				Jue Soo, 10%, generated two of the new exercises.	Y
				Chris Thompson, 5%, edit of manuscript.	N
				Tina Overton, 10%, input into analysis and manuscript.	N
6	Inquiry-, problem-, context- and industry-based laboratories: An investigation into the impact of large-scale, longitudinal redevelopment on student perceptions of teaching laboratories.	Submitted	60%. Data collection, analysis and primary author.	Angela Ziebell, 20%, generated most of the new exercises.	N
				Chris Thompson, 5%, edit of manuscript.	N
				Tina Overton, 15%, input into analysis and manuscript.	N
7	Curtailing marking variation and enhancing feedback in large-scale undergraduate chemistry courses through reducing academic judgement: A case study.	Published	60%. Data collection, analysis and primary author.	Mary-Rose Carroll, 15%, aid in mass development of spreadsheets and Moodle reports.	N
				Angela Ziebell, 12.5%, aid in development of spreadsheets and edit of manuscript.	N
				Chris Thompson, 5%, edit of manuscript.	N
				Tina Overton, 7.5%, edit of manuscript.	N

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: [redacted] **Date: 15/2/19**

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature: [redacted] **Date: 15/2/19**

Acknowledgements

Hahahahahahahahahahahahahaha! Well, this was a twist I for one didn't see coming. Most people finish a PhD in their given field and generally refer to the experience as one of the toughest of their lives. They celebrate its completion and vow to take vengeance on whatever scientific conundrum eluded them for several years. Apparently, I decided that that path was for quitters.

Making the decision to undertake another PhD was certainly a unique concept. Most people struggled to understand how anyone could be so masochistic while others simply wondered what would happen to my titles (Dr George-Williams, PhD perhaps?). Whilst important questions, neither was particularly relevant to the why of it all – to become employable and to follow my passion for teaching.

With regards to teaching, I would like to give a significant shout out to Mary-Rose and the technical staff at Monash University. It was always great fun to teach into the laboratories (not labs – happy Tina?!) and I was eternally grateful for the funds and continuing teaching experience. Next would have to be Kellie Vanderkruk, whose kindness and support enabled me to break into the undergraduate chemistry tutorials – a source of great experience and wealth.

With regards to employability, anyone and everyone involved in my research group needs to be thanked immensely. Michelle, Kathy, Jared, Stevie D, Mahbub, Jue, Megan, Elizabeth, Camilla and Wei – your support was very much appreciated alongside our constant debates on whether or not a given statistical test was truly okay to use. Furthermore, you managed to embrace my unique sense of humour and, at the very least, appeared to enjoy our shared company. I know I certainly did and I will miss you all dearly.

Of course, I can't go any further without raising Chris and Tina:

Chris, whilst sighting you was just as likely as encountering the Loch Ness monster, your continuous encouragement and support did not go unnoticed. Thank you for the generosity of your kindness and I look forward to working with you in the future (even at separate institutions!).

Tina, you're an absolute legend! When I joined the research group in 2016 I had absolutely no idea the incredible researcher/European Rockstar that I was working with. Your prestige and incredible talent, alongside your unwavering determination, support and almost indescribable warmth and affection, have made this PhD a truly amazing experience. Every success I have had over the last few years can 100% be traced back to you. I cannot begin to describe the magnitude of my gratitude to you and everything that you have done. Thank you from the very core of my heart.

Before I wrap up this list of acknowledgments, I have one last individual to cover – ~~my cat, Link~~ my wonderful husband, Jamie. It's truly difficult to describe what it has meant to me to know that whenever I was coming home, I was coming home to you. Whilst you're undeniably an utter loon who refuses to clean even the most basic of things, you fill my life which constant joy, laughter and love. Just by being you, you made this experience infinitely more doable and I am eternally grateful to have you around. Thank you, you nut case.

Contents

Chapter 1 - Introduction

Transforming Laboratory Learning (TLL) was a programme at Monash University that ran from February 2016 to February 2019. Its overall goal was to use the laboratory programme to address a perceived lack in the employability skills of students who studied chemistry at Monash University. This was highlighted by the results of Sarkar, Overton, Thompson, and Rayner (2016), which showed that, whilst graduate employers were relatively pleased with graduating students' scientific knowledge, their technical, planning and interpersonal skills were often seen to be lacking. Furthermore, this problem formed a central concern of the 2016 Grattan report 'Mapping Australian Higher Education' (Norton & Cakitaki, 2016) showing that this issue appeared to affect all tertiary institutions in Australia, across all science disciplines.

The central premise of the Transforming Laboratory Learning programme was that the issue of student employability could be potentially addressed through the undergraduate laboratory programme. This proposal required a consideration of the current aims of teaching laboratory activities to be able to amend them in such a manner as to combat the issue of skills deficits. The role of laboratory instruction in teaching science has been a topic of discussion for well over a century. Laboratory classes have been operating since the mid-1800s and early 1900s in the traditional science disciplines of physics (Phillips, 1981), chemistry (Bowers, 1924) and biology (G. L. Miller, 1992). As stated by Shah (2007) – '*Today, it would be rare to find any science course without an element of project work*'. Commonly, the aims of laboratory exercises (ICSU-CTS, 1979; A. H. Johnstone & Al-Shuaili, 2001; P. A. Kirschner & Meester, 1988; Linn, 1997; Reid & Shah, 2007; Shymansky & Penick, 1979) can be considered as:

1) A chance for students to learn science in a more tactile, engaging way.
2) Complementing underlying scientific theory.
3) Developing technical skills.
4) Imparting scientific methodology.
5) Enhancing transferable/soft skills (communication, time management, etc.).

1.1 Arguments for and against the use of laboratory activities

Not all academics agree with the necessity of laboratory classes for student learning. Specifically discussing chemistry, Hawkes (2004) raised a commonly overlooked argument about the lack of need for manipulative skills for non-chemistry majoring students who are forced to take a first-year chemistry practical chemistry course. Hawkes' argument was that when one considers the cost of running an undergraduate course, students who aren't majoring in chemistry will often be better served by encountering additional content in lectures or an increased number of tutorial hours. An investigation undertaken by Sundberg and Armstrong (1993) into the state of biology laboratory programmes highlighted that many American teaching laboratories were forced to run on a budget of $10/student (approx. $10000 for a typical first-year class of 1000 students). If one considers the cost of consumables, facilities, room bookings and staff time, this was a very tight budget with which to operate an impactful teaching laboratory programme.

The issue of financial concerns is further complicated by the common observation raised by Hawkes that students often do not successfully learn the desired content in these traditional laboratory exercises (presuming that that is an aim of a given teaching laboratory). Indeed, a large analysis of physics laboratories by Holmes and Wieman (2018) of student performance in exams revealed that laboratories specifically designed to teach content had no measurable effect on the students' final score in subsequent assessment. This finding was shown to be consistent across three institutions, nine units, seven teaching staff and approximately 3000 students. Additional focus groups performed with the same students highlighted that the only thinking that students felt they undertook during these laboratories '*was in analysing data and checking whether it was feasible to finish the lab in time*'. It is important to note, however, that unlike Hawkes, Holmes and Wieman do not argue for the removal of teaching laboratories but rather a shift in focus to allow students more decision-making opportunities within a laboratory activity.

A very similar argument can be found in the chemistry education literature, through the work of Seery, Agustian, and Zhang (2018). In this study, the researchers agree with the argument put forward by Woolnough and Allsop (1985) that educators should '*stop using practical work as a subservient strategy for teaching scientific concepts and knowledge*'. Serry, Agustian and Zhang further propose an overall framework for teaching laboratories that first begins by focusing on explicit instruction of practical skills before becoming more student-directed in subsequent activities and later year levels. Once again, the argument was not for the removal of teaching laboratories, but rather a significant overhaul to focus on explicit technique development, enhancement of scientific thinking skills and an appreciation of methodological development.

Wilcox and Lewandowski (2017) collected survey data that appeared to show negative outcomes from traditional concept-focused laboratories. The researchers measured the responses of 4915 students over 67 institutions and found that students who completed concept-focused laboratory activities became less likely to respond in an 'expert like' manner – i.e. they lost beliefs and perceptions more akin to practicing scientists because of the concept-focused laboratory activities. Students who undertook more skills-focused laboratory activities not only responded in a more 'expert like' manner but were also less likely to be affected by their gender identity. It was shown that a gap in the responses of male- and female-identifying students (in favour of male-identifying students) was eliminated by the conclusion of more skill-focused courses. This further cements the argument that laboratory activities are still vital experiences, but the way in which they are delivered may need to shift.

Currently, many (if not most) teaching laboratories around the world utilise expository (or cookbook/recipe) experiments, which, as Domin (2007) states, rely '*almost exclusively on laboratory manuals to create a situation where students perform the activity by following a prescribed procedure to experience a pre-determined outcome*'. These experiences are often criticised for invoking little critical thought (Hodson, 1990), and often students are left with little to no understanding of the underlying science (Letton, 1987). Whilst these exercises are often well laid out processes, they are also thought to '*encourage cheating or copying rather than thought or effort*' (Carnduff & Reid, 2003). It should also be noted that upon questioning, students often cannot recall the aim of a laboratory activity immediately after completion (Kirschner & Meester, 1988). With these concerns in mind, it would appear that expository laboratory exercises are not succeeding at achieving many of the previously mentioned aims of the laboratory sessions.

1.2 Student perceptions of the aims of laboratory activities

How students perceive the aims or goals of teaching laboratories has been investigated multiple times in the literature. Boud, Dunn, Kennedy, and Thorley (1980) used a survey of closed questions with a Likert scale and found that students tended to rate the development of practical skills and enhancement of theoretical knowledge above the development of problem-solving skills or scientific methodology. Graduates and practicing scientists (i.e. those employed to undertake science in industry or universities) focused more on developing observation skills and critical awareness. Clearly, there was a mismatch between student views and practicing scientists.

A series of 13 interviews were undertaken in 2008 by Russell et al. to further investigate the viewpoints of students towards teaching laboratories (Russell & Weaver, 2008). The responses of the students indicated that they were predominately focused on finishing the class to obtain marks to the exclusion of other goals (e.g. theoretical learning or skill development). A similar study was also undertaken in 2015 (DeKorver & Towns, 2015) by Towns et al., wherein video recordings were obtained showing students focusing on simply finishing the task as quickly as possible. This result is also common in secondary schools (Lynch & Ndyetabura, 1983; Wilkinson & Ward, 1997), indicating similar mismatches between students and teaching staff around the aims of teaching laboratories.

The expectations of students towards what they think, feel and do in a teaching laboratory (i.e. the cognitive, affective and psychomotor domains, respectively) has also been recently studied. In particular, the work of Bretz *et al.* (Galloway & Bretz, 2015a, 2015b; Galloway, Malakpa, & Bretz, 2016), utilised Novak's theory of meaningful learning (i.e. that, to learn, one must align the three aforementioned domains) to generate the Meaningful Learning in the Laboratory Instrument (MLLI). This survey consisted of 31 closed questions and specifically focused on how student expectations towards their own thoughts and feelings could change over a given semester. The responses of students to the survey were then split into three categories, questions that related to the cognitive domain, questions that related to the affective domain and questions that overlapped both the affective and cognitive domains. It was noted that students generally started with positive expectations towards the laboratory activities but became more negative when asked to reflect back on a set of laboratory activities that they had just completed in a given unit. These results indicated that a) students tended to feel that the laboratory activities did not fully meet their expectations and b) the MLLI survey was able to detect and measure the variation in student expectations towards how they would think and feel during laboratory activities.

1.3 Academic perceptions of the aims of laboratory activities

The perceptions of academic staff towards laboratory aims has already been investigated predominately in the US (by Bruck and Towns et. al.), either through interviews (Bretz, Fay, Bruck, & Towns, 2013; L. B. Bruck, Towns, & Bretz, 2010) or the Faculty Goals Survey (Bretz, Galloway, Orzel, & Gross, 2016; A. D. Bruck & Towns, 2013). The interviews indicated that teaching staff perceived laboratory aims as the development of transferable skills (such as teamwork, independence, critical thinking and scientific communication), the enhancement of theoretical understanding or the mastery of practical skills. The survey indicated that the aims raised appeared to change depending on the course being discussed or even the year level. Therefore, it would appear that the laboratory aims raised by teaching staff were changeable and dependent on context. There are few studies that directly assess or measure the perceptions of students and teaching staff of the aims of teaching laboratories on a large-scale beyond the US context.

1.4 Inquiry-based learning.

Inquiry-based learning (IBL) is a pedagogical approach that encapsulates any learning activity in which students are encouraged to ask their own questions and seek their own answers. It generally involves posing questions, problems or scenarios to students that they are expected to answer for themselves rather than simply following a list of instructions. Inquiry-based experiments attempt to create this student-focused, question-driven environment either through removing knowledge of the final answer, the data analysis procedures or even the individual steps normally provided in the laboratory manual (Cummins, Green, & Elliott, 2004). Hence, inquiry-based learning is a broad term that can vary greatly depending on the degree of control given to the student (Banchi & Bell, 2008). This can be seen in the model of Buck, Bretz, and Towns (2008), which was a variation of the original work of Herron (1971), which consists of four main levels of inquiry. Figure 1 depicts the four levels (Level 0 – 3) and is followed by a more detailed description with respect to laboratories in Table 1.

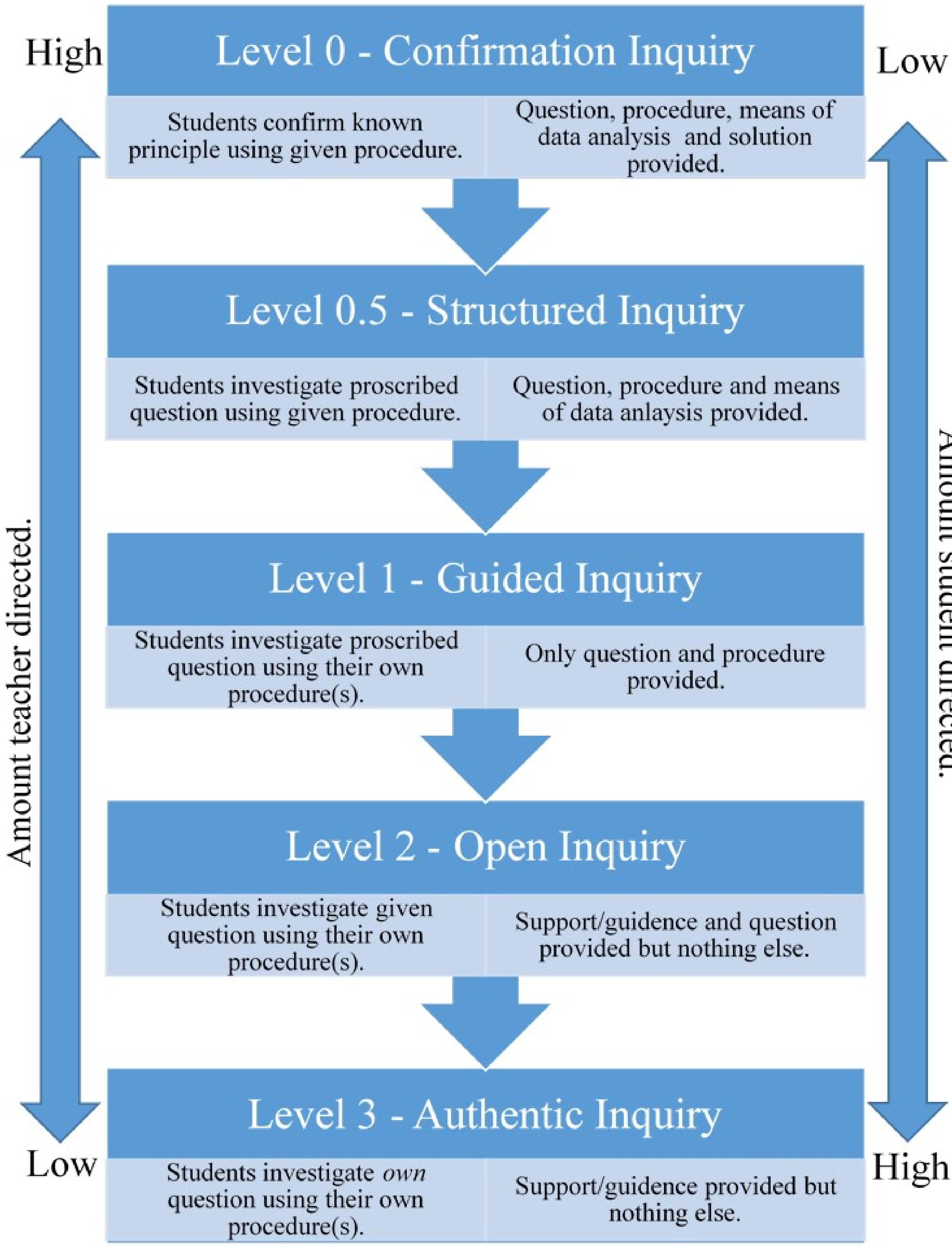


Figure 1 - The four levels of inquiry as per Buck, Bretz and Towns (2008).

Table 1 - Detailed descriptions of the levels of inquiry proposed by Buck, Bretz and Towns (2008).

Level	Description
0 - Confirmation	These are experiences where students only confirm an existing principle or theory using a known method and a provided means of data analysis, just like in an expository laboratory.
0.5 - Structured	Experiments that utilise structured inquiry are similar to confirmation experiments, only with the knowledge of the final answer/solution removed, allowing the student to discover the answer for themselves. Note that the means to analyse the data is still prescribed to the students.
1 - Guided	Similar to Structured Inquiry but the means of data analysis is either removed or provided to students through a guided discussion with teaching staff.
2 - Open	In these instances, the students are no longer following a prescribed method and neither the knowledge of the final answer/solution nor the means of data analysis are provided to the students. However, possible paths of investigation are provided to the students in order to still direct or guide the experience.
3 - Authentic	In the highest tier of inquiry, nothing is provided to the students save a general topic and what equipment/chemicals are available for use in their experiments. They are expected to devise their own topic and hypothesis, provide a methodology and discover the unknown answer on their own.

As previously mentioned, confirmation laboratories are incredibly common in many tertiary institutions throughout the world. A common example is where students titrate a weak acid with a strong base and compare their results to the provided expected concentrations. These laboratories typically require little from students in terms of engagement or input, resulting in a potentially non-impactful experience.

In contrast, structured and guided inquiry experiences can offer students an opportunity to answer questions through their own means. For example, Hulien et al. (2015) described a structured experience in which students identify a range of common household solids through the use of X-ray powder diffraction. The identity of the products is unknown and students are expected to research common household products in order to arrive at the correct answer. It was noted that students were highly engaged and motivated as well as also becoming much more confident with the use of the X-ray powder diffraction technique as students '*now use this technique more frequently in settings where they are asked to choose for themselves how to analyse samples*'.

A higher level guided example can be considered through the work of Kulevich, Herrick, and Mills (2014) who described an undergraduate laboratory wherein students discovered the factors that control buffers, up to and including the mathematical laws that govern them. Specifically, students were tasked with investigating the properties of weak acid/base solutions and were only given limited guidance on the analysis of their data. It was noted that the students '*confidence with preparing buffered systems improve[d]*' and it was believed that students obtained a strong '*understanding of topics related to acid−base equilibrium*'.

An example of the next level of inquiry can be considered through the more open experiment devised by Schepmann and Mynderse (2010). In this case, students optimise the use of the Grubb's catalyst through altering a range of reaction conditions (that they choose from) and student derived methodologies. As such, the degree of student control is much higher than in the previous examples. This experiment showed students increased their ability to undertake experimental planning and were highly engaged with the process, referring to it as '*important and worthwhile*'. However, it should be noted that students also started to raise concerns about the uncertainty of the final answers.

Lastly, an authentic inquiry example can be noted in the experiment described by Stout (2016) in which students were tasked to '*Find an interesting chemical, biological, or environmental issue involving carbon dioxide. Devise a researchable question relative to it and an experiment to provide the data necessary to answer [their] question*'. The results indicated much greater gains with respect to independent thinking, an awareness of scientific methodology, and was noted to result in students who performed better in later research environments, even in other units. Clearly, this highly inquiry-based activity resulted in a large number of positive outcomes, especially with regards to broader skill development.

In a recent meta-analysis of inquiry-based activities, Lazonder and Harmsen (2016) investigated the effects of guidance in 72 different studies. It was found that a greater use of guidance lead to '*a more* *proficient use of inquiry skills*', better student performance (by approximately half a standard deviation) and was more likely to result in the successful completion of desired learning outcomes (but only if the learning outcome was inquiry-focused rather than content-focused). This is consistent with the laboratory examples discussed above and potentially indicates the generalisability of the use of inquiry-based learning. It is important to note that recent work highlights that students prefer the more guided experiences as compared to the open-inquiry laboratories (Chatterjee, Williamson, McCann, & Peck, 2009) but concerns have been raised about deferring to student preference as they tend to choose tasks that require less effort (Harris, 1997) rather than those that result in the highest potential learning gains.

Another example of inquiry-based learning considered on a large scale can be seen through the work of Healy and Jenkins (2009). This report, generated for the Higher Education Academy in the UK, summarises many case studies in which inquiry-based learning was achieved using research focused activities. Of interest is the argument put forward that the inclusion of inquiry- or research- based practices are needed as students 'face an uncertain employment market' and need help learning to 'cope with uncertainty, ambiguity, complexity and change'. They also noted the large number of small-scale interventions (i.e. only included the top 5-10% of students) with few cases where inquiry- or research- based activities were disseminated to most of the undergraduate cohort.

It is interesting to note that of the examples shown, several of the laboratory activities tended to have strong apparent connections to the students' daily lives (e.g. common household solids/drugs or the effects of carbon dioxide in the real world) rather than simply focusing on complex theoretical concepts. The use of these connections to potentially enhance student engagement will be discussed in the next section.

1.5 Context-based learning.

Context-based learning (CBL) refers to any learning experience in which the material has been connected to the students' personal experiences or to the real world (Pilot & Bulte, 2006). However, there are many different forms of context-based learning which are dependent upon the ways in which the learning materials are delivered. A detailed discussion of this can be found in the work of Gilbert (2006) who described the use of context-based learning through four models:

1) '*Context as the direct application of concepts*'. In this model, contexts are added after the students have already been taught an underlying theoretical concept (e.g. learning about buffers in blood after learning about weak acids and bases). This was criticised by Gilbert (2006) as being 'tacked on' and the lowest form of context-based learning.
2) '*Context as reciprocity between concepts and applications*'. This model allows for a two-way connection between concepts and application. For example, the context of an environmental scientist working in their field and the theoretical concepts required to undertake their work. This can induce complications around language use (e.g. 'pure water' has a different meaning to an environmental chemist as compared to a physical chemist) and it can be difficult for students to understand why a given theoretical concept matches a given application or field.
3) '*Context as provided by personal mental activity*'. Through this model, a strong sense of narrative or story-telling is invoked. Here, the actions of a given scientist or a research group are discussed in the context of both their discoveries and the scientific understanding of the time. This is particularly powerful when connected to more modern research as students can more readily understand the societal and scientific contexts of the time period. It should be noted however that students are still external observers throughout these learning activities.
4) '*Context as the social circumstances*'. This model allows for students to become participants in the learning process whilst also engaging in the contextualisation of the activity. Here, students investigate a given subject matter, such as Genetically Modified Foods, and participate in a teacher-guided investigation onto the topic. Teachers are no longer the source

of 'right' or 'wrong' answers and students must investigate the given contextualised concept using more internal motivation.

The importance of context in chemical education was further considered by Mahaffy (2004) particularly in regards to Johnstone (1991) who stated that for students to understand a given theoretical concept in chemistry they needed to understand the topic on three levels - symbolic (e.g. drawings of chemical structures), macroscopic (e.g. changes in physical properties or colour) and microscopic (e.g. changes in bond angles or electron densities). Mahaffy extended this work to include the need for students to connect to the material through a 'human element' – i.e. to contextualise the materials in real-world scenarios that have meaning to the learners. These two principles are shown below in Figure 2. Mahaffy also explained a range of ways in which this tetrahedron could be used to guide student learning such as using student-generated visualisations of theoretical concepts, although no effect of these interventions was discussed.

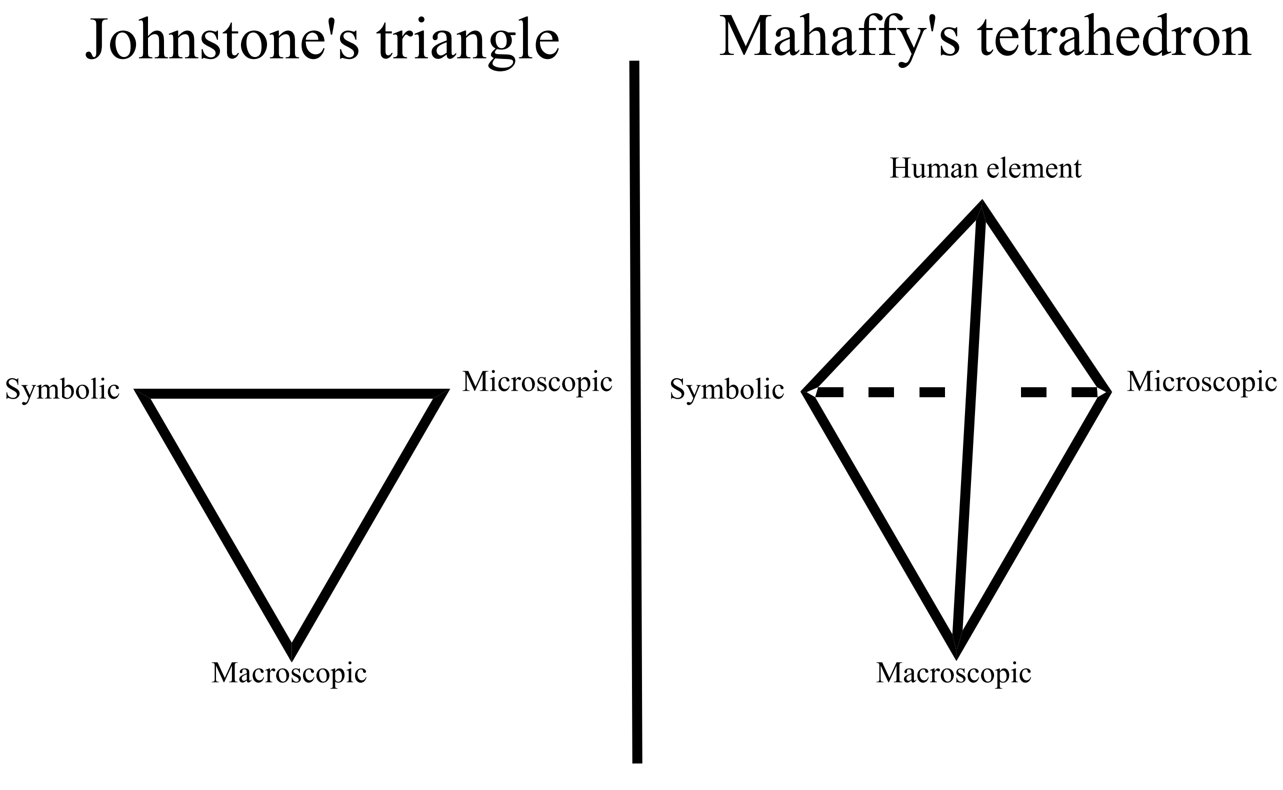


Figure 2 - The levels of student understanding required for students to fully understand a given theory according to Johnstone (1991) and later extended by Mahaffy (2004).

An example of using context-based learning on a unit/course level can be considered through the work of Bulte, Westbroek, de Jong, and Pilot (2006). In this case, a unit/course was designed around the context of water quality in which secondary school students investigated a range of water samples of a neighbouring suburb. The concepts covered included dissolved materials, concentrations, determination of solutes alongside accuracy and precision during experimental measurements. The context particularly focused on chemistry in society and was designed to personally connect with students, particularly in asking them if they would drink the water themselves. Over several iterations, it was noted that student engagement was high, and they enjoyed the context of the activities and the inquiry-based nature of the tasks. It was also noted that students were more able to articulate the theoretical understanding as noted through video interviews and surveys. However, it was noted that students failed to see the point of the assessment (a traditional laboratory report) and required strong support and guidance throughout the activities to continuously appreciate the context of each new step and sub-topic. This example highlights the potential outcomes on a specific unit/course scale.

Relevant results can be found in the work of Broman and Parchmann (2014) who investigated students' ability to answer a range of problem sets that were strongly contextualised beyond simply theoretical concepts. The topics covered ranged from intermolecular bonding to organic chemistry mechanisms whilst the contexts included, drugs, fats, soaps, energy drinks and fuels. The ability of the students to solve open inquiry-based questions with these contexts in mind was noted through recorded 'think-aloud' interviews with 20 upper secondary students. It was noted that:

> *'Students' apparent use of the contextualization of our tasks, especially regarding the personal context has been supporting and not distracting to the students. This assures that the context in itself is helpful when solving the problem, not being merely a decoration or a motivational trick'*

This is an important finding in that students were still able to undertake significant learning with regards to theoretical concepts when strong contexts were utilised. It was again noted that the context helped to increase student engagement further solidifying the potential benefits of context-based learning.

The use of context-based learning on a large-scale can be considered through the work of Avargil, Herscovitz, and Dori (2013), in which a new context-based chemistry curriculum was designed, developed, and delivered throughout all secondary schools in Israel between 2000 and 2011. It was found that teaching staff became highly motivated by the new contextualised curriculum, with the number of teachers electing to undertake the altered curriculum increasing exponentially before it was made mandatory in 2011. However, many teaching staff struggled with the new curriculum as they "*were used to teaching their students how to solve chemistry problems in an algorithmic way*". Hence, the 'buy-in' of teaching staff is an important challenge when implementing new context-based activities. Additionally, failure rates of students undertaking the new curriculum were noted to be a third of those who continued to study the traditional curriculum whilst assessment was able to be more focused on increasingly higher-order application focused questions.

Another large-scale example exists in the 'Salters' approach' in the UK as described by Campbell et al. (1994). The Salters' approach, so named after the original benefactor of the project, the Salters' Institute for Industrial Chemistry, was a large-scale curriculum reform in the UK during the mid-late 1980s and early 1990s. Its original focus was on chemistry courses for middle school students (13-year-olds) but later expanded to secondary schools, other sciences (biology and physics) and finally to advanced courses for 17-18-year-olds finally totalling six different courses (Bennett & Lubben, 2006). It has also been adopted outside of the UK in Belgium, Hong Kong, New Zealand, Russia, Scotland, Slovenia, Spain, Swaziland and the USA. The Salters' approach was generally designed to increase the context-based nature of the science taught in schools, particularly with regards to helping students to develop an appreciation of the connection between chemistry, their daily lives, society and the natural environment. Specifically referring to the advanced chemistry course, teachers were noted to feel that they found the course:

‘*more motivating to teach; that their students were more interested in chemistry … [the students] were better able to engage in independent study and take more responsibility for their own learning. However, they reported that they found the course more demanding to teach.*’ - (Bennett & Lubben, 2006, p 1007)

Students participating in the middle and secondary schools:

‘expressed higher levels of interest in the course and commented positively on the wide range of activities … [but they] expressed more concern than students on the more conventional course about their abilities to cope with revision and tests.’ - (Bennett & Lubben, 2006, p 1009)

Overall, it would seem apparent from the literature that context-based learning is able to result in higher student engagement, increased performance on subsequent assessments and potentially broader skill development. However, this can come at a cost of additional effort for teaching staff and may result in student concerns around traditional examination or other assessment practices (e.g. laboratory reports). With regards to laboratories, some recent examples in university-level undergraduate chemistry units/courses can be seen in Table 2. These examples highlight how applicable and diverse contextualised teaching laboratories can be, and how popular context-based learning has become as a pedagogy for modern laboratory activities in a similar vein to its large-scale use in other forms of content delivery in schools around the world. The articles tended to highlight the increased engagement of the students and their enhanced understanding of the desired scientific content, which matches with the previously discussed findings of context-based learning.

Table 2 – A range of exemplar contextualised laboratories.

Reference	Scientific content	Context	Major findings
Esson, Scott, and Hayes (2018)	Solubility of paints.	Paint removal from vandalised artwork (a real-life case study).	100% of students enjoyed the experiment and identified the main lesson 71% of students agreed or strongly agreed that 'This experiment helped me develop my ability to hypothesize'
Rajapaksha et al. (2018)	Gas Chromatography	Determination of Xylitol in fresh and chewed gum to determine toxicity for dogs.	It was noted that 'Students were excited to find gum chewing a part of the planned exercise' 77% picked this laboratory exercise as their favourite over the entire semester.
Samarasekara, Hill, and Mlsna (2018)	Paper chromatography.	Organic acids in commercially available wines and fruit juices.	Student survey results showed that: "the experiment worked well so that [students] got good results" (average of 4.0 "agree" on a Likert scale of 1−5) "the experiment [was] interesting to perform" (average of 4.0 out of 5) "[students] would recommend others to do the lab" (average of 4.1 out of 5). Subsequent assessment indicated that students understood the key learning outcomes (>84% of students answered the key questions correctly)

1.6 Problem-based learning.

Whilst inquiry- and context-based learning are undeniably impactful, it is important to consider their impact when utilised together. For example, a subset of inquiry-based learning, known as problem-based learning (Hmelo-Silver, 2004), incorporates both context- and inquiry-based learning together. Problem-based learning (PBL) experiences are a subset of inquiry-based experiences which present the students with a real-life scenario or problem, thus bringing together context-based learning and inquiry-based learning. The students are required to solve the task given far less overall guidance than traditional classes or expository laboratory exercises (Schmidt, Rotgans, & Yew, 2011), typically being asked to identify what information they need for themselves (Torp & Sage, 1998). PBL has been used extensively in training for medical professionals for decades with students noted to be more prepared for learning activities, more comfortable with conflicting information or uncertain answers and even to be more prepared once they entered clinical work after their university degrees (Donner & Bickley, 1993). Hence, PBL experiences are a combination of context-based learning and either guided, open or authentic inquiry.

Many examples exist in the literature of either IBL or PBL where individual laboratory activities have been modified to adopt these new pedagogies. Examples include:

1) A student-driven investigation into potential art forgeries (Nielsen, Scaffidi, & Yezierski, 2014). Students utilise Raman spectroscopy and decide on the experimental conditions required to detect the use of different dyes and pigments on supplied samples. A major finding of this work was that the students were routinely forced to overcome obstacles that they were not pre-warned about. Hence, students developed the skills required to problem-solve under considerable time pressures. Furthermore, as noted through the assessment, students showed a significant ability to analyse and explain their complex procedures and decisions, indicating a use of higher order thinking beyond simply recall of theoretical concepts.

2) A class-wide synthesis of a polymer with a wide range of molecular weights and properties (Mc Ilrath, Robertson, & Kuchta, 2012). Students designed tests to investigate the isolated polymers and discuss and integrate discrepancies and contradictions in results collated from the entire class. The main finding of this work was the large increase in student engagement. Students were seen to be highly enthusiastic about the experiment and took ownership of the methods used to determine the strength of the final polymer such as discovering that some polymers formed through low catalytic loading '*were unbreakable when heaved down multiple flights of stairs, using a table vice, or slamming in a door*'.

3) An investigation into the identity and amounts of dissolved metals in simulated hazardous wastes (Dunnivant, 2002). Whilst basic procedures were given, students had to adapt them to each sample provided. A range of techniques were utilised, and students were tasked with investigating and explaining discrepancies between the results obtained from different instrumentation. Once again, students were forced to overcome obstacles that they were unaware of and were encouraged to undertake significant trial and error to find the best technique/method possible for their given sample. Whilst this caused significant frustration for the students, it was noted that the laboratory activity was '*voted the most enjoyable and best learning experience of the labs*'. Furthermore, through dealing with complex issues and multiple or ambiguous results, it was believed that '*the students are better prepared to analyze real-world samples upon graduation*'.

Overall, these examples highlight significant student engagement with PBL laboratory activates alongside a significant development of skills with regards to handling unknown or complex situations. However, it was also highlighted that students sometimes struggled with the new procedures due to the uncertainty of the task on hand. Issues with regards to PBL experiences for both students and teaching staff who struggle to adapt to the new learning environment are common in the literature and tend to focus on perceived difficulty and a noted decrease in specific knowledge retention (Basey, Maines, Francis, & Melbourne, 2014; Cummins et al., 2004; Schmidt et al., 2011; Tosun & Taskesenligil, 2013). Another major challenge in PBL experiences is the students' ability to seek out relevant information in order to complete the more open-ended task, which requires significant scaffolding and guidance beyond simply setting a PBL task (Shultz & Li, 2016).

Regardless of these issues, PBL laboratory exercises have been shown to increase both the students' general problem-solving ability but also their metacognitive skills (specifically how students can control their thinking during a specific task) (Sandi-Urena, Cooper, & Stevens, 2012). Overall, PBL activities appear to be more suited to the development of problem-solving or scientific processing skills (Tosun & Taskesenligil, 2013) at the expense of focusing on specific theoretical content and simple retention.

1.7 Cognitive load theory.

The response to PBL and IBL has not always been favourable, as typified in the published discussion of Kirschner, Sweller, and Clark (2006). This paper focuses in detail on a major issue with any form of learning that requires minimal guidance by the teachers, namely the issue of cognitive overload for the students. Cognitive load refers to the number of discrete pieces of information that one can have in their working (or short-term) memory at any given moment (Kirschner et al., 2006). Typically, this number (G. A. Miller, 1956) is around seven ± two and if a cognitive overload occurs during instruction, the learner experiences reduced performance on the given task (Mayer & Moreno, 2003). This can simply be measured through the amount of effort a given learner has to expend in order to obtain the desired learning objectives (Paas & Van Merriënboer, 1994; Paas & Van Merriënboer, 1993). Minimal guidance experiences that utilise higher level IBL/PBL tasks, require students to not only think about the scientific content at hand but also the new procedure which they must personally generate – a situation that can easily lead to cognitive overload.

Whilst cognitive load is undeniably a concern when designing IBL/PBL tasks, there are ways in which it can be overcome. Firstly, one must consider the different forms of cognitive load individually:

1) Intrinsic – The ‘difficulty’ of the assigned task (e.g. basic arithmetic as compared to advanced theoretical mathematics).

To a first approximation, the ‘difficulty’ of a given task cannot be altered by teaching staff as it is inherent within the task itself, the level of prior learning a student has undertaken and individual student ability. The difficulty of a given task is generally determined by the level of ‘element interactivity’ (Paas, Renkl, & Sweller, 2003) wherein more complicated tasks (e.g. learning sentence structure whilst learning a new language) have higher elements of interactivity than more simple tasks (e.g. learning individual words in a new language). Truly minimal guidance experiences will fail if they are beyond the current ability of the student to achieve due to the high level of prior learning required and the large degree of element interactivity in the task (e.g. how individual variables will affect one another and the results). The extent to which a student can successfully learn new material is often referred to as their ‘zone of proximal development’ (Vgotsky, 1978) which implies that a given task should not be too far outside the ability level of the students (Pea, 2004) if successful learning is to occur. For example, an activity with little to no procedural guidance is bound to fail if students are not already competent with both the theoretical knowledge, the necessary practical skills required and an appreciation for the way in which variables will interact (such as temperature, concentration, reaction time and so on).

2) Extraneous – The way in which information is presented to students (e.g. the readability of a laboratory manual).

This is the most easily influenced source of cognitive load, as the presentation of the learning materials presented to the students is entirely developed and disseminated by teaching staff. This is particularly powerful when combined with any pedagogy that requires students to prepare for a given activity *before* they arrive, such as the use of pre-laboratory quizzes (Johnstone, 1997; Reid & Shah, 2007). Whilst the exact composition of pre-laboratory quizzes is currently being discussed in the literature (Agustian & Seery, 2017), their ability to reduce extraneous cognitive load in the laboratory (as students are already familiar with the laboratory manual) is well documented (Gregory & Trapani, 2012; Koehler & Orvis, 2003; McKelvy, 2000; Schmid & Yeung, 2005).

3) Germane – The ease or difficulty of connecting the learning task to prior learning through the generation of schemas.

Schemas are mental subroutines (or a collection of complicated facts) that speed up or ease difficult mental tasks or, as per Paas et al. (2003), 'cognitive constructs that incorporate multiple elements of information into a single element with a specific function'. They result from constant exposure to problems wherein the learner becomes an 'expert' at a given task insofar as it becomes almost rote or automated in subsequent use. The cognitive load required to form new schema, or to fold new information into a previously formed schema, adds to the cognitive load during any given learning activity. Considering IBL/PBL laboratory activities often aim to enhance a student's ability to design an experiment or to deal with unknowns rather than to develop theoretical knowledge or specific practical skills, the germane cognitive load in these experiences can be quite high. Hence, it is ideal to limit the number of schemas that a student would be attempting to form during such an activity. It is important to note that an IBL/PBL task may still be difficult and therefore require a high degree of intrinsic cognitive load, even if the task covers previously learned material and has a lower germane load.

There is an additional concern here if one also considers the use of teams or groups in undergraduate laboratories with regards to cognitive load. Kirschner, Sweller, Kirschner, and Zambrano (2018) extended the framework of cognitive load theory to also consider the effect of team/group work on the cognitive load of students. Overall, it was another potential issue with students now having to successfully communicate with one another and to draw upon a collective working memory, i.e., the amount of information the team or group could handle before experiencing cognitive load. It was suggested that assigning roles to students (e.g. data recorder, team leader etc.), using consistent teams over several learning tasks (to induce familiarity between the learners) or avoiding large teams/groups (i.e. three or more students) could help manage this load, but also that not all tasks should simply be converted into team/group work if the inherent cognitive load of the task is already considered to be high. With regards to laboratories and the use of teamwork to reduce costs through a reduced need for glassware and reagents, this provides a strong argument for individual or paired activities as much as possible. This is particularly relevant when undertaking inquiry/problem-based tasks as the cognitive load in those tasks is likely to already be quite high.

Overall, when incorporating IBL/PBL activities, strong consideration of the cognitive load (whether individual or group) is required to avoid students failing to achieve the learning objectives of a given learning task.

1.8 Assessment issues in laboratories.

Whenever any new pedagogy is incorporated, subsequent assessments should also be altered to align with the new focus – a process often referred to as constructive alignment (Biggs, 1996). As such, the assessment tasks should match the new contexts or student-based projects to fully support the new activities and to avoid a mismatch between the new contexts and the assessment tasks (e.g. a scientific report when the context is working for a corporate entity). For example, the assessment can be tailored to a specific activity, such as the use of executive summaries (i.e. short summaries to commercial clients, business partners, or governmental agencies) or templates that match official government documentation. This is known as 'authentic assessment' (Gulikers, Bastiaens, & Kirschner, 2004), wherein the assessment is altered to match a real-world or professional experience. Through their literature review, Darling-Hammond and Snyder (2000) highlighted that authentic assessment can:

1) motivate student learning by connecting the assessment task to real, engaging scenarios.
2) provide a deeper understanding of student learning as students often struggled to apply knowledge to real-world scenarios when they didn't fully understand the fundamental theory.
3) aid in teacher development, particularly with regards to improving their pedagogical practices and helping to focus their classes on specific real-world contexts.
4) provide evidence of educational reform which can, in turn, generate greater reform in other learning environments.

Additional research by Diller and Phelps (2008) highlighted that authentic assessment can lead to students gaining a greater appreciation for learning beyond simple assessment performance. Alongside this finding, it was also noted that students tended to develop a greater range of higher order skills such as critical thinking, the sense of place in society, information literacy and communication. Clearly authentic assessment can have a plethora of positive outcomes and its use in guiding student learning after the completion of laboratories is of importance. Indeed, several examples in the literature in which students either undertake class debates on the environmental impacts of varied synthetic pathways (Pilcher, Riley, Mathabathe, & Potgieter, 2015) or produce non-scientific reports (Erhart, McCarrick, Lorigan, & Yezierski, 2015) has resulted in greater student learning, especially with regards to gaining an appreciation of societal, environmental or commercial concerns.

Another aspect of assessment that requires consideration is the variation of assigned marks when large numbers of teaching staff are involved in the marking process. It is a common finding that when large numbers of teaching staff are involved in assessing the work of many students, there is likely to be considerable variation within the marks assigned that is dependent upon the markers themselves (i.e. a single piece of work may receive vastly different marks from different teaching staff). This is not a new issue, with studies into this phenomenon first particularly notable in the 1980s (R. C. Bell, 1980; Byrne, 1980; Collier, 1986; Edwards, 1979; Hall & Daglish, 1982). Whilst this issue can exist in any case where large numbers of teaching staff are utilised, this is particularly observed in teaching laboratories due to the large turnover of staff (Smith & Coombe, 2006).

Two common means for dealing with the issue of marking variation is to either undertake structured training (or calibration) *before* the assessment or to attempt to moderate marks *after* assessment has already taken place. Whilst there is evidence of the success of either calibration (Bird & Yucel, 2013) or moderation (Zahra et al., 2017) in reducing marking variation, arguments in the literature against the use of such processes (Bloxham, 2009; Bloxham, den-Outer, Hudson, & Price, 2016) highlight the large amount of effort and commitment on behalf of the teaching staff involved (i.e. large amounts of time or personal commitment).

It is argued that a large amount of this marking variation is likely due to the core beliefs of the teaching staff themselves (i.e. what they believe makes for high or low quality work) which tends to be based on their previous educational, teaching and cultural backgrounds (Hunter & Docherty, 2011). Hunter and Docherty argue that this large variation in the core beliefs of teaching staff would appear to lead to markers using significantly different levels of academic judgement – i.e. markers are potentially using differing degrees of internal judgement, rather than explicit marking criteria or guidelines, in order to assign marks to a given assessment piece. Furthermore, marking variation was also noted to be dependent on the discipline being assessed, with more quantitative subjects (e.g. chemistry) exhibiting a higher spread of marks with greater marking variation noted between markers when compared to more qualitative subjects (e.g. history or English) (Bridges et al., 1999). This finding is considered an artefact of the lower degree of certainty in the more qualitative subjects (i.e. there is rarely an exact correct answer) which leads to markers less likely to assign either very high or low marks – which in turn narrows the marking range and subsequent marking variation noted.

Overall, it is clear that marking variation is a highly complex and somewhat contentious issue. This is further compounded when feedback is taken into consideration. Firstly, students are not unanimous in the type of feedback they desire, with requests for either exact guidance (i.e. a detailed list of every mark gained and lost with reasons noted) or more general feedback provided (i.e. an overall mark and comment given) noted in the responses of students (Bell, Mladenovic, & Price, 2013). Next, electronic means of feedback are becoming increasingly common, with students stating similar levels of satisfaction when compared to hand-written feedback and teaching staff reporting reduced marking time (Sopina & McNeill, 2015). There also exists a large range of choices for electronic feedback including automated (i.e. computer generated or assigned) (Watt, Simpson, McKillop, & Nunn, 2002), audio/video feedback (Lunt & Curran, 2010) or specifically designed computer programs (Campbell, 2005; Heinrich, Milne, Ramsay, & Morrison, 2009). It is worth noting that these articles all raise the issue of retraining teaching staff to use the new assessment procedures alongside the significant time-investment often required.

Whilst these studies often noted students were more likely to access electronic feedback, it is important to consider that how a student utilises the feedback is more important than whether they simply accessed the feedback (Crisp, 2007). Typically, the ‘transmission’ model is utilised (i.e. where a list of comments is simply provided to the students) as opposed to a ‘dialogue’ model in which students obtain feedback through a continuous conversation with teaching staff (Ajjawi & Boud, 2017, 2018; Carless, 2015). As noted by Ajjawi, Boud and Carless, the dialogue model results in a stronger teacher-student interaction and results in feedback that is noted to alter student behaviour. However, it is clear through these articles that whilst the dialogue model is ideal and founded upon sound pedagogical principles, this model requires significant training of teaching staff and would likely limit the amount of assessment achievable within a given time frame.

Whilst automated marking processes, authentic assessment, and inquiry/context-based learning are undeniably impactful, it is important to consider their impact when utilised together. For example, a subset of inquiry-based learning, known as problem-based learning (Hmelo-Silver, 2004), incorporates both context- and inquiry-based learning together and can readily matched with altered marking procedures and authentic assessment.

1.9 The scope of the TLL programme

Interestingly, this need to prepare students for the workforce brings the focus of the laboratory experiences back to their roots in the mid-1800s, which were typically used to teach technical skills required for industry[i] and research (Elliott, Stewart, & Lagowski, 2008; Good, 1936). Examples can be found in the literature of changes that have been made to align teaching with industrial requirements on various scales:

1) An individual laboratory exercise.

In this example (Erhart et al., 2015), students are tasked by a fictional company (CitrusTech) to undertake non-destructive testing of lemons using NMR and MRI techniques. During the first week of the experiment students are tasked with discovering the parameters that effect the output of an NMR machine in a relatively theory-focused experiment. However, in the following week, the students apply their understanding of NMR to an MRI machine, in the hopes of investigating the quality of lemons that may have been affected by frosts during their growth. After completion of the laboratory, students are instructed to then submit a technical report to the company that is written in such a manner that a lay person could easily understand both the results and the use of NMR/MRI. It was found that students were able to communicate a strong understanding of the use of both an NMR/MRI machine on a detailed technical level. Additionally, through generating a report designed for a lay audience, students further developed skills around non-scientific communication. This was assessed through a rubric designed to focus on these key learning outcomes. This is a strong example of contextualising the laboratory activity in an industrial setting in both the set-up (working for

---

[i] NB: In this case (and throughout the thesis), the term 'industry' refers to any and all commercial endeavours as well. This includes Innovation, Enterprise, Start-up activities as well as large scale production of materials.

CitrusTech) and the assessment (submitting a non-scientific technical report). Other similar examples exist in the literature such as investigating methods to clean up simulated chemical spills on behalf of an environmental agency (Hicks & Bevsek, 2012) or, outside of the laboratory, group learning activities investigating the industrial requirements for refrigerants or sodium chloride, chlorine gas and hydrogen gas (Lennon, Freer, Winfield, Landon, & Reid, 2002).

2) An individual unit/course.

At the University of Pretoria (Pilcher et al., 2015), a concerted effort was undertaken to overhaul a 3rd-year organic chemistry laboratory course to incorporate inquiry-based learning in an industrial context. Students were given a brief from the fictional company 'Chem-Co Ltd' and asked to investigate the synthesis of a target organic molecule that is used in pharmaceuticals, agrochemicals, dyestuff and even the fragrance industries. Students are given three different routes to investigate and must present their findings in a mock board meeting with the chemical company. Following this, students were invited to participate in class debate/discussion about which synthetic pathway is most feasible in terms of cost, safety and potential scalability. Throughout this presentation, it was clear that students had given a large amount of consideration to the environmental and commercial aspects of chemistry. In several cases, students were even noted to raise highly specific items such as the environmental issues of by-products and waste or the cost of heating a reaction or the salary of a given scientist. Survey responses showed students were highly engaged with the process and felt their technical skills were enhanced by the longer multi-week laboratory activity. Additionally, whilst many students raised the difficulty of the program, they were highly engaged in the process and routinely raised it as an example of the best laboratory experience that they had to date. As with the CitrusTech example, this laboratory activity showcases the use of an industrial context throughout both the setting and the subsequent assessment but, in this case, on the much larger scale of an entire unit/course.

Industry-focused examples are limited in the literature and never appear to encompass multiple year levels of the same undergraduate laboratory programme (e.g. all chemistry teaching laboratories over all year levels were up for consideration). This challenge brings us to the aim of this work, to research the impacts of the Transforming Laboratory Learning project in the school of chemistry at Monash University which ran from 2016-2018 and aimed to:

i) Increase the amount of real-life context throughout the laboratory activities and their assessment. Where possible, this would focus on industrial/workforce environments.

ii) Introduce new inquiry/problem-based laboratory activities or increase the amount of inquiry in some of those that remained.

iii) Ensure that the changes made were independent of the original researchers and could continue to run (and be refined) after 2018.

The impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments was measured through students' perceptions of their laboratory experiences. The research themes were framed around the areas of student perception that were a) most likely to be altered by the large-scale inclusion of the new activities and b) could result in an increase in student employability through their increased ability to articulate the skills developed throughout their laboratory experiences. Thus, research areas were identified which are listed shown in Figure 3:

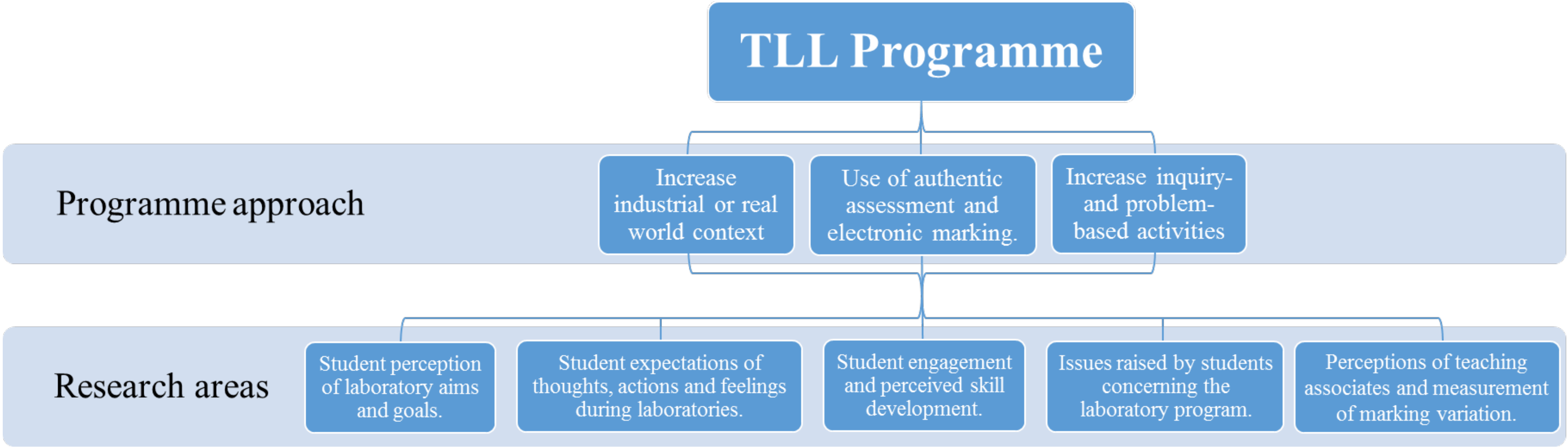


Figure 3- An overview of the approach of the TLL programme and the associated research.

These themes were investigated through the following research questions:

1. What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' perception of laboratory aims and their expectations of their thoughts, actions and feelings during laboratories? (Baseline data only: How does this compare to the perception of teaching associates and academics?)

2. What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' level of enjoyment of the laboratory exercises?

3. What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the development of students' employability skills and their recognition of these skills?

4. What was the effect of reducing academic judgment on the marking variation and feedback provided by the teaching associates?

Chapter 2 - Research Framework and Methodology

The central hypothesis of TLL was that the programme will increase the employability of undergraduate students using authentic, inquiry-based laboratory exercises and industry contextualised experiences (Figure 4). The use of these two pedagogic approaches should result in more engaged students with an increased understanding of scientific methodology and enhanced transferable skills, or at least an increased ability to articulate their developed skills.

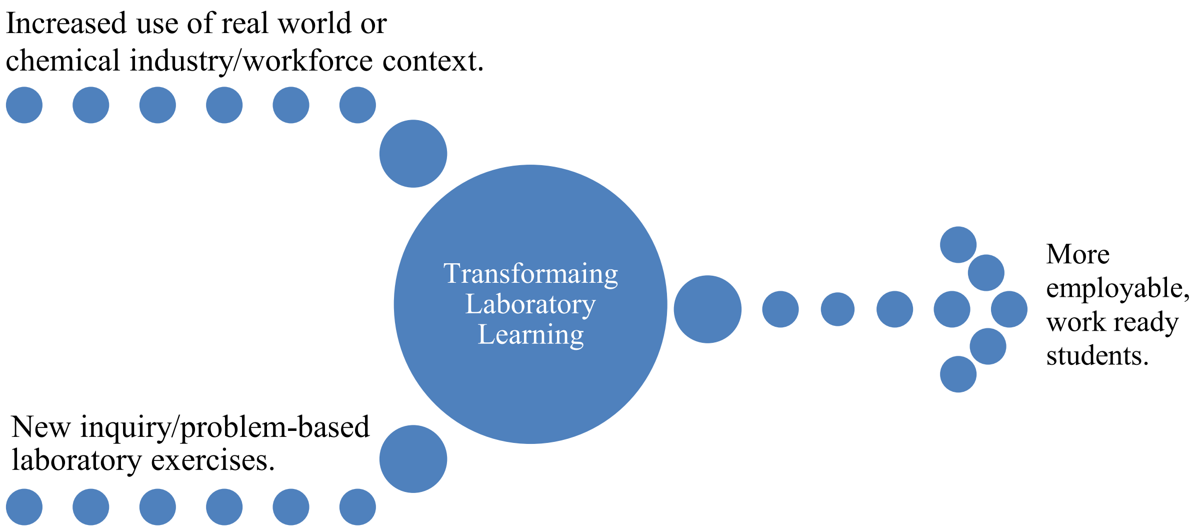


Figure 4 - A pictorial representation of the goals of the Transforming Laboratory Learning (TLL) programme.

TLL was underpinned by the theory of Constructivism which '*underlies the assumption that learning is an active process where knowledge is constructed based on personal experiences and the continual testing of hypotheses*' (Leal Filho & Pace, 2016). Hence, to learn, one must be able to actively reconcile new information with previous experiences and either incorporate the new information, even perhaps changing one's beliefs, or to undertake further testing of the new information through active inquiry. This is especially relevant for learning in a laboratory environment where students will ideally connect to the new contexts provided either through their connections to their personal daily lives or their beliefs and perceptions of the tasks performed in the modern workforce (whether that be in industry, academia or elsewhere). This would ideally allow students to reconcile the new information with their own previous personal experiences. Furthermore, the continual testing of hypotheses, through the inclusion of more inquiry- or problem-based activities, will potentially allow students an opportunity to undertake their own inquiry into the tasks and topic on hand. Theoretically, this would allow the students to achieve a deeper level of learning as predicted through Constructivism.

It is also worth noting that many of the activities encouraged teamwork and communication skills through the use of enforced groupwork and assessments. Through this, students were required to work together and develop a shared understanding of the task on hand – which is a concept more akin to Social Constructivism (i.e. the construction of knowledge through interactions with other people (Hodson & Hodson, 1998)). However, the TLL programme sought to increase of the employability of *individuals*, rather than groups of students, and, as such, the effect on the ability of students to work together and to construct knowledge did not form a major focus of this work.

As Constructivism (and to a lesser extent, Social Constructivism) was used to guide the generation of new laboratory activities, it was also utilised to drive the research questions. The degree to which the students were able to connect to their own personal experiences was measured through direct questions related to their engagement with the material and how much they felt the new laboratories connected to the real world or to the workforce. Additionally, questions also centred on the degree of control students believed they had with the direction of the given laboratory activity, which is of course crucial if students are to be allowed to undertake their own inquiry into new information. Furthermore, the effect of the programme on students' predictions for future laboratory activities was used to investigate whether the students were able to reconcile the new experiences with whatever preconceptions they may have held about the aims and goals of teaching laboratories. This was particularly important with regards to the students' perceptions of their own developed skills, as their ability to articulate their own skill development may be crucial for future employment and job interviews.

In order to address the research questions outlined in Section 1.6, a mixed methods approach was adopted. In this case, the mixed method approach utilised qualitative data (answers to open-ended questions or recorded audio during interviews) and quantitative data (answers to surveys with fixed answers/scales) which were collected concurrently (Creswell, 2009). This mixed methods approach was believed to provide a more complete measure of human behaviour (Gay, Mills, & Airasian, 2006; Morse, 2003) where any limitation in one method should theoretically be (partially) covered by the other. The alignment of the research questions with the methods in shown in Table 3.

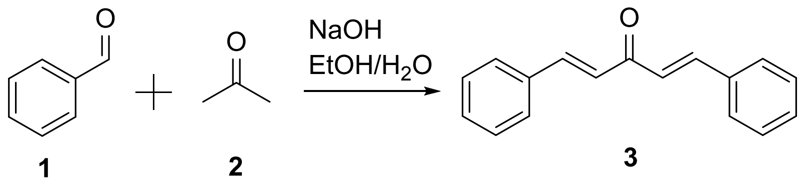
Table 3 - The research methods used to address the key research questions.

Question	Research method
1. What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' perception of laboratory aims and their expectations of their thoughts, actions and feelings during laboratories? (Baseline data only: How does this compare to the perception of teaching associates and academics?)	(Baseline) Large scale paper-based surveys to students and teaching associates. (Baseline) One-to-one recorded interviews with academic staff. Annual large-scale paper-based surveys to students.
2. What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' level of enjoyment of the laboratory exercises?	Large scale paper-based surveys to students after completion of individual experiments before and after the TLL programme. Audio recorded focus groups with students after completion of units/courses before and after the TLL programme.
3. What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the development of students' employability skills and their recognition of these skills?	Large scale paper-based surveys to students after completion of individual experiments before and after the TLL programme. Audio recorded focus groups with students after completion of units/courses before and after the TLL programme.
4. What was the effect of reducing academic judgment on the marking variation and feedback provided by the teaching associates?	Surveys to teaching associates. Tracking of student marks.

In order to better conceptualise the TLL programme, three examples of laboratory activities generated will be discussed:

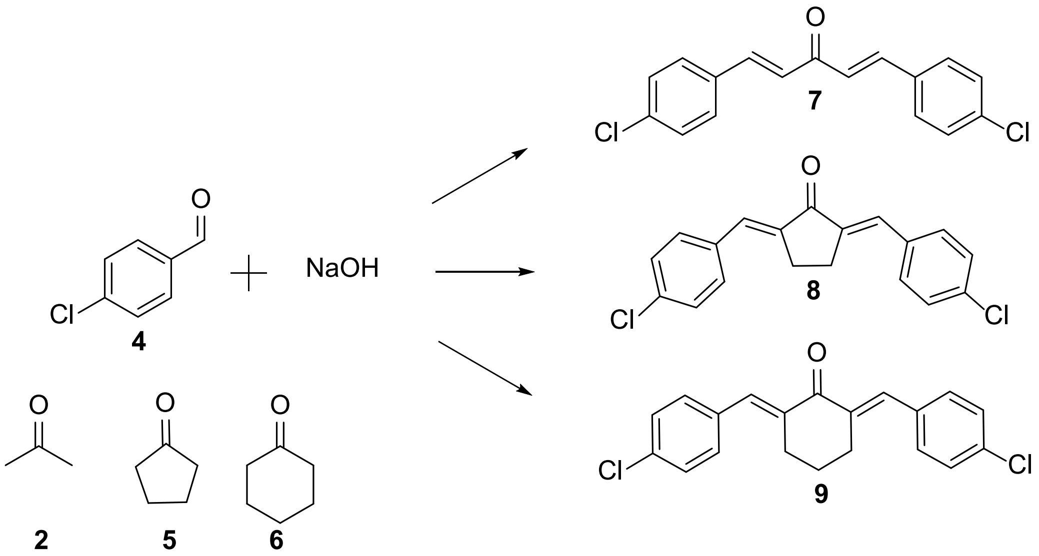
2.1 Example 1 – Making effective sunscreen ingredients using Claisen-Schmidt condensation (Appendix 1)

This activity began as a traditional 2nd year organic chemistry laboratory exercise on the reaction of benzaldehyde **1** and acetone **2** to form the Claisen-Schmidt condensation product **3** (Scheme 1).



Scheme 1 – The reaction of benzaldehyde **1** with acetone **2**.

From the reaction scheme alone, it is perhaps not surprising that the original design of this laboratory activity was very expository. Students were instructed to simply make the product, isolate and then characterise it. No effort was made to contextualise the material in everyday use nor was the method linked to industrial procedures. The key to redesigning this activity was the discovery that the product can be used as an active ingredient in sunscreens which was noted through an article in the Journal of Chemical Education (Huck & Leigh, 2010) which highlighted that these materials could be made even more applicable as sunscreen reagents through changing the ketone that is used (in this case, acetone **2**). Lastly, it was noted that these new products would require heating to form and would, therefore, be costly to make on an industrial scale. This last point led to a direct connection to a Melbourne-based skin care company, Rationale™, who agreed to have their name and logo branded onto the new laboratory activity, which utilised the following reaction pathway (Scheme 2):



Scheme 2 – The expanded reaction of 4-chlorobenzaldehyde **4** with a range of ketones (**2**, **5** and **6**) to form condensation products **7-9**.

In the new laboratory activity, students were given information about the sunscreen context, the use of various ketones and the difficulty of formation without the use of heat. An authentic learning environment was created by referring to the students as Monash Consulting – a fictional company that employed the students as researchers. They were instructed, in an 'authentic' request from the skincare company, Rationale™, to investigate the synthesis of chlorinated derivatives **7-9** that would be formed at room temperature in order to avoid the energy costs of forming these materials on an industrial scale. The students worked in groups of three, with each student forming one of the three chlorinated products through a provided method modelled on actual research papers (i.e. not written in a stepwise fashion).

At the end of the experiment, students shared their results with one another and then, individually, wrote a typical scientific report alongside an executive summary (i.e. a short non-scientific communication) to Rationale™, which provided a more authentic assessment task. It is also worth noting that they were asked to consider the application of these materials through UV-analysis which is a technique not used in the original activity. Overall, the new laboratory:

- Required the same practical techniques as in the original activity and expanded to include an additional technique (UV-analysis).
- Provided context for the synthesis of these molecules (sunscreens).
- Utilised authentic assessment (Executive Summary).
- Allowed for teamwork due to the group synthesis of a range of compounds.
- Expanded the focus of typical chemistry reactions to industry through a direct connection (Rationale™) and a consideration of energy costs through avoiding the need for heat input.

2.2 Example 2 – Investigation into the Efficacy of a Digestive Enzyme (Beano™) (Appendix 2)

Whilst it was considered ideal for a laboratory activity to be directly connected to an industry or real company, this was not always feasible. In these cases, either a fictitious company was generated or some other real-life context was included. For example, a new laboratory was generated for a 2nd year Biological Chemistry unit/course that investigated the use of a commercial enzyme (Beano™) which can be used to help individuals whose high legume intake results in flatulence or similar gastro-intestinal discomfort. In this case, it was planned to incorporate a large amount of inquiry into the laboratory such that the students could decide which investigation into the enzyme to undertake (e.g. effect of pH or effect of the presence of ethanol). This is an example of a problem-based task due to both the strong context and the use of inquiry.

The technique used to measure the breakdown of complex sugars into glucose (as performed by Beano™) was measured by a standard glucometer used by people who have diabetes, which most students had no experience of. Therefore, a consecutive two-week model was used in which students first undertook an expository exercise designed to allow them to become proficient with the use of a glucometer and in basic analytical experimental design e.g. the generation of standard solutions and calibration curves. The second week was then a research project in which students could choose from a list of six research questions or choose their own if reasonable.

This example highlights that many of the goals of TLL (increasing inquiry/problem/context/industry-based experiments) can still be achieved without a direct industrial link. This outcome of this particular laboratory has been published (George-Williams, Ziebell, Thompson, & Overton, 2018b) and has also been reproduced in Chapter 3.

2.3 Example 3 – Electronic Waste (Appendix 3)

The previous two examples were 2^{nd}-year laboratory activities where students are expected to already have some fundamental theoretical and practical knowledge to draw from. Clearly, this assumption is less reasonable in first-year laboratories and thus traditional, expository experiences may seem more appropriate to this environment. Whilst this is a reasonable assertion to make, there are still ways to increase the impact of these experiences through context- and inquiry-based learning.

At Monash University, there existed a common first-year laboratory activity investigating the colour of transition metal ions in solution. This activity simply had students adding small amounts of solutions together in small glass test tubes and recording their observations. This culminated in an assessment piece that predominately required students to simply list chemical equations for every chemical reaction that had occurred throughout the activity. There were several major issues with this activity:

- There were too many combinations of solutions to consider, which easily led to cognitive overload.
- Beyond theoretical understanding, there was virtually no application to the real world explained to the students.
- Both the assessment and the practical task were highly repetitious with little critical thought required by the students.

To address these issues, a new laboratory activity was sought that still covered the fundamental theory required (transition metals in solution) but was contextualised to the real world and allowed for student-based decisions. The result was the Electronic Waste activity where students were first informed of how metals leach from old electronic equipment into soil. They were then tasked with obtaining observations on how four metal ion solutions react with acids, bases and ammonia solutions in order to determine the identity of metals in an unknown solution (which contained two metals of varying concentrations). The students were also provided with a multi-directional flowchart in which they filled out observations and the identities of the complexes rather than simply following a list of instructions. Students could choose which reaction to undertake at any given time and were encouraged to repeat any observations that were either ambiguous (either due to students missing an outcome or a difficult to interpret reaction result) or that they had failed to properly record.

These changes resulted in a laboratory activity that covered far fewer concepts than in previous years (as fewer reactions were covered), was contextualised (electronic wastes) and allowed for student control (open flowchart) and inquiry (identifying unknown metal solutions). Whilst the assessment still called for students to correctly identify chemical equations, the focus on the laboratory became using observations to determine solution composition. The outcome of this laboratory has been submitted for publication (George-Williams, Ziebell, Thompson, & Overton, 2018a) and can also be found in Chapter 3.

2.4 Development of marking criteria and procedures

Throughout the TLL programme, attempts to address the variation in marking and variation between teaching associates were undertaken in the first-year laboratories. There were three major interventions:

1) Enhanced marking criteria and training. This included adding extra detail into the marking criteria (which often just consisted of a total mark assigned to a section with no further advice)

and the implementation of a specific marking activity and discussion during the annual training for teaching associates.

2) The use of an automated Excel spreadsheet. This entailed the use of an electronic rubric that was housed in Excel. Each item in the rubric had a dropdown box next to it that assessors could simply click to indicate that they felt the work matched the description in that cell. The rubric auto-marked for the assessor and was provided to students as feedback to their work alongside any additional comments.
3) The use of automated Moodle reports (NB: Moodle was the online Learning Management System used at Monash during this study). This final intervention took most of the questions given to students and placed them into Moodle quizzes. Hence, this marking was fully automated by a computer. Only the discussion, conclusion and laboratory notebooks were still marked by teaching associates.

2.5 Participants

Data for this project was primarily gathered from students enrolled in all undergraduate chemistry units/courses at Monash University, Australia. Students at this institution are generally local students (although enrolment is open to any applicant) who have achieved high marks during their high school experience. Monash University is a large institution with 1900+ students enrolled in undergraduate chemistry units/courses over all year levels. Many students complete one-two first-year chemistry units/courses as a prerequisite for other majors (e.g. biochemistry, biology or physics). They can then choose to complete two higher year units/courses for a minor in chemistry or six higher year units/courses for a major. Students can then graduate with a Bachelor of Science degree or return for a research-focused 'honours' year. Typically, a single unit/course is delivered over a single semester and consists of two hours of lectures, a one-hour tutorial, one hour of directed independent study and the equivalent of three hours of laboratory time per week. The codes for the units/courses, which were within the scope of TLL, are shown in Table 4.

Table 4 - The units/courses within the scope of TLL. Those in italics indicate units/courses where more than half of the laboratories were redesigned or replaced by the TLL programme.

Course Code	Unit/course name
CHM1011	Chemistry I
CHM1021	Chemistry II
CHM1051	Chemistry I - Advanced
CHM1052	Chemistry II - Advanced
CHM2911	*Inorganic and organic chemistry*
CHM2922	Spectroscopy and analytical chemistry
CHM2942	*Biological chemistry*
CHM2962	*Food chemistry*
CHM2951	Environmental chemistry - water
CHM3180	*Materials chemistry*
CHM3911	Advanced physical chemistry
CHM3922	*Advanced organic chemistry*
CHM3930	Medicinal chemistry
CHM3941	Advanced inorganic chemistry
CHM3952	Advanced analytical chemistry
CHM3960	Environmental chemistry
CHM3972	Sustainable chemistry

Data was also collected from teaching associates and academic staff at Monash University. Teaching associates were typically either PhD candidates, Honours students or a small number of experienced teaching associates who predominately derive their annual income from this role.

Additional data was collected at the University of Warwick in the UK and from the University of New South Wales (UNSW) and the University of Sydney in Sydney, Australia. The other Australian universities in this study are very similar to Monash University, in terms of the type of enrolled students, the flexibility of their studies, as well as the nature of teaching associates. In contrast, the University of Warwick students choose their science major before starting first-year studies and focus on that branch of science throughout their studies.

2.6 Data Collection

All procedures undertaken and described herein were done so with the full approval of the Monash University Human Research Ethics Committee (MUHREC), with application number 2016000584. The University of Warwick accepted the result of the MUHREC approval and required no further consultation. A timeline is shown below (Figure 5) and will be described in further detail throughout this sub-section:

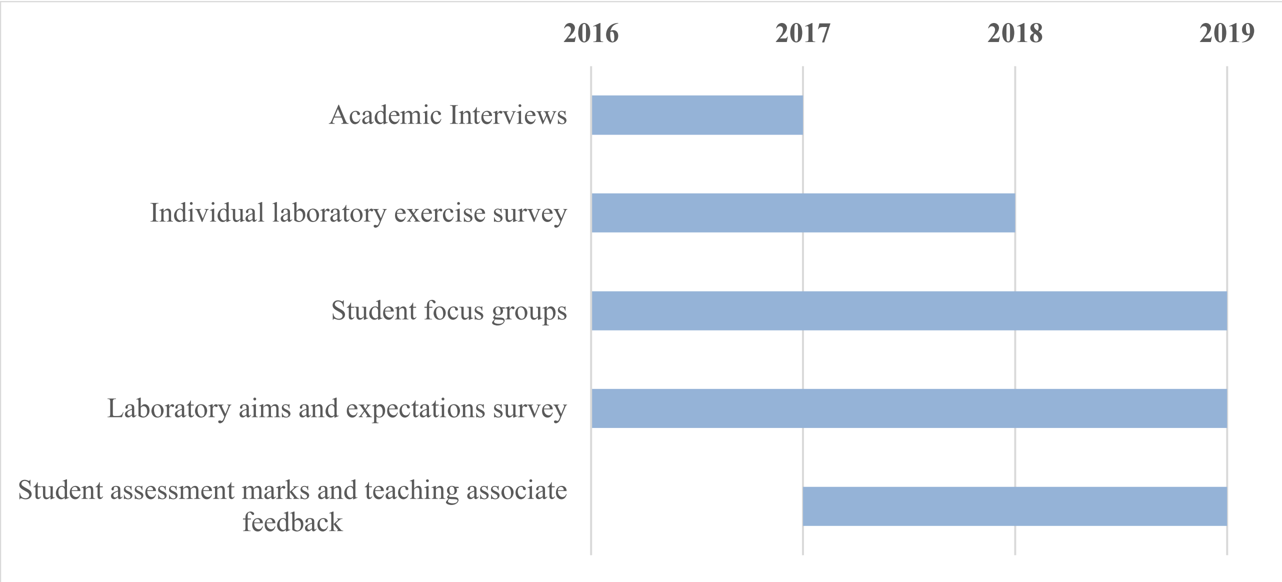


Figure 5 - The timeline of the data collection process.

2.6.1 Individual laboratory exercise survey

To measure the development of students' employability skills and the students' enjoyment with the laboratory exercises, a survey was needed to investigate these issues following an individual laboratory experience. A survey was taken from an available PhD thesis (Russell, 2008) and the following changes made:

i) The formatting was altered to coincide with the other surveys utilised.

ii) Several questions were removed (for example, it was considered unlikely that a single laboratory exercise would cause a student to desire a Masters or PhD degree).

iii) Any reference to ‘chemistry course’ was changed to ‘lab’ in order to ensure the student correctly thought about the individual laboratory experience.

iv) A distractor question (‘Please select Agree and Disagree to this question’) was added to enable the removal of inappropriate responses, such as selecting ‘Agree’ to all statements on the survey.

v) Eight stems were altered to their negative counterpart (e.g. ‘This lab experience made me learn’ became ‘This lab experience *did not*’ make me learn’).

vi) Six new closed questions were added to further probe student enjoyment.

vii) Three open questions were added to provide more in-depth responses regarding student perceived skill development and enjoyment.

This final survey is shown in Appendix 4. This survey was distributed to students at the completion of 13 different laboratory exercises. The number of responses in shown in Table 5.

Table 5 - The number of responses to the individual laboratory survey from 14 different laboratory exercises.

	Year level	Experiment name	Number of responses	% of students
Pre-TLL	3rd	Anthracene oxidation	91	76
	3rd	EAS	40	38
	3rd	Isomerisation	49	61
	2nd	Macrocycles	37	46
	2nd	Panacetin	184	74
	2nd	Proteins (2016)	70	58
	2nd	Rearrangement	51	49
Post-TLL	3rd	Nylon	69	58
	2nd	Proteins (2017)	138	69
	2nd	Enzymes	28	47
	2nd	Food Project	97	49
	2nd	Sunscreen	51	49
	2nd	Electronic Waste	160	59
	1st	Pseudenol	65	63

2.6.2 Student focus groups

Volunteers for focus groups were identified via messages on the specific units’/courses’ Moodle news feed or through direct emails (teaching associates). The volunteers were invited to participate in an audio-recorded session in exchange for refreshments. In some cases, only one or two volunteers participated in the focus groups and these sessions were run as informal interviews. Student focus groups were asked three open questions pertaining to their laboratory experience over an entire semester:

1) Which laboratory experiences did you enjoy? Why?
2) Which laboratory experiences didn't you enjoy? Why?
3) What skills do you think you developed throughout these laboratory experiences?

The facilitators of the focus groups occasionally included members of the research team that the students may have known from their classes. However, if a researcher also acted as their laboratory Teaching Associate (and was therefore responsible for their marks) a replacement was found. The students were encouraged to be honest and to provide both positive negative feedback as necessary. They were also informed that any names would be redacted to encourage an honest discourse.

The validity of these questions was measured after the first focus group was completed. As the responses of students matched the desired focus, these questions were deemed to be valid. Focus groups were run a week after the completion of all lectures and laboratory exercises. The cohorts and the number of participants is shown in Table 6. Focus groups were only undertaken for units/courses in which more than half of the laboratory activities were generated or redeveloped by the TLL programme except for CHM2942, as this course was changed at the start of the TLL programme and no 'pre' focus group could be undertaken.

Table 6 - The number of participants per recorded focus group.

Pre-TLL		Post TLL	
Cohort (Year)	Number of participants	Cohort (Year)	Number of participants
CHM2911 (2016)	6	CHM2911 (2017)	8
CHM2962 (2016)	2	CHM2962 (2017)	5
CHM3180 (2016)	2	CHM3180 (2017)	3
CHM3922 (2016)	7	CHM3922 (2018)	7

2.6.3 Laboratory aims and expectations survey

In order to investigate student and staff perception of the aims of teaching laboratories and the actions, thoughts and feelings of students during a typical teaching laboratory, a survey was generated by combining a single open question 'What do you think the aims of a practical chemistry course are?' with the Meaningful Learning in the Laboratory Instrument (MLLI (Galloway & Bretz, 2015a)) as well as some demographic questions (age, gender, domestic/international enrolment and overall course enrolment). Other small changes were made to the MLLI survey such as the use of a Likert scale rather than an electronic slider and the modification of a distractor question to 'Please select Agree and Disagree for this question'. The survey (Appendix 5) was administered to all students, over year levels 1, 2 and 3, enrolled in chemistry units/courses during early semester 2 2016 (late July/August). The units/courses surveyed at this time were CHM1021, CHM1052, CHM2941, CHM2962, CHM3180, CHM3922, CHM3952 and CHM3972. The survey was typically handed out during induction periods where students were not under a significant time pressure and the students were informed the survey was both anonymous and non-compulsory. The same procedure was undertaken in early 2017 at the University of New South Wales, albeit at the end of a normal lab rather than during induction. At the University of Warwick, the survey was distributed during free lunches in late 2016 and early 2017 during full day laboratory sessions. Lastly, data from a similar study performed at the University of Sydney was collected but that version of the survey utilised a ten-point scale (compared to strongly disagree … strongly agree). The survey was also delivered to students at Monash University during early semester 2 in 2017 and 2018 (late July/August), but only to level 2 and 3 units/courses (CHM2922, CHM3952 and CHM3972). The number of responses from each university can be found in Table 7.

Table 7 - The number of responses to the laboratory aims and expectations survey.

Institution	*N*
Monash University (2016)	1334
Monash University (2017)	275
Monash University (2018)	203
The University of New South Wales	712
The University of Warwick	292
The University of Sydney	869

The same survey was also administered to teaching associates during an induction period just prior to semester 1, 2017. However, in the wording of the parent question for the MLLI portion of the survey ‘When performing experiments in a chemistry laboratory course, I expect …’, ‘I expect’ was changed to ‘I think the students’ and the following direction given to the teaching associates:

‘When filling out the following survey, please ensure you answer with what you think the students will actually be doing during any given practical exercise. Try not to answer with what you would like the students to be doing.’

Minor changes were also made to the questions themselves such as ‘I expect to be frustrated’ became ‘I think the students will be frustrated’. The teaching associates were informed the survey was both anonymous and non-compulsory. The survey was not distributed to teaching associates at the University of Sydney and the teaching associates at the University of New South Wales choose not to respond. The amended survey was also given to academic staff at Monash University and the University of Warwick alongside an electronic version (as a Google form) sent to a variety of academics in the UK and Australia. The total number of responses are given in Table 8.

Table 8 - The number of teaching staff responses to the altered laboratory aims and expectations survey as per university.

	Monash University	The University of Warwick	The University of New South Wales	Assorted Australian and UK universities
Teaching associates	111	32	-	-
Academic staff	13	10	12	67

2.6.4 Academic Interviews

Academic staff at Monash University, the University of New South Wales and the University of Warwick were also invited to participate in recorded one-to-one interviews. Academic staff were asked the exact same open question as the students and teaching associates (‘What do you think the aims of a practical chemistry course are?’) alongside one additional question – ‘With those aims in mind, do teaching laboratories at your institution succeed at meeting those aims?’). Typical interviews resulted in 10-25 minutes of recorded audio which were then transcribed before analysis. 13 interviews were recorded at Monash University, 12 at the University of New South Wales and nine recorded at the University of Warwick.

2.6.5 Student assessment marks and teaching associate feedback

To monitor changes to attempts to curtail marking variation between markers, student marks in first-year laboratories were collected *via* Moodle. The marks were downloaded as an Excel spreadsheet from Moodle and all identifying information immediately removed. An electronic Google form survey was also sent out to teaching associates in late-2017 and mid-2018 to collect their perspectives around these changes. 22 (out of 26) and 35 (out of 53) responses were collected, respectively. The closed questions are shown in Appendix 6 and each closed question was immediately followed by an open box and a prompt to the teaching associate to explain why they choose the answer that they did.

2.7 Data analysis

Once collected, the data was split into qualitative (i.e. open responses) and quantitative (frequency of responses) in order to analyse for further discussion.

2.7.1 Theme extraction from qualitative data

Qualitative data was generated from open questions on the surveys, recorded focus groups or one-to-one interviews (both with students and teaching staff). The data was treated under an inductive coding approach (Thomas, 2006), where themes were extracted from the answers given through multiple readings of the transcribed text and careful consideration of overlapping responses. Once a list of themes was collated, any themes that were seen to be redundant (i.e. they covered similar concepts) were merged together. These themes were then each assigned a two-letter code (e.g. TU for enhancing, consolidating or strengthening theoretical understanding) and then assigned to the same transcribed data used to generate the themes (e.g. survey responses), a process known as coding

(Saldaña, 2015). For example, consider the following two responses from first year Monash students when asked 'What do you think the aims of a practical chemistry course are?':

'Apply concepts being taught in lectures'

'Learn critical thinking skills'

The first option was coded to the theme 'to apply theory in the 'real world' or introduce applications' whilst the second was coded to the theme 'to enhance critical thinking, problem solving and other cognitive skills'. Once themes were extracted from the data, and preliminary coding completed by the original researcher, the same data set and themes were distributed to between three and six other chemical education researchers who were then asked to attempt to code the data themselves using the provided themes. This was done to measure inter-rater reliability which can be considered as the following, as written by Gwet (2014):

'The extent to which these … categorizations coincide represents what is often referred to as inter-rater reliability. If inter-rater reliability is high, then … raters[ii] can be used interchangeably without the researcher having to worry about the categorization being affected by a significant rater factor.'

Simply put, inter-rater reliability is an attempt to measure whether coding (and the themes used to do so) are highly dependent on the original researcher who raised them. In this case, the codes were only utilised if greater than 80 percent agreement was obtained (as per McHugh (2012) et al.) between the raters (other chemical education researchers). If this percentage agreement was not reached, the themes are revisited by the original researcher and modified until a higher percent of agreement was reached. It is worth noting that whilst there are statistical means to calculate other measures of inter-rater reliability (such as Cohen's kappa (Viera & Garrett, 2005), which is more laborious to calculate),

[ii] NB: In the context of this research, raters refers to the other chemical researchers asked to code the data with the previously derived themes.

McHugh states that ‘if raters are well trained and little guessing is likely to exist, the researcher may safely rely on percent agreement to determine inter-rater reliability’, which holds true in this case.

Lastly, when large numbers of responses were collected (such as to the open questions on the surveys) the number of times a given theme was coded in a given set of responses was then tallied to generate a final result (e.g. 10 out of 50 responses raised the application theme). In this case, comparisons between the frequency of a given theme raised by different cohorts or groups (e.g. first-year students versus 2nd year students) could be made to determine if the appearance of a given theme was dependent on specific cohorts. Hence, numerical data was extracted from the qualitative analysis which was analysed using the quantitative analysis protocol discussed in the next section.

2.7.2 Statistical testing of quantitative data generated from surveys.

Quantitative data from Likert surveys were treated as ordinal data except for the application Cronbach’s alpha (as calculated by the statistical analysis program - Statistical Package for the Social Science, or SPSS, version 23). Cronbach’s alpha is a measure of the internal consistency which, in this case, refers to how well the questions utilised all relate to the same topic or construct. A very internally consistent result, above the literature benchmark of 0.7 (Santos, 1999), indicates that few or no questions deviate from the topic being investigated. Typically, Cronbach’s alpha requires the use of continuous data and is known to underestimate the internal consistency of a series of questions which utilize a Likert scale (Gadermann, Guhn, & Zumbo, 2012). Whilst other measurements do exist to counteract this (for example Zumbo’s alpha (Zumbo, Gadermann, & Zeisser, 2007)), they are often laborious and require the use of highly specialised programs. Hence, Cronbach’s alpha was still utilised in this study as, at worse, it would only result in underestimating the internal consistency of the data collected.

Likert data was converted to values[iii] (Strongly disagree = 1, Disagree = 2, Neutral = 3, Agree = 4 and Strongly Agree = 5) and the frequency of a response tabulated per each item on the surveys. For visual analysis, these frequencies were also converted to an overall percentage of the number of participants to allow easy comparison between groups of different sample size. However, for statistical analysis, the raw frequency values were compared without further alteration in order to measure significant differences in responses between various categories or cohorts (e.g., 16-18 year-olds vs 19-21 year-olds). Before analysis of how specific cohorts responded to individual items on a survey, omnibus testing (in this case, through an *F*-test) was performed in order to measure if a cohort responded differently to ALL questions in the survey (e.g. if male-identifying students generally choose to disagree rather than specifically disagreeing to given item). If no significant difference was noted, then a comparison of how a cohort responded to each individual item could be considered. If a significant difference was noted, the relative 'importance' or size was measured through calculation of a strength of association. This test generates a value between −1 and +1, where the absolute values (i.e. the magnitude of the value rather than the sign) can be assigned to a specific 'effect size'. Here, eta-squared (η^2) was utilised and if the value was found to be <0.04, then further analysis was undertaken as if no significant difference was originally noted.

Once the *F*-test had shown that further analysis could be undertaken, non-standard tests (i.e. not *t*- or z-tests) were required due to the non-continuous and potentially non-parametric nature of the data. Hence, significant differences measured by SPSS were calculated using either:

1) A Pearson's chi squared test. This test is typically used to compare categorical data sets against one another (e.g. whether male or female identifying students are enrolled or not in a given course). However, this test can also be used for ordinal, ranked data where the difference

[iii] NB: The conversion into numbers was used to ease data entry - letters or other symbols could have also have been used at this stage. As such, these numbers were never directly processed (e.g. no average value was calculated).

between ranks is undefined (e.g., the difference between agree and strongly agree has no measurable value). The test also assumes the groups being compared are independent of one another (e.g., gender identity) and allows for non-parametric data (i.e. where the data is not necessarily normally distributed) (McHugh, 2013). The Pearson's chi squared test also requires that at least 20% of the possible items (e.g., strongly agree to question one) must have at least five responses. Violation of this assumption can be countered by the use of a less sensitive Fischer exact test (Graham, 1992), which measures significant differences regardless of the number of responses received. This test was used to measure the responses of cohorts to the laboratory aims and expectations survey and the workplace readiness survey.

2) A Wilcoxon test. This test holds the same assumptions as the Pearson's chi squared test except for the dependence of the compared groups (Lowry, 2014). Hence, this test was used to measure the responses of the same group of students as they experienced a variety of laboratory experiences.

In cases where there were large discrepancies between group sizes (e.g. 1000+ first year responses and only approximately 200 3rd year responses), a subset of the larger group was used (e.g. every fifth first year response) in order to ensure accurate statistical testing was achieved. Additionally, the traditional use of a 95% confidence interval (or p=0.05) was complicated whenever a large number of items were present in the surveys (24-30 questions in total) which could readily result in a Type I error (i.e. a false positive). Hence, in these cases, the Holm-Bonferroni correction was utilised which typically changed the value to a 99.5% confidence interval (or p=0.005). With regards to effect size for the Pearson's chi squared test, Cramer's V (Sheskin, 2003) (φ_c) was utilised which was originally generated with specific ranges as shown in Table 9.

Table 9 - The original ranges for the φ_c measurement of effect size.

φ_c range	Effect size
0.100-0.300	Small effect
0.301-0.500	Medium effect
>0.501	Large effect

However, it is possible that these ranges underestimate the effect of many interventions measured. Subsequent work performed by Hattie (2008), specifically focused on Cohen's *d*, indicated that these values should in fact be considerably lower. This was later extended (Fritz, Morris, & Richler, 2012; Lenhard & Lenhard, 2016) to *r* and, as the original ranges for *r* matched those for φ_c (Cohen, 1988), the same altered ranges were used in this analysis. As such, the ranges were redefined and are shown in Table 10. For the Wilcoxon test, *r* was instead utilised but matched the same ranges provided in Table 10.

Table 10 - The altered ranges for the 'effect size'

φ_c range	Effect size	Explanation
0-0.100	'Student' effect size	This refers to the natural variation in any group of students. For example, a more motivated student may respond more positively than a less motivated student.
0.101-0.200	'Teacher' effect size	This refers to the effect of a particularly motivated teacher over the course of a single year (i.e. this effect size could be achieved given time/motivation).
>0.201	Zone of desired effect	This refers to interventions that have an immediate impact and are where educators should typically focus their efforts.

The ranges shown in Table 10 were utilised either when the frequency of raised themes was being compared (from responses to open questions) or the responses from the closed question were dependent (i.e. when the same group/cohort were responding to old and new laboratory activities). These ranges were not used for the closed questions between independent groups, however, as the degrees of freedom (df) further complicates the use of φ_c, whose cut-off values are dependent on the degrees of freedom utilised. The degrees of freedom was determined as four options on a Likert survey minus one (in this case, 5 - 1 = 4) multiplied by the number of categories being compared minus one (which was always two in this study, thus 2 - 1 = 1), and so df = 4 × 1 = 4). The corrected effect sizes are shown in Table 11.

Table 11 - The ranges of possible φ_c values and the associated effect sizes.

φ_c range	Effect size
0.050-0.150	Small effect
0.151-0.250	Medium effect
>0.251	Large effect

2.7.3 Statistical testing of quantitative data generated from student assessment marks.

Before analysis, students' marks were assigned into the same groups that they were organised into for their weekly laboratory classes. This data was then taken from Excel and analysed in SPSS to generate an average mark for each group of students. This data was imported back into Excel where each group average was noted to either be 'high' (i.e. 10% above the entire cohort average) or 'low' (i.e. 10% below the entire cohort average). The number of times a group average was 'high' or 'low' was tallied and split into four sub-groups:

1) Baseline. The sub-group of results before any interventions took place.
2) Increased Detail. The sub-group after the amended training programme was implemented and enhanced marking criteria were distributed.
3) Excel. The sub-group that utilised the Excel spreadsheet for marking.
4) Moodle. The sub-group that were assessed using the automated Moodle marking system.

Whether or not 'high' or 'low' averages were significantly different between the subgroups was determined using a Pearson's chi squared test (p=0.05) with the calculation of φ_c used for measurement of the effect size. The effect size ranges used were those values shown in Table 11.

2.8 Limitations

The internal bias within the framework of the project (i.e. the goal of TLL to increase the employability of the students) impacted the questions asked in surveys, focus groups and interviews. Whilst attempts were made to include broad questions that covered as many aspects of the laboratory environment as possible (such as using surveys that contain questions beyond the scope of employability), the bias cannot be fully ignored.

Validity and reliability were not measured directly but should be accounted for in the use of multiple data collection techniques such as surveys (three different types), focus groups and interviews. This process is known as triangulation (Given, 2008) and is routine in social science. The broad range of measurements provided plentiful data that highlighted any irreproducible themes extracted from the analysis. Furthermore, the surveys utilised were based directly on already tested and validated instruments (the MLLI survey and the Russell survey) with any alterations made unlikely to have affected these previous tests.

The next consideration is whether the data collected is truly representative of the cohorts being investigated. When surveys were utilised, the response rate was generally between 30-80% indicating a large degree of variability in the percentage of students responding. This variability in sample size can be countered through a consideration of the work of Barlett, Kotrlik, and Higgins (2001) who provided acceptable sample sizes to statistically represent various populations. At Monash University the population sizes considered varied from approximately 250 students (enrolled in 3rd year units/courses) to almost 1300-1400 enrolled in first year units/courses, which required at least 80 and 110 responses respectively to be considered statistically representative at the 5% level of significance. In all cases studied throughout this work, the sample sizes were always above these acceptable levels before analysis was undertaken.

It is also worth discussing the way in which the surveys were delivered to students and teaching staff. The surveys were always anonymous and tended to be given to students and teaching staff in paper format, either in times where few other time pressures existed (e.g. during an induction) or where food was provided during a scheduled break in class times. Whilst students and teaching staff were strongly encouraged to participate, all respondents were informed that the survey would have no impact on either their academic standing or their employment at Monash University.

The issue of which cohorts were selected to be surveyed is another limitation of this work, particularly when considering individual laboratory activities before and after the TLL programme. The experiments chosen to represent the 'Pre-TLL' experiments were chosen due to convenience and timetabling. Hence, it is possible that different results may have been obtained if other experiments were investigated. Furthermore, when large cohorts were an issue, a cross-section of the cohort was surveyed to ease data collection. This is a process known as convenience sampling (Henry, 1990) and could potentially have led to less statistically valid data or non-representative data.

The last point to cover is which students chose to participate in the recorded focus groups. The students were invited to a free lunch/dinner either through an electronic message on the respective unit/course Moodle page and anyone could choose to respond. Hence, this typically resulted in the most engaged students participating in the process. The sample sizes are small (2-8) and cannot be considered truly representative of their respective cohorts. However, as this data was compared to the much larger number of survey responses, this data is considered valid through comparison (i.e. through triangulation).

For the quantitative data, the major issues result from the statistical tests chosen to measure significant differences between cohorts. The use of either the F-test or the Pearson's chi squared test on the ordinal data resulting from the Likert surveys does not match the data type these tests are intended for (continuous and categorical, respectively). However, no such test currently exists that is specifically designed for ordinal data, so the choice of any given statistical test will come with this same issue.

Finally, it must be noted that the majority of this work (with the exception of the baseline studies in discussed in Chapter 3) has taken place at a single, Australian university – Monash University. Hence, it is not possible to generalise these results to all other institutions. Furthermore, the Australian university model also complicates matters with students able to enrol in a large number of non-chemistry units/courses, which results in students of vastly different backgrounds and interests in any given chemistry unit/course. This also means that later analyses which focus on students enrolled in units/courses designed for a given year level are likely populated by a small proportion of students who are taking the unit/course out of sequence (e.g. a 2nd year level student enrolled in a 3rd year level course, or vice versa). Whilst this cannot be fully accounted for, it was considered that the number of students proceeding through the units/courses in the intended pathway (e.g. a 2nd year level student enrolled in a 2nd year level course) was high enough to allow for the analysis undertaken.

Chapter 3 - Student and teaching staff perceptions of laboratory aims and expectations before the large-scale inclusion of impact inquiry/problem/context/industry-based experiments.

The first research questions to be considered were 'What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' perception of laboratory aims and their expectations of their thoughts, actions and feelings during laboratories? How does this compare to the perception of teaching associates and academics?' (Page 30, Research Question 1). In order to achieve this, a survey was generated that contained a single open question (What do you think the aims of a practical chemistry course are?) and 31 closed questions from the Meaningful Learning in the Laboratory Instrument (Galloway & Bretz, 2015a) (Appendix 5).

The survey was used to gather data from 1917 students and 118 teaching associates across three universities (Monash University and the University of New South Wales in Australia and the University of Warwick in the UK). For the open question, the responses of 34 academic members of staff at the same three universities were collected during interviews. For the closed questions, an additional 873 student responses were collected from another Australian university, the University of Sydney. Furthermore, another 68 responses from academic staff over a range of universities in Australia or the UK were collected through an electronic form of the survey that just contained the closed questions alone. These responses were collected through a broadcast message to colleagues through personal contact lists.

The results of these surveys have been analysed and published as two peer-reviewed journal papers, one describing the findings or the data generated by the open question (Stephen R. George-Williams et al., 2018) results and another describing the findings of the analysis of the responses to the closed questions (George-Williams et al., 2019). Both papers are included in the following pages.

Chemistry Education Research and Practice

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Educ. Res. Pract.*, 2018, **19**, 463

'What do you think the aims of doing a practical chemistry course are?' A comparison of the views of students and teaching staff across three universities

Stephen R. George-Williams, [a] Angela L. Ziebell, [a] Russell R. A. Kitson, [b] Paolo Coppo, [b] Christopher D. Thompson *[a] and Tina L. Overton *[a]

The aims of teaching laboratories is an important and ever-evolving topic of discussion amongst teaching staff at teaching institutions. It is often assumed that both teaching staff and students are implicitly aware of these aims, although this is rarely tested or measured. This assumption can lead to mismatched beliefs between students and teaching staff and, if not corrected for, could lead to negative learning gains for students and become a source of frustration for teaching staff. In order to measure and identify this gap in a manner that could be readily generalised to other institutions, a single open question – 'What do you think the aims of doing a practical chemistry course are?' – was distributed to students and teaching staff at two Australian universities and one UK university. Qualitative analysis of the responses revealed that students and teaching staff held relatively narrow views of teaching laboratories, particularly focusing on aims more in line with expository experiences (*e.g.* development of practical skills or enhances understanding of theory). Whilst some differences were noted between students at the three institutions, the large amount of similarities in their responses indicated a fairly common perception of laboratory aims. Of the three groups, academics actually held the narrowest view of teaching laboratories, typically neglecting the preparation of students for the workforce or the simple increase in laboratory experience the students could gain. This study highlights gaps between the perceptions of students and teaching staff with regards to laboratory aims alongside revealing that all three groups held relatively simplified views of teaching laboratories.

Received 15th September 2017,
Accepted 18th January 2018

DOI: 10.1039/c7rp00177k

rsc.li/cerp

Introduction

The role of laboratory instruction in teaching science has been a topic of discussion for well over a century. Indeed, the argument of whether or not teaching laboratories should focus on theory or practical skills dates as far back as the early 1800s when Michael Faraday produced a book titled *Chemical Manipulation* (Faraday, 1830). It was unique for its time as it 'did not discuss mathematical equations, argue for new chemical laws, or try to interrelate experimental data and theoretical ideas' (DeMeo, 2001). It is important to discuss the aims of teaching laboratories as, in the words of Shah (2007), 'It would be rare to find any science course in any institution of education without a substantial component of laboratory activity'.

[a] *School of Chemistry, Monash University, Victoria, 3800, Australia. E-mail:* [redacted]
[b] *Department of Chemistry, University of Warwick, Gibbet Hill, Coventry, CV4 7AL, UK*

Originally, teaching laboratories in the mid-1800s typically aimed to teach technical skills required for industry and research (Good, 1936; Elliott *et al.*, 2008). This has changed over time, becoming more complex and including more diverse aims (ICSU-CTS, 1979; Shymansky and Penick, 1979; Kirschner and Meester, 1988; Linn, 1997; Johnstone and Al-Shuaili, 2001; Reid and Shah, 2007) which can be broadly considered as (but not limited to):

(1) A chance for students to learn science in a more tactile, engaging way.
(2) Complementing underlying scientific theory.
(3) Developing technical skills.
(4) Imparting scientific methodology.
(5) Enhancing transferable/soft skills (communication, time management, *etc.*).

The importance of communicating these aims to students is well known and forms the basis of the passionate plea of Reid and Shah (2007) – 'There is a need for a clarification of aims and objectives, and these need to be communicated to learners'. Significant strides towards measuring the perceptions of academic staff around the aims and goals of teaching laboratories has been undertaken in the US. They have been investigated either through interviews (Bruck *et al.*, 2010; Bretz *et al.*, 2013) or the Faculty Goals Survey (Bruck and Towns, 2013; Bretz *et al.*, 2016).

The interviews (Bruck *et al.*, 2010; Bretz *et al.*, 2013) highlighted that academic staff held a range of aims ranging from the development of transferable skills (such as teamwork, independence, critical thinking and scientific communication), the development of practical skills and imparting theoretical understanding. The results of the survey (Bruck and Towns, 2013; Bretz *et al.*, 2016) identified aims that appeared to be course dependent. For example, when comparing organic chemistry to general chemistry, laboratory writing was rated more important, whilst teamwork was less so. It was also noted that the stated aims changed with year level.

It is important to identify the perspective of the students themselves, as their positive perceptions of teaching laboratories are closely linked with their success with respect to both affective and cognitive learning goals (Rentoul and Fraser, 1979; Fraser, 1981). Several studies have investigated the expectations of students towards how they will think and feel during a laboratory exercise, particularly through the lens of Novak's theory of meaningful learning (Galloway and Bretz, 2015a, 2015b; Galloway *et al.*, 2016). Whilst the underlying cognitive and affective expectations of the students were successfully mapped, these studies did not actively address what students' perceive the underlying aims of the teaching laboratory to be.

The student held beliefs of the importance of a range of identified laboratory aims was investigated by Boud *et al.* (1980). It was found that chemistry students perceived aims such as the development of practical skills or connection to theory as more important than the development of problem solving skills or 'the use of labs as a process of discovery'. These results were compared to both graduates and practising scientists and employers, who focused more on developing observational skills and critical awareness, with enhanced student–teacher relations and enhancing engagement as less important. It is worth noting that this study only focused on a single institution and potentially restricted students to a given list of aims rather than allowing the students to raise their own.

In 2008, an effort was made to avoid leading the students through 13 interviews with university undergraduate students (Russell and Weaver, 2008). A grounded theory approach was taken to the analysis of the students' responses after they were asked about their laboratory experiences and what they perceived the point of them to be. Their responses indicated that a significant mismatch existed between the opinions of teaching staff and students, with students primarily focused on merely completing the task for assessment purposes to the exclusion of all else. Studies can also be found in the secondary schools (Lynch and Ndyetabura, 1983; Wilkinson and Ward, 1997), indicating mismatches between students and teaching staff around the perceptions of the aims of laboratory activities.

A more recent study (DeKorver and Towns, 2015) sought to investigate students' personal goals throughout a given laboratory experience. Video recordings of laboratory sessions were collected and supplemented with interviews with students. It was noted that students tended to focus on 'affective goals', namely through finishing the required tasks in a short time period. This was found to be at odds with any psychomotor or cognitive goals that the students may have held. Whilst important, this study was undertaken with a limited number of participants in general (or first year) chemistry and is, therefore, difficult to generalise to a large population. Furthermore, the students' responses were focused on a single experiment, rather than on all laboratory experiences.

There are few studies that directly assess or measure the perceptions of students and teaching staff of the aims of teaching laboratories on a large scale beyond the US context. Hence, this study sought to investigate the contemporary perceptions of the aims of teaching laboratories held by:

(1) students at all year levels at three different institutions. This ensures a large scale, international study that may potentially be generalised to other contexts.

(2) teaching associates and academic members of staff at three institutions.

Method

This study was undertaken at three institutions, two within Australia (Monash University and the University of New South Wales) and one in the United Kingdom (the University of Warwick). The aim of this study was to compare students' and teaching staff's perceptions of the aims of practical laboratory activities within degree programmes. This was investigated using a qualitative analysis of responses to either a paper-based survey or an audio-recorded one-to-one interview. Further funding is also gratefully acknowledged from the Monash Warwick Alliance Seed fund. Ethics approval was obtained at Monash University and was accepted by the ethics approval boards and the other two institutions.

Data collection

Undergraduate students and teaching associates (sometimes referred to as laboratory demonstrators) were asked to answer a single open question – 'What do you think the aims of doing a practical chemistry course are?'. The survey also included some general demographic questions about age and gender, and domestic/international enrolment for students, or the amount of teaching or industry experience for teaching associates.

It was made clear to both teaching associates and students that the survey was not compulsory and would not affect either their academic standing or employment in anyway.

At Monash University, all students enrolled in chemistry courses, at any year level, were given the opportunity to complete the survey. In total, two first year courses, three second year courses and four third year courses were surveyed. There were 1600–1800 students enrolled in these courses in mid-2016. Overall responses rates varied from Teaching associates were asked to complete the survey at the end of a compulsory training session in early 2017. There were approximately 120 teaching associates present who taught across all year levels.

At the University of New South Wales (UNSW), students were asked to complete the survey at the end of their first teaching laboratory in semester 1, 2017. Unlike Monash University, access was not readily available to all chemistry courses so three second year and two third year courses were surveyed. The number of enrolled students was also 1600–1800 students but only 1300–1400 received the survey. Approximately 120 teaching associates at UNSW received the survey during their compulsory training session in early 2017. These particular teaching associates only taught at the first-year undergraduate level.

Responses from students at the University of Warwick were obtained during three separate events where a free lunch was provided, scheduled on days when students were undertaking laboratory exercises early in their academic year (late 2016). Third year students were not surveyed until May of 2017 due to scheduling commitments. There were approximately 490 students enrolled in the first three year levels at the University of Warwick at the time of data collection. Approximately 60 teaching associates were encouraged to complete the survey during the same laboratory sessions as the students.

Lastly, academic members of staff at all three institutions were asked the same single question as students and teaching associates, but during an audio-recorded one-to-one interview. Another question, 'With those aims in mind, do teaching laboratories at your institution succeed at meeting those aims?' was also raised during this interview. Staff were approached *via* email and were not compensated in any way.

Research theoretical framework

The primary theoretical framework underpinning this study is Constructivism which postulates that learning constantly evolves and is heavily reliant upon day-to-day experiences (Leal Filho and Pace, 2016). Hence, in the case of a respondent's perceptions of the aims of a laboratory programme, it is postulated that their responses will be mediated by their previous experiences. Therefore, the responses from participants to being asked to reflect on the purpose of laboratory learning will be as a result of any prior understanding that they have built for themselves. The non-leading nature of the open question ensures the respondent will draw from their own personal experiences and understanding rather than being prompted by the survey, the researchers or by interactions with others at the time of the response.

Data analysis

Of the surveys and interviews actually completed, 1917 undergraduate students (1108 from Monash University, 523 from UNSW and 283 from the University of Warwick), 118 teaching associates (91 from Monash University and 26 from the University of Warwick) and 34 academic members of staff (13 from Monash University, 12 from UNSW and nine from the University of Warwick) were transcribed verbatim. Typically, 40–75% of any cohort completed the survey. Whilst courses at Australian Universities are typically designed for students of a given year level, they are actually open to students of all year levels. Therefore, whilst for example most students in a second year course would be second year students, there may be some variation. The data were then analysed for emerging themes using the qualitative analysis program NVivo (version 11.3).

Analysis was first attempted on the largest dataset, which was students enrolled in first year at Monash University (n = 782). In order to ensure that the results were unaffected by the bias of a single researcher, themes extracted from the data by a sole researcher were then used by a team of six chemical education researchers to code a sample of the transcribed data (50 responses). Throughout this process, subtle differences in themes were noted requiring the addition of some new themes or splitting of existing large themes (such as students raising the development of specific practical skills *versus* simply becoming accustomed to the overall laboratory environment, which were both originally assigned to the development of practical skills).

The themes were subsequently reconsidered and revised and were again used to code the data, this time by three researchers over two iterations. This resulted in an inter-rater reliability (*i.e.* the percentage of times multiple researchers/raters choose the same theme) of greater than 90%. Hence, the values in this article may vary by up to 10%. These final themes were then used to code the rest of the responses from all three institutions.

All transcribed responses were coded to the themes generated and NVivo provided the number of participants who raised a particular theme. These data were then expressed as a % of participants who raised the theme and were presented graphically. In order to determine the significance of any differences between either universities, year levels or demographics, the coded data was analysed with SPSS using a Pearson's Chi Squared test to ensure differences held to the 95% confidence interval (*i.e.* $p < 0.05$). Cramer's V was calculated in order to measure the effect sizes of any differences (small, 0.05–0.25, medium, 0.25–0.5, or large, >0.5).

Results and discussion

The cohorts of students who responded to the survey and their demographic data are shown in Table 1. From the analysis of the largest subset of the qualitative data (Monash first years), and the subsequent inter-rater reliability tests, 11 major themes describing the aims of teaching laboratories emerged from the responses (Table 2). An additional theme labelled as 'other/unassigned' was used for a large variety of responses that were considered either nonsensical or irrelevant. The percentage of respondents raising a theme is shown in Table 3.

First year students

The most common aims raised by first year undergraduate students at each of the institutions are shown in Fig. 1. The six aims, TU (enhancing theoretical understanding), AP (application of theory), PS (developing practical skills), EX (gaining general laboratory experience), WF (preparation for the workforce) and TS (developing transferable skills), were all identified by at least 10% of the cohort (and therefore considered to be meaningful)

Table 1 The demographic breakdown of the students who responded to the survey at Monash University, UNSW and the University of Warwick

		1st year students	2nd year students	3rd year students
Monash University (*N*)		782	187	139
Gender (%)	M	46	51	60
	F	52	47	37
	Other	2	2	2
Age (%)	16–18	46	6	1
	19–21	49	85	71
	22+	5	9	28
Enrolment (%)	Domestic	92	92	90
	International	8	8	10
		1st year students	2nd year students	3rd year students
UNSW (*N*)		368	107	53
Gender (%)	M	52	43	51
	F	48	57	45
	Other	0	0	4
Age (%)	16–18	65	12	0
	19–21	29	79	80
	22+	6	9	20
Enrolment (%)	Domestic	88	75	85
	International	12	25	15
		1st year students	2nd year students	3rd year students
University of Warwick (*N*)		148	58	77
Gender (%)	M	45	49	54
	F	55	51	46
	Other	0	0	0
Age (%)	16–18	51	0	0
	19–21	48	100	70
	22+	1	0	30
Enrolment (%)	Domestic	90	95	90
	International	10	5	10

at one or more of the three institutions. The other six aims, CX (contextualising theory), EG (enhancing student engagement), SR (imparting safety/responsibility), SM (developing scientific methodology), LV (learning by varied means) and OT (other) were raised by too few students to be considered relevant in this case.

Possible demographic effects in the data were considered, such as gender identity, age and domestic or international enrolment. To ensure enough responses in each sub-group, these effects were investigated in the largest cohort, the Monash University first year undergraduates.

Gender identity was noted to have minimal effects upon the data, with only two themes showing any significant differences. These were more female-identifying respondents raising application of theory (AP, $p = 0.002$, small effect, $V = 0.114$) and more male-identifying respondents raising developing practical skills (PS, $p = 0.008$, small effect, $V = 0.096$). Neither age (determined through the use of categories, 16–18, 19–21 or 22+) nor domestic/international enrolment showed statistically significant differences. It should be noted that as only 63 international students responded, this result could underestimate differences in their viewpoints. Overall, demographic effects were considered minimal and likely to be so in all other comparisons made throughout this study.

The fact that five of the six major aims (application of theory, gaining general laboratory experience, enhancing theoretical understanding, developing practical skills and preparation for the workforce – AP, EX, TU, PS and WF) were raised by more than 10% of the students who responded in each institution indicates that this result is likely to be generalisable between Australia and the UK as it appears independent of university, country, culture or prior schooling. This would imply that students generally appreciate that teaching laboratories aim to (a) provide a chance to apply theoretical

Table 2 The themes generated through inductive analysis of the responses from first year students at Monash University

Code	Theme	Examples
TU	Aid in theoretical understanding, retention and consolidation.	'Consolidating what was taught in class', 'assist in helping students learn concepts in the course'
AP	Allow students to apply, use or visualise theory.	'Applying what we've learnt in lectures', 'To put into use theory in lecture'
WF	Prepare students for their future careers in industry or research.	'Get a job', 'In order to help prepare us for future careers in laboratories in industrial and other areas'
EX	Increase students' practical or laboratory experience/exposure/confidence.	'Provide practical experience', 'Becoming used to working in a laboratory-based environment'
PS	Enhance students' practical laboratory skills and equipment/instrument use.	'Build up practical skills', '...learning lab techniques'
TS	Enhance students' general transferable skills.	'To develop critical thinking and analysis skills', 'Getting proficient in doing... experiments within a certain time limit'
CX	To contextualise theory in the real world.	'Familiarising us with real-world chemistry', 'applying this knowledge to everyday life'
EG	Enhance students' engagement with the subject.	'Making it more interesting', 'To make the content more engaging...'
SR	Enhance students' understanding and practice of safety and responsibility.	'To familiarise myself with safe lab procedures...', 'Gain practical knowledge in basics of working safely in labs'
SM	Enhance students' understanding of scientific methodology.	'Understand the methodology of a scientific experiment', 'understand methods for conducting experiments...'
LV	Allow students to learn chemistry in different ways tactile/hands-on, visual, *etc.*	'Applying concepts learned... in a 'hands-on' way', 'hands-on learning'
OT	Other/unassigned.	Varied-forced requirement, desire for marks, negative comment, advancement of science or unexplainable.

Table 3 The percentages of students raising a given theme as per university and year level

Codes	Monash University 1st year (%)	Monash University 2nd year (%)	Monash University 3rd year (%)	UNSW 1st year (%)	UNSW 2nd year (%)	UNSW 3rd year (%)	The University of Warwick 1st year (%)	The University of Warwick 2nd year (%)	The University of Warwick 3rd year (%)
Top six aims raised									
TU	49	32	31	39	34	28	43	34	32
AP	30	28	34	31	49	36	31	21	35
WF	20	30	30	20	32	30	42	41	34
EX	22	37	32	27	13	23	31	24	19
PS	23	36	47	38	36	50	49	53	60
TS	12	16	25	7	14	11	29	50	60
Bottom six aims raised									
CX	5	5	8	4	0	8	3	3	6
EG	5	0	3	5	2	6	5	7	3
SR	5	5	6	7	3	6	4	2	9
SM	4	2	1	2	1	2	0	0	3
LV	5	6	1	3	3	4	3	7	3
OT	7	3	3	2	3	10	0	0	6

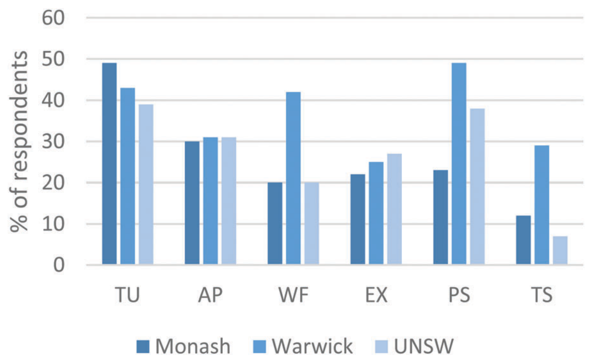


Fig. 1 The percentage of first year students raising the six most common themes at all three institutions.

knowledge, (b) provide general laboratory experience, (c) enhance and consolidate theory, (d) impart technical skills and (e) prepare students for the workforce.

Interestingly, the sixth major aim, enhance transferable skills (TS), was only raised by more than 10% of the respondents at two of the institutions, Monash University (by 12% of the cohort) and the University of Warwick (by 29% of the cohort). This large variation would suggest that the students' perception of the importance of developing transferable skills is more variable than the other skills commonly identified. This may be the result of the different university systems in Australia and the UK. In Australia, students undertake a far more generalised degree compared to the focused nature of the UK system. Hence, UK students may be more focused on developing skills required in the workforce. Another possibility is that the UK higher education system is simply more focused on the development of employability skills in higher education, particularly as a result of the Dearing report (Dearing and Education, 1997), which highlighted the need for these skills approximately 30 years before this study.

There were some subtle differences between the prevalence of the common aims raised by students from different institutions. The most common aim identified by the Monash University cohort was enhancing theoretical understanding (TU, ~49% of respondents) and indicated that, for these students, the main aim of teaching laboratories was the enhancement of their theoretical understanding. After TU, the appearance of the other aims decreases, ending with the least prominent aim (developing transferable skills, TS) at about 12% of the total number of respondents. It would seem that these students did not consider the development of transferable skills as a notable aim of teaching laboratories. Furthermore, even with employability becoming a major focus of many universities for many years (Taylor, 1986; Boden and Nedeva, 2010; Bennett *et al.*, 2015), the WF aim was raised by only 20% of the students. Even something that may be considered fundamental to working in a laboratory, such as PS, was only raised by 23% of students. These students had experienced one semester of university chemistry teaching laboratories, so these results are unlikely to be a result of inexperience with the university system.

A Pearson's Chi squared test was used to measure the differences between the responses from students at all three institutions. When comparing the responses from UNSW to the Monash University students, the enhancing theoretical understanding (TU) aim was less emphasised ($p = 0.002$, small effect, $V = -0.091$) and the gaining general laboratory experience (EX) aim and developing practical skills (PS) aims are more prominent (small effects, $p = 0.020$, $V = 0.069$ and $p < 0.0005$, $V = 0.157$ respectively). This result is potentially an artefact of the delivery of the laboratory activities at UNSW, which assess practical 'core skills' thereby placing more emphasis upon the practical environment and the skills developed within it. There was no significant difference between the two universities for the AP and WF themes.

When considering responses from the University of Warwick, all six aims were raised by more than 20% of the respondents, with three aims (developing transferable skills, developing practical skills and preparation for the workforce – TS, PS and WF) being raised by more than 40%. The prevalence of the TS, PS and WF aims was significantly different to Monash University ($p < 0.0005$, small-medium effects, $V = 0.208$, 0.213

and 0.193 respectively). These results indicate that students at the University of Warwick begin their undergraduate careers with a broader view of the aims of a laboratory program. They appear to see the development of practical skills, the enhancement of theoretical understanding or preparation for the workforce as equally valid. These aims are then followed to a lesser extent by the application of theory, gaining general laboratory or practical experience and the development of transferable skills. It is difficult to say if this is due to the UK university system compared to the Australian university system, but it is possible that the students at Warwick, by being required to choose their specialisation so early, are simply more engaged and have more appreciation for the learning potential within chemistry teaching laboratories. Australian students often take first-year chemistry subjects as required elements for other degree paths and it could be this difference between the two systems that is responsible for the different views of the students.

Overall, this study suggests that students identify very similar aims of laboratory programs in chemistry regardless of education system. However, the degree to which they focus on the individual aims can vary.

Higher year students

To measure the impact of their time at University, the responses of second and third-year students were gathered and are shown in Fig. 2–4. As before, only aims that were raised by more than 10% of the students appear in the graphs.

For all institutions, the six main aims (enhancing theoretical understanding, developing practical skills, application of theory, gaining general laboratory experience, preparation for the workforce and developing transferable skills – TU, PS, AP, EX, WF and TS) were generally raised more than 10% of the time in any year level (except for TS from UNSW first-years). Hence, it is reasonable to suggest that a typical student will expect to be addressing these aims throughout their teaching laboratory experiences over their undergraduate careers. Of the six main aims, three (application of theory, gaining general laboratory experience and preparation for the workforce – AP, EX and WF) appeared to show erratic changes that did not correlate with the year level of the students. The other three main aims showed clearer trends, with the enhancing theoretical understanding (TU) aim decreasing in prominence whilst the developing practical skills (PS) and developing transferable skills (TS) aims increased. Hence, it would seem that as they progress through their degrees students began to focus less on developing theoretical understanding and more on obtaining practical and transferable skills. The statistical significance of these changes were evaluated using a Pearson's Chi squared test.

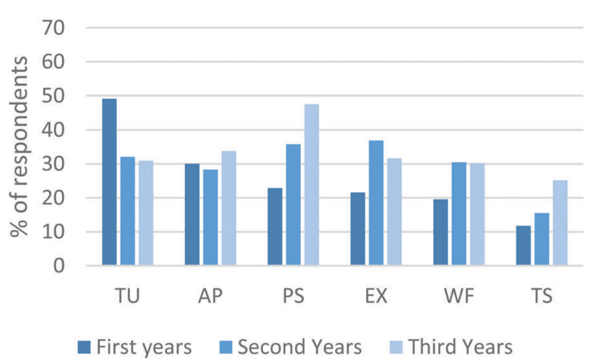


Fig. 2 The percentage of students at Monash University raising one of the six main aims.

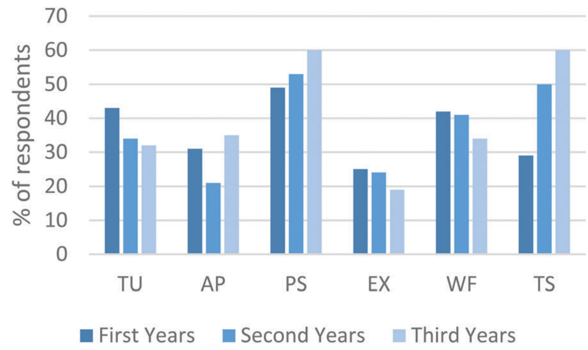


Fig. 3 The percentage of students at the University of Warwick raising one of the six main aims.

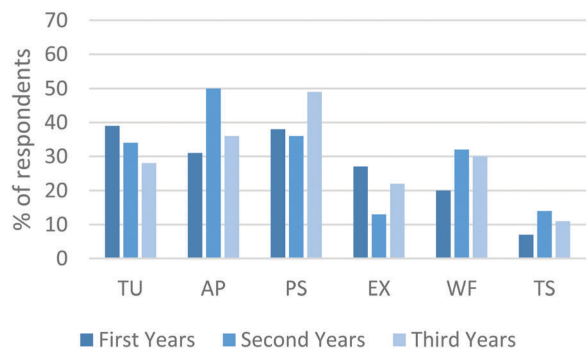


Fig. 4 The percentage of students at UNSW raising one of the six main aims.

For Monash University the appearance of the enhancing theoretical understanding (TU) aim significantly decreased ($p < 0.0005$, $V = -0.129$) whilst the developing practical skills (PS) ($p < 0.0005$, $V = 0.201$), gaining general laboratory experience (EX) ($p = 0.007$, $V = 0.089$) and developing transferable skills (TS) ($p = 0.001$, $V = 0.121$) aims significantly increased (small-medium effects). For UNSW, none of the changes were significantly different between year levels. Lastly, for the University of Warwick, the only significant change was the large increase in the amount of students raising the TS aim (medium effect, $p < 0.0005$, $V = 0.276$).

The apparent lower level of change at UNSW and the University of Warwick could be due to a lesser impact of the laboratory program (compared to Monash University) on the students' perceptions of laboratory aims. Regardless, the constant appearance of the six main aims (with the exception of the developing transferable skills (TS) aim from the UNSW first years) alongside the consistent trends noted, implies that these results could be generalisable to many other universities.

These results also appear to match those reported by Boud *et al.* (1980) with students more focused on practical skills and theory development/connection as opposed to transferable skills, such as problem solving. Interestingly though, the responses of these students contradicts the results of Russell and Weaver (2008), with students showing appreciation for aims outside of simple assessment. Overall, this large scale study would appear to show students have a broader view of laboratory aims than the Russell and Weaver study, but the same somewhat narrow view highlighted in the original survey by Boud *et al.* (1980). Additionally, these results differ from those noted by DeKorver and Towns (2015), with students freely raising a range of aims beyond the narrow view of simply finishing the experiment on time or to complete forced assessment. Furthermore, many of the additional cognitive or psychomotor aims were raised with no more than the original prompt further highlighting their importance in the students' minds.

It is worth noting that understanding of the scientific method, or the ability to plan and undertake an experiment, are themes that were never raised by more than 10% of the responding students. This may highlight a need for either a greater number of experiences that focus on these aims (*e.g.* inquiry or discovery experiences) or a more overt conversation with the students to emphasise the importance of these aims.

Teaching associates

The perceived aims of the laboratory experience as viewed by teaching associates could significantly impact on the engagement and overall learning of the students. As per Dobson *et al.* (2012), teaching associates 'set the tone for the type of learning that goes on in the laboratory'. Tables 4 and 5 show the demographic breakdown of the teaching associate cohort as well as the number of times a given theme was raised (Table 6). Note that there were too few responses from the UNSW teaching associates to be considered representative, hence this set is not shown.

Fig. 5 shows the results when teaching associates were asked the same open question as the students. As before, only aims that were raised by more than 10% of the teaching associates are shown, with two exceptions shown for comparison.

Table 4 The demographic breakdown of the teaching associates who responded to the survey at Monash University

Monash University		
N		91
Gender (%)	M	54
	F	45
	Other	1
Age (%)	19–21	15
	22–24	33
	25+	52
Teaching experience	< one year	44
	≥ one year	56
Industry experience	< one year	69
	≥ one year	31
Occupation	Postgraduate student	75
	Other	25

Table 5 The demographic breakdown of the teaching associates who responded to the survey at the University of Warwick

University of Warwick		
N		26
Gender (%)	M	57
	F	40
	Other	3
Age (%)	22–24	50
	25+	50
Teaching experience	< one year	30
	≥ one year	70
Industry experience	< one year	78
	≥ one year	22
Occupation	Postgraduate student	100

Table 6 The number of responses (converted to percentages) of teaching associates raising a given theme by university

	Monash University		University of Warwick	
Codes	*N*	%	*N*	%
TU	23	25	10	38
AP	37	41	9	35
CX	2	2	0	0
WF	18	20	8	30
EX	21	23	4	15
PS	57	63	15	58
TS	39	43	2	8
EG	5	5	2	8
SR	13	14	2	8
SM	0	0	1	4
LV	0	0	0	0
OT	10	10	0	0

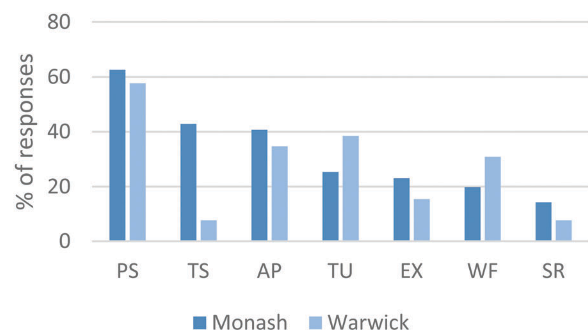


Fig. 5 The percentage of teaching associates raising one of the seven main aims by university.

As with the student data, any potential demographic effects on the responses of the teaching associates was first considered, such as gender, prior teaching experience or time spent working in industry. Only the Monash data allowed for this, due to a smaller number of responses from Warwick University. With regards to gender identity, variations were found in the developing practical skills (PS) aim (27% more male identifying responses, medium effect, $p = 0.010$, $V = 0.271$) and the application of theory (AP) aim (27% more female identifying responses, medium effect, $p = 0.011$, $V = 0.267$) with both differences being significant. These results suggest that male-identifying teaching associates were more focused on the development of the students' practical skills whereas female-identifying teaching associates were more focused on the students' application of theory.

Whether the teaching associates had worked in industry for more than one year had no significant effect upon the aims raised by respondents. However, teaching associates with more than one year of teaching experience were 19% more likely to raise the enhancing theoretical understanding (TU) aim (small-medium effect, $p = 0.041$, $V = 0.218$) and 23% less likely to raise the preparation for the workforce (WF) aim (medium effect, $p = 0.008$, $V = -0.285$). It would appear that teaching experience resulted in a shift in focus from workforce preparation towards enhancing theoretical understanding. Overall, demographic effects were more pronounced in the responses of the teaching associates compared to the students but, for the purpose of subsequent analysis of this study, they will be treated as a single group apart from their institution. No literature examples of these differences could be found by the authors at this time.

The responses of teaching associates from Monash University and the University of Warwick indicate that five of the themes raised by them were also raised by the students (developing practical skills, application of theory, enhancing theoretical understanding, gaining general laboratory experience, and preparation for the workforce – PS, AP, TU, EX and WF). There are no significant differences between the teaching associate cohorts according to the Pearson's Chi squared test. The high levels of the PS aim ($\sim$60%) indicates that many teaching associates at both universities mainly see teaching laboratories as a chance for students to develop practical skills.

The next aims (application of theory, gaining general laboratory experience, enhancing theoretical understanding and preparation for the workforce – AP, EX, TU and WF) vary slightly in prominence, ranging from 15% (University of Warwick, EX aim) to 41% (Monash University, AP aim). This indicates that these aims, whilst still relevant, are a secondary focus for teaching associates. Following these five aims, there are stark differences between the two cohorts. The teaching associates at Monash University raised a new aim, SR (Safety and Responsibility). It is reassuring to see this as the teaching associates are usually directly responsible for the safety of the students. However, its placement as the least prominent aim could be considered a matter for concern. The other major difference was that the teaching associates at the University of Warwick raised the developing transferable skills (TS) aim less than 10% of the time. This is particularly interesting, as students at the University of Warwick were very likely to raise this aim in their final year which indicates a major inconsistency between student and teaching staff expectations.

The teaching associates at Monash University held views generally consistent with students in the final years of their degrees. This could be due to the teaching associates' potential position as role models shaping students to eventually have viewpoints that matched their own. If this were true, one would expect to see the students at the University of Warwick responding in a manner more consistent with their teaching associates as well, which is not the case (particularly in the developing transferable skills, TS, aim). However, in Australia the role of the teaching associate is well established and teaching associates take more ownership and responsibility for the laboratory teaching environment that their counterparts in the UK. Furthermore, teaching associates in the UK are less varied than those in Australia, who are from more varied backgrounds and generally have more teaching experience. It is possible that this leads to stronger role modelling in Australia and, hence, more similarities between teaching associates' views and those of their students.

In general, the data from the teaching associates at both institutions highlighted a range of similarities and differences between student and teaching staff perceptions of the aims of teaching laboratories. Importantly, the results were different for the two universities, implying that one cannot assume the mind-set of the teaching associates; it must be investigated at each institution. As teaching associates have a potentially high impact on a learning experience, the influence of the beliefs of the teaching associates should not be underestimated. This influence could be addressed through enhanced training of teaching staff (Dobson *et al.*, 2012) or better communication with students, depending on which group held aims more consistent with those desired by a given institution.

Academic staff

Whilst teaching associates arguably spend the most face-to-face time with students in the laboratories, it is generally the academic staff that direct and design the overall teaching laboratory activities. Through controlling which activities will be undertaken, academics (either intentionally or unintentionally) can determine which laboratory aims will be focused on. Hence, gathering the views of academics on teaching laboratories is important and was achieved through interviews rather than a written response to a survey. Two main questions were asked, 'What do you think the aims of doing a practical chemistry course are?' and 'With those aims in mind, do teaching laboratories at your institution succeed at meeting those aims?'

Through the inherent nature of an interview, much longer responses were obtained from the academics than either the teaching associates or the students. Assignment of themes was simplified as they typically justified their statements, shown in the following in depth example (as well as the assigned codes):

'I think that, er, that we want students to be learning a set of techniques that they're going to be using… It is also about learning concepts and encountering concepts in a practical setting. It's about, um, learning to work safely… It's also about learning to work cooperatively, certainly in an environment that is a bit closer to, um, a working environment… I think we need to think about the broader aspects of learning and exploring within a safe environment and working together are increasingly important.' (PS, AP, SR, WF and TS)

Of the 34 academic members of staff, only their professional titles are shown in Table 7 as no demographic evidence was collected during the interviews. Table 8 shows the number of times a given theme was raised.

Results of the analysis of the academics' responses are shown in Fig. 6. Again, only themes raised by more than 10% of the cohort are shown and, due to a limited number of responses from each individual university, the results responses from all three institutions are combined. The graph shows yet another variation from the student responses, with only four major aims being raised. These four aims do however coincide with those raised by both the students and the teaching associates. The most consistently raised one was PS, the development of practical skills. Other skills, such as transferable skills, appeared but to a lesser extent, and more in line with enhancing theoretical knowledge and applying that theory to real examples. The notable lack of the preparation for the workforce (WF) or gaining general laboratory experience (EX) aims indicates that many academics do not see the teaching laboratory as environment in which students could be more prepared for the workforce or as a chance to gain general laboratory experience (although this may be simply due to the subtle difference between this and simply developing practical skills).

Although there are too few responses from any institution to allow for generalisation, additional themes, such as understanding of scientific methodology or independent inquiry (SM) were raised by some individuals. This is exemplified in the following quote and shows that some academics do hold a broader view of the aims of the teaching laboratory:

Table 7 The role titles of the academic members of staff who participated in the interviews at all three universities

Title	Percentage
Lecturer	13
Senior lecturer	16
Associate professor/reader	39
Professor	32

Table 8 The number of responses (converted to percentages) of academics raising a given theme

Codes	*N*	%	Codes	*N*	%
TU	10	30	TS	11	33
AP	17	52	EG	2	6
CX	0	0	SR	1	3
WF	2	6	SM	2	6
EX	2	6	LV	0	0
PS	22	67	OT	1	3

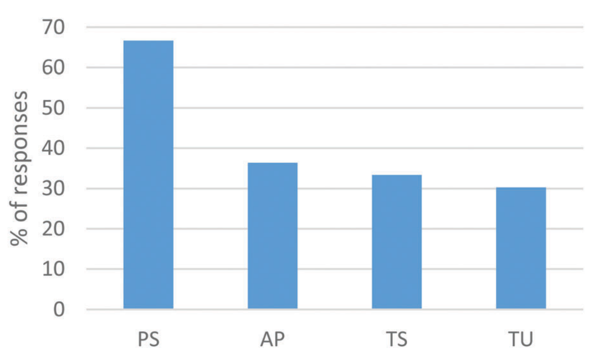


Fig. 6 The percentage of academics raising one of the four main aims.

'. . . I guess the main purpose is to get students to start thinking about independent inquiry. . . they get to try out following their own idea and that's when it becomes quite independent, the nature of it. I think it's quite nice. It's a nice transition from a first or second year lab to like a research project.'

Of the academics interviewed, most held the view that whilst many laboratory activities were well implemented, some activities did not meet the aims that they themselves raised, and typically cited they were too traditional or expository. This is exemplified in the following quote:

'I still think there's an element of spoon-feeding, especially in year one. I do not know year two as well but I still think we're not allowing them to have enough fun.'

This situation is further complicated by the fact that academic staff are often either in disagreement about laboratory aims or are given insufficient guidance, as exemplified by the response:

'I do not think we have any clear guidance or leadership on what our labs should aim for. I do not think we actually have a clear strategy on that. I think there's a lot of disagreement about what they should be used for, um, among the staff.'

These results are in strong agreement with the faculty-focused work conducted in the US (Bruck *et al.*, 2010; Bretz *et al.*, 2013; Bruck and Towns, 2013; Bretz *et al.*, 2016). These studies highlighted a general consensus on the need to develop transferrable and practical skills whilst also imparting theoretical knowledge. Furthermore, the underlying development of scientific methodology or experimental design was generally lacking from both the participants in this and those earlier studies. Some US academics did raise the aim of preparing students for research but, as the number who raised this is unknown, it is difficult to determine if this a greater theme than that noted here.

Not only is it important for academic staff to attempt to come to an agreement about the aims of teaching laboratories, it is also important that they ensure that these aims are fully conveyed to the students and teaching associates. Either simply adding these aims to laboratory manuals or online resources is unlikely to be effective, and will require constant discussion and reinforcement with the students and teaching associates. If it is desired that students fully engage with teaching laboratories in a manner directed by teaching staff, then the value of these critical conversations cannot be underestimated. Without them, the situation represented in this article with academics, teaching associates and students all holding relatively narrow views of the aims of teaching laboratories (likely born from expository experiences) is likely to continue. It is also important for academics themselves to have conversations with one another in order to expand their viewpoints of teaching laboratories beyond the relatively simplistic aims raised in this article.

Limitations

The responses from the students and teaching staff over the three universities generated a large amount of very rich data which directly related to the central aim of this work; to investigate what students and teaching staff perceived the aims of teaching laboratories to be. However, as in all studies, there are limitations to this work that need to be addressed when discussing analysis of the results. These include, but are not limited to: the delivery of the surveys and interviews, number of responses, the method of interpretation and demographics (*e.g.* gender, enrolment or age).

The number of responses at the University of Warwick appear low due to a lower number of enrolled students. The number of responses at UNSW are also lower than Monash University due to the inherent time pressure of their teaching laboratories, resulting in fewer students having the time to complete the survey. Neither effect was considered to negate the results obtained, as there was, in all cases, a response rate greater than approximately 30% and even as high as 80%. The number of teaching associates responding at the University of Warwick also appeared low, but again represented a significant percentage of the entire cohort (approximately 43%). Hence, the number of responses was not considered a major issue in this case.

The method of interpretation would likely be a major source of error, with some theme assignments shifting by up to 10% upon consultation with other chemical education researchers. However, it is believed that the iterative nature of the theme generation negated these issue to significant degree, as reflected in the high level of inter-rater reliability.

Demographic effects were considered throughout the study and could potentially effect the results to varying degrees. Having being measured, it should be noted that the changes discussed throughout this article were generally unlikely to be the cause of the overall changes noted.

Conclusions

The perceptions of 1917 undergraduate students, 118 teaching associates and 34 academic members of staff were transcribed verbatim from surveys and interviews focusing on the open question – 'What do you think the aims of doing a practical chemistry course are?'. These responses were sourced from two Australian universities (Monash University and UNSW) as well as one UK university (the University of Warwick). Inductive analysis resulted in 11 themes being found, with only 4–6 being raised by more than 10% of any of the respective subgroups. These aims were quite narrow and primarily focused on those more in line with expository experiences, such as developing practical skills, applying theory or enhancing theoretical understanding. Other aims, such as the development of transferable skills, preparation for the workforce or gaining general laboratory experience were raised to a lesser degree, with the last two aims not raised by more than 10% of the academic members of staff. This study showed that whilst differences did exist between the perceptions of teaching staff and students, all three groups would likely benefit from either a greater number of conversations around teaching laboratory aims, or simply a larger variety of teaching experiences. Due to the large numbers of respondents alongside the use of three different international institutions, it is believed that this result is applicable to many modern universities around the world.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors would like to acknowledge and thank, first and foremost, the students, teaching associates and academics who participated in this work, whose honest feedback will result in a far stronger learning experience. The authors would also like to acknowledge the technical staff at Monash University, the University of New South Wales and the University of Warwick, without whom, this work would not have been possible. The authors acknowledge Monash University for funding, and hosting, the Transforming Laboratory Learning program. Further funding is also gratefully acknowledged from the Monash Warwick Alliance Seed fund. Ethics approval was obtained from the Monash University Human Ethics Research Committee, application number 2016000584.

References

Bennett D., Richardson S. and MacKinnon P., (2015), Enacting strategies for graduate employability: how universities can best support students to develop generic skills, Sydney: Australian Government Office for Learning and Teaching.

Boden R. and Nedeva M., (2010), Employing discourse: universities and graduate 'employability', *J. Educ. Policy*, **25**(1), 37–54, DOI: 10.1080/02680930903349489.

Boud D. J., Dunn J., Kennedy T. and Thorley R., (1980), The Aims of Science Laboratory Courses: a Survey of Students, Graduates and Practising Scientists, *Eur. J. Sci. Educ.*, **2**(4), 415–428, DOI: 10.1080/0140528800020408.

Bretz S. L., Fay M., Bruck L. B. and Towns M. H., (2013), What Faculty Interviews Reveal about Meaningful Learning in the Undergraduate Chemistry Laboratory, *J. Chem. Educ.*, **90**(3), 281–288, DOI: 10.1021/ed300384r.

Bretz S. L., Galloway K. R., Orzel J. and Gross E., (2016), Faculty Goals, Inquiry, and Meaningful Learning in the Undergraduate Chemistry Laboratory, *Technology and Assessment Strategies for Improving Student Learning in Chemistry*, American Chemical Society, vol. 1235, pp. 101–115.

Bruck A. D. and Towns M., (2013), Development, Implementation, and Analysis of a National Survey of Faculty Goals for Undergraduate Chemistry Laboratory, *J. Chem. Educ.*, **90**(6), 685–693, DOI: 10.1021/ed300371n.

Bruck L. B., Towns M. and Bretz S. L., (2010), Faculty Perspectives of Undergraduate Chemistry Laboratory: Goals and Obstacles to Success, *J. Chem. Educ.*, **87**(12), 1416–1424, DOI: 10.1021/ed900002d.

Dearing R. and Education G. B. N. C. o. I. i. H., (1997), *Higher education in the learning society [Dearing report]*, Leeds: National Committee of Inquiry into Higher Education.

DeKorver B. K. and Towns M. H., (2015), General Chemistry Students' Goals for Chemistry Laboratory Coursework, *J. Chem. Educ.*, **92**(12), 2031–2037, DOI: 10.1021/acs.jchemed.5b00463.

DeMeo S., (2001), Teaching Chemical Technique. A Review of the Literature, *J. Chem. Educ.*, **78**(3), 373, DOI: 10.1021/ed078p373.

Dobson I. R., O'Toole P., Australian Council of Deans of Science (ACDS), Educational Policy Institute, Monash University and Centre for Population and Urban Research, (2012), *Demonstrator Development: Preparing for the Learning Lab*, Australian Council of Deans of Science.

Elliott M. J., Stewart K. K. and Lagowski J. J., (2008), The Role of the Laboratory in Chemistry Instruction, *J. Chem. Educ.*, **85**(1), 145, DOI: 10.1021/ed085p145.

Faraday M., (1830), *Chemical manipulation: being instructions to students in chemistry, on the methods of performing experiments of demonstration or of research, with accuracy and success*, Murray.

Fraser B. J., (1981), Learning Environment in Curriculum Evaluation: A Review, *Evaluation in Education: An International Review Series*, **5**(1), 3–93, DOI: 10.1016/0191-765X(81)90014-8.

Galloway K. R. and Bretz S. L. (2015a), Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory, *J. Chem. Educ.*, **92**(7), 1149–1158, DOI: 10.1021/ed500881y.

Galloway K. R. and Bretz S. L. (2015b), Measuring Meaningful Learning in the Undergraduate Chemistry Laboratory: A National, Cross-Sectional Study, *J. Chem. Educ.*, **92**(12), 2006–2018, DOI: 10.1021/acs.jchemed.5b00538.

Galloway K. R., Malakpa Z. and Bretz S. L., (2016), Investigating Affective Experiences in the Undergraduate Chemistry Laboratory: Students' Perceptions of Control and Responsibility, *J. Chem. Educ.*, **93**(2), 227–238, DOI: 10.1021/acs.jchemed.5b00737.

Good H. G., (1936), On the early history of Liebig's laboratory, *J. Chem. Educ.*, **13**(12), 557, DOI: 10.1021/ed013p557.

ICSU-CTS, (1979), *Learning strategies in university science*, Cardiff: University College Cardiff Press.

Johnstone A. H. and Al-Shuaili A., (2001), Learning in the laboratory; some thoughts from the literature, *Univ. Chem. Educ.*, **5**(1), 41–50.

Kirschner P. A. and Meester M. A. M., (1988), The Laboratory in Higher Science Education: Problems, Premises and Objectives, *High. Educ.*, **17**(1), 81–98, DOI: 10.1007/BF00130901.

Leal Filho W. and Pace P., (2016), *Teaching Education for Sustainable Development at University Level*, Springer.

Linn M. C., (1997), The Role of the Laboratory in Science Learning, *Elem. School J.*, **97**(4), 401–417, DOI: 10.1086/461873.

Lynch P. P. and Ndyetabura V. L., (1983), Practical work in schools: an examination of teachers' stated aims and the influence of practical work according to students, *J. Res. Sci. Teach.*, **20**(7), 663–671, DOI: 10.1002/tea.3660200707.

Reid N. and Shah I., (2007), The role of laboratory work in university chemistry, *Chem. Educ. Res. Pract.*, **8**(2), 172–185, DOI: 10.1039/B5RP90026C.

Rentoul A. J. and Fraser B. J., (1979), Conceptualization of Enquiry-Based or Open Classroom Learning Environments, *J. Curriculum Stud.*, **11**(3), 233–245, DOI: 10.1080/0022027790110306.

Russell C. B. and Weaver G., (2008), Student perceptions of the purpose and function of the laboratory in science: a grounded theory study, *Int. J. Scholarsh. Teach. Learn.*, **2**(2), 9, DOI: 10.20429/ijsotl.2008.020209.

Shah I., (2007), *Making University Laboratory Work in Chemistry More Effective*, MPhil thesis, University of Glasgow, Scotland.

Shymansky J. A. and Penick J. E., (1979), Use of systematic observations to improve college science laboratory instruction, *Sci. Educ.*, **63**(2), 195–203, DOI: 10.1002/sce.3730630207.

Taylor J., (1986), The employability of graduates: differences between universities, *Stud. High. Educ.*, **11**(1), 17–27.

Wilkinson J. and Ward M., (1997), A comparative study of students' and their teacher's perceptions of laboratory work in secondary schools, *Res. Sci. Educ.*, **27**(4), 599–610, DOI: 10.1007/bf02461483.

Chemistry Education Research and Practice

PAPER

View Article Online
View Journal | View Issue

Cite this: *Chem. Educ. Res. Pract.*, 2019, **20**, 187

Investigating student and staff perceptions of students' experiences in teaching laboratories through the lens of meaningful learning

Stephen R. George-Williams, * Dimitri Karis, Angela L. Ziebell, Russell R. A. Kitson, Paolo Coppo, Siegbert Schmid, Christopher D. Thompson and Tina L. Overton

How students behave and learn in the teaching laboratory is a topic of great interest in chemical education, partly in order to justify the great expense of teaching laboratories. Much effort has been put into investigating how students think, feel and physically act in these unique learning environments. One such attempt was made through the generation and utilisation of the Meaningful Learning in the Laboratory Instrument (MLLI). This 30 question survey utilised Novak's theory of Meaningful Learning to investigate the affective, cognitive and psychomotor domains of the student learning experience. To date, this survey has been used to great effect to measure how students' perception of their own feelings and actions will change over the course of a semester. This study reports the use of a modified MLLI survey to probe how the expectations of students change over their undergraduate degree. To increase the generalisability of the outcomes of the study data was gathered from four universities from Australia (Monash University, the University of New South Wales and the University of Sydney) and the UK (the University of Warwick). Students were found to start their university careers with very positive expectations of their teaching laboratory experiences. Their outlook became somewhat more negative each year that they were enrolled in the program. A further modified MLLI survey was presented to teaching associates and academic staff. Teaching staff were shown to have far more negative expectations of the students' feelings and actions, with academic staff more likely to believe that students do not undertake many items of positive meaningful learning. Overall, this study highlights the large gap between the expectations of teaching staff and students which, if left unaddressed, will likely continue to cause great frustration for both teaching staff and students.

Received 21st July 2018,
Accepted 21st September 2018

DOI: 10.1039/c8rp00188j

rsc.li/cerp

Introduction

In almost every institution that teaches chemistry throughout the world, one would expect to encounter a teaching laboratory that complements the lectures and tutorials delivered to the students. Whilst it is commonly believed that these teaching laboratories aid student learning, scant meaningful evidence has been obtained to support such a claim (Hofstein and Lunetta, 1982; Hofstein and Lunetta, 2004). Indeed, there have been arguments over the need for teaching laboratories (Hawkes, 2004; Morton, 2005; Sacks, 2005; Stephens, 2005) with concerns raised that not all students continue in chemistry and therefore do not require the practical skills developed in laboratories. Furthermore, the laboratory teaching experiences themselves are often criticised for being too expository or recipe-based (Letton, 1987; Hodson, 1990), *i.e.* they rely heavily on laboratory manuals and cause students to simply follow a procedure (Domin, 2007). Overall, there is a need to investigate the value of the learning undertaken in teaching laboratories.

Galloway and Bretz (2015a) sought to meet this need through the generation of the Meaningful Learning in the Laboratory Instrument (MLLI). This 31 item survey consisted of a range of questions that were generated through the lens of Joseph Novak's Theory of Meaningful Learning and Human Constructivism (Novak, 1998). This theory focuses on the concept that true learning requires the overlap of the affective, cognitive and psychomotor domains of the students' thoughts and actions. Whilst many surveys exist in the literature with a focus on teaching laboratories, these tend to focus on just the cognitive domain (Grove and Bretz, 2007) or just the affective domain (Bauer, 2005, 2008; Xu and Lewis, 2011). The MLLI survey was the first survey to 'focus solely on learning in the laboratory and to expressly operationalize a theory of learning' (Galloway and Bretz, 2015a).

School of Chemistry, Monash University, Victoria, 3800, Australia.
E-mail: [redacted]

The original use of this survey focused on the changes in the students' expectations in relation to teaching laboratories after one semester, in either first year general chemistry or first year organic chemistry. It was generally found that the expectations of the students were not being met by the experiments that they undertook (*e.g.* students felt that they were not thinking about what the molecules are doing). A second study was undertaken at a national level (Galloway and Bretz, 2015b), including 15 different institutions and the responses of 3853 students. With this data, the researchers were able to support the supposition that the mismatch between student expectations and experiences was a widely observed issue.

The aforementioned surveys primarily focused on the perspectives of students. The responses of teaching staff have also been investigated in the literature either through interviews (Bruck *et al.*, 2010; Bretz *et al.*, 2013) or the Faculty Goals Survey (Bruck and Towns, 2013; Bretz *et al.*, 2016). It is worth noting, however, that these investigations tended to have a wide focus on the overall aims or goals of teaching laboratories rather than on the specific actions or feelings of students as raised by the MLLI survey. These studies indicated that academic staff tended to focus more on cognitive or psychomotor goals compared to affective ones.

Whilst these previous studies highlight the large amount of work already undertaken in this field, the results are difficult to compare to one another due to the different means of measurement and underlying focus. Additionally, the responses of students have tended to be sourced from first year cohorts and the viewpoint of teaching associates or laboratory demonstrators is currently missing from the literature.

To address these issues, this study sought to utilise the MLLI survey to investigate the perceptions of both students and teaching staff of how students will act and feel during teaching laboratories. The study also investigated the responses of students in upper year levels in order to measure the longitudinal impact of multiple chemistry courses. Finally, data was collected at four different institutions in two different education systems in order to increase the generalisability of any conclusions drawn. Data was also collected from teaching staff in order to compare their views of student experiences with those of the students themselves.

Method

The aim of this study was to compare students' and teaching staffs' perceptions of the cognitive, psychomotor and affective expectations of students during teaching laboratories. This was investigated using a quantitative analysis of responses to either a paper-based or online survey.

Data collection

Undergraduate students and teaching associates (sometimes referred to as laboratory demonstrators) were asked to answer the modified MLLI survey (Galloway and Bretz, 2015a) in paper format. The scale was modified from the original electronic slider (0–100%) into a five point Likert scale (strongly disagree, disagree, neutral, agree and strongly agree) at Monash University, the University of New South Wales (UNSW), and the University of Warwick and a 1–10 scale (*i.e.* on a scale of one 1–10 how much do agree with the statement) at the University of Sydney. The modified survey also included some general demographic questions about age and gender, course choice and domestic or international enrolment for students, or the amount of teaching or industry experience for teaching associates. Academic staff were asked to either complete another modified survey in paper format or through an online Google form. Teaching associates and academics were approached *via* email. All participants were informed that the survey was voluntary and would not affect either their academic standing or employment.

At Monash University, all students enrolled in chemistry courses, at any year level, were given the opportunity to complete the survey. In total, two first year courses, three second year courses and four third year courses were surveyed. There were a total of around 1700 students enrolled in these courses in mid-2016. Teaching associates were asked to complete the survey at the end of a compulsory training session in early 2017. There were approximately 120 teaching associates present who taught across all year levels.

At UNSW the students were asked to complete the survey at the end of their first teaching laboratory in semester 1, 2017. Unlike Monash University, access was not readily available to all chemistry courses so only three second year and two third year courses were surveyed. The number of enrolled students was similar to Monash University but only around 1350 received the survey.

The responses for the University of Sydney were collected during a typical laboratory session early in semester 1, 2017. Students were asked to complete the survey before the commencement of their experiments for the day. The number of enrolled students was similar to Monash University.

Responses from students at the University of Warwick were obtained during three separate events where a lunch was provided, scheduled on days when students were undertaking laboratory exercises early in their academic year (late 2016). Third year students were not surveyed until May 2017 due to scheduling commitments. There were approximately 490 students enrolled in the first three year levels at the University of Warwick at the time of data collection.

Monash University, the University of Sydney and the University of New South Wales represent three of the eight Australian universities known as the 'Group of Eight'. These universities routinely perform very well on international teaching and research ranking platforms. The University of Warwick is a member of the 'Russell Group' in the United Kingdom. It is a highly prestigious university that also performs highly on international teaching and research platforms.

Research theoretical framework

The primary theoretical framework of this work is Constructivism which postulates that learning is an active process that builds upon the prior experiences of the learner (Leal Filho and Pace, 2016).

Hence, in the case of a respondents' perceptions of how students act and feel during teaching laboratory experiments, it is postulated that their responses will be a direct result of their experiences prior to answering the closed questions.

Data analysis

The surveys completed by 3202 undergraduate students, 143 teaching associates and 102 academic staff were transcribed into Excel after recoding (*e.g.* Strongly Disagree = 1, Disagree = 2, Neutral = 3, Agree = 4 and Strongly Agree = 5). Data from the University of Sydney was originally on a 1–10 scale, which was then recoded (1–2 became 1, 3–4 became 2, 5–6 became 3, 7–8 became 4 and 9–10 became 5) in order to allow for common methods of analysis.

To ensure the questions within the survey held a reasonable amount of internal consistency, a Cronbach's Alpha was calculated by SPSS for all student responses and found to be 0.756. As this value was over the literature threshold of 0.7 (Nunnally and Bernstein, 1994), and was likely underestimated as per the use of Cronbach's alpha on ordinal data (Gadermann *et al.*, 2012), the internal consistency of the survey was considered reasonable.

In order to determine the significance of any differences between teaching staff and students, universities, year levels or demographics, the coded data was analysed as frequencies (*e.g.* the number of respondents selecting agree) with SPSS. Using overall data an omnibus test was performed in the form of an *F*-test with effect sizes measured through a calculation of eta-squared (η^2). For example, the responses from 16–18 year olds to all questions was directly compared to how all students (of any age level) responded. Further comparisons of how the 16–18 year olds answered individual questions was only performed if the *F*-test showed no significant difference or exhibited an effect size <0.04. Hence, this test would show that any differences noted later on are more likely due to how the cohort responded to a specific question rather than how they responded to any question. Comparison of group responses to individual questions was achieved through using a Pearson's chi-squared test to check that differences held to Holm–Bonferroni corrected confidence interval (*i.e.* $p \leq 0.05$ became $p \leq 0.002$ for the first nine questions compared). Cramér's $V(\phi_C)$ was also calculated in order to measure the effect sizes of any differences (df = 4; small, 0.05–0.15, medium, 0.016–0.25, or large, >0.25) (Sheskin, 2003).

Limitations

The first limitation of this study is the change in scales utilised throughout this work as compared to the original electronic slider (0–100%) from the MLLI survey. This change complicates direct comparison between the results generated in this study with those previously reported. This change was made to allow for a paper format rather than an electronic version in order to increase the number of responses received from students and teaching staff. As such, the transcribed data was treated differently to the original work *i.e.* a question by question analysis with a Pearson's chi-squared test. It is worth noting that a factor analysis did not show factors that aligned with original ones raised by Galloway and Bretz (2015a) (affective, cognitive and affective/cognitive). This is possibly due to the students at these universities failing to connect which items are connected to either the cognitive, psychomotor or affective domains *i.e.* the students potentially exhibit poor meaningful connections to the questions being raised. Therefore, direct comparison between the factors raised in the original results of the MLLI survey and this study will not be achievable.

The second limitation of this work is the variation in the percentage of respondents from any given group of students. These ranged from 22% to 80% of the various cohorts surveyed. The approach of Barlett *et al.* (2001) can be used to evaluate whether each data set is representative of the respective cohort. However, the application of these minimal sample sizes is complicated, as this study uses ordinal data rather than the continuous or categorical data considered in the literature. If one were to utilise the lower acceptable sample sizes for the continuous data with an alpha of 0.05 (*i.e.* a 5% error), there are sufficient responses in eight of the twelve subgroups (*e.g.* first year students at Monash University) to statistically represent those subgroups. Of the four that fail to meet these minimum requirements (UNSW, third year students; the University of Sydney, second and third year students; University of Warwick, second year students), they tend to fall short by only 10–15 responses. As such, whilst these data sets may not be fully representative of their respective cohorts, their use as comparison data likely mitigates this issue.

The number of responses from teaching associates at Monash University (111) are considered statistically representative of that group. However, the number of responses at the University of Warwick is significantly less than required for statistical representation (32). Therefore, the data from the teaching associates at the University of Warwick is unlikely to be truly representative. The responses from academic staff were collected at a range of institutions in Australia and the United Kingdom and are, therefore, non-representative of any given institution or country.

The third limitation of this work is the change in scale utilised in the data collected from the University of Sydney (1–10 rather than strongly disagree to strongly agree). This was simply due to multiple research groups gathering data individually prior to later collaboration. This could potentially result in erroneous conclusions being drawn from that data. However, as the data were being used as a comparison data set, this effect was believed to have been minimal.

The fourth limitation is the statistical test used to compare the data sets. The use of an *F*-test to measure overall variation (an omnibus test) is typically used to compare the mean values of multiple data sets, which is usually considered poor practice for ordinal data. However, measuring the same data through a Pearson's Chi-squared test (which required strict control of sample sizes), generally revealed either no significant differences or differences with little to no practical relevance (*i.e.* $\phi_c < 0.1$). Furthermore, the Pearson's Chi-squared test utilised is actually for categorical data (Pearson, 1900) whereas the data produced by this survey are ordinal. In some cases, another common statistical test, the Mann Whitney *U*, can be used for ordinal data (Nachar, 2008). However, this test cannot be used to compare data sets that do not have the same overall shape (*e.g.* a bimodal response compared to a unimodal response). To remove this error, the Pearson's chi-squared test was used. It is possible that the choice of statistical test affected the results, but a comparison of outcomes of applying both tests on a single data set (male-identifying first year students *vs.* female-identifying first year students at Monash University) showed minimal differences between them.

Finally, as this survey was sourced from the literature with no changes made to the parent statements (other than changing the focus for teaching staff), it was believed the original validity and reliability measures undertaken by the original authors of the survey would suffice in this case. As such, no attempts were made to measure or ensure the validity or reliability of this survey.

Results and discussion

The cohorts of students who responded to the survey and their demographic data are shown in Table 1.

Table 1 The demographic breakdown of the student cohorts by university, year level, gender identity, age and enrolment

		1st year students	2nd year students	3rd year students
Monash University (*N*)		965	210	159
Gender (%)	M	46	51	60
	F	52	47	37
	Other	2	2	2
Age (%)	16–18	46	85	1
	19–21	49	6	71
	22+	5	9	28
Enrolment (%)	Domestic	92	92	90
	International	8	8	10
UNSW (*N*)		419	238	55
Gender (%)	M	52	43	51
	F	48	57	45
	Other	0	0	4
Age (%)	16–18	65	12	0
	19–21	29	79	80
	22+	6	9	20
Enrolment (%)	Domestic	88	75	85
	International	12	25	15
USyd (*N*)		728	68	77
Gender (%)	M	40	59	47
	F	58	40	53
	Other	2	1	0
Age (%)	16–18	79	51	6
	19–21	16	41	81
	22+	5	7	13
University of Warwick (*N*)		148	58	77
Gender (%)	M	45	49	54
	F	55	51	46
	Other	0	0	0
Age (%)	16–18	51	0	0
	19–21	48	100	70
	22+	1	0	30
Enrolment (%)	Domestic	90	95	90
	International	10	5	10

Generally, there were approximately equal numbers of male-identifying students and female-identifying students. Of the students enrolled, 10% were noted to be international students. The majority of the students were aged 16–18 in their first year of study, with older students enrolled in higher year levels as expected. As this data matched the demographic enrolment of the individual cohorts, the data appears to be representative of the student bodies, at least in terms of gender, age and enrolment.

Demographic effects

The survey was first used to investigate demographic effects on the largest data set available – the Monash University first-year cohort. When enrolment and gender were considered, only one question was noted to have significantly different responses. Furthermore, Cramér's *V* indicated only a small effect size for the individual question *i.e.* whilst there was a measurable significance difference, the actual difference was quite small. Investigating the effect of age resulted in no questions showing significant differences. Hence, gender, age and enrolment were not considered significant sources of variation in this study.

Overall, demographic categories did not appear to have particularly notable effects on the responses. Hence, demographics will not be raised again and are considered an unlikely source of any changes noted.

First year students

Comparisons of the responses of the first year students from three of the institutions were made. Each institution's results were compared with those from Monash University, as this was the largest data set. The data from the University of Sydney was excluded from this analysis as the change in scale (a 1–10 scale *versus* strongly disagree–strongly agree) meant that any discrepancies in the data could not be deconvoluted from this.

Before a direct comparison could be made, the large difference in the sample size between Monash University first year students (965) and UNSW first year students (419) needed to be taken into account. Hence, only every second result from Monash University was utilised in the analysis (463 in total). This method of analysis has been shown to be valid and accounts for the sensitivity of the Pearson's chi-squared to large variations in sample sizes (Barlett *et al.*, 2001). Only two questions showed a significant difference between student responses ($p \leq 0.0005$) and both exhibited only a medium effect size ($\phi_c = 0.182$ and 0.159). Overall, these results indicate a minimal difference between the responses from first year Monash University and UNSW students. As these universities are both high ranking international universities within the same education system, this lack of difference is not surprising.

To compare the results of Monash University and the University of Warwick, the sample sizes (965 and 148 respectively) were again accounted for. Therefore, every fifth response from Monash University was used for the analysis (193 in total). Again, only two questions showed a significant difference

Table 2 Questions that showed a significant difference when comparing the responses of Monash University first year students (MUFYS) with first year students enrolled at the University of Warwick (UWFYS), as measured by the Pearson's chi-squared test. Due to the 95% confidence interval chosen, only changes above 5% are noted

Question showing significant difference	>5% decrease in response (MUFYS compared to UWFYS)	>5% increase in response (MUFYS compared to UWFYS)	p value	ϕ_C
To be confused about how the instruments work.	Neutral	Agree	<0.0005	0.275
To be excited to do chemistry.	Strongly agree	Neutral	<0.0005	0.253

($p < 0.0005$), with both exhibiting a large effect size ($\phi_c = 0.253$ and 0.275). These results are shown in Table 2.

These results appear to show that the first year cohort at the University of Warwick had very similar expectations to those at Monash University or UNSW. The only notable differences were that students at the University of Warwick may be more likely to expect excitement, or to be confused by instrumentation. This is potentially the result of the UK university system where students choose a single discipline (*e.g.* chemistry) compared to the Australian students who undertake a more general education.

Higher year students

The responses from the first year students were compared to the second and third year students at each of the four institutions. In order to compare the overall effect of each laboratory program, changes in the responses to specific questions that resulted in at least a medium effect size between the first year and the third year data are shown in Table 3. It is worth noting that subsets of the data were again required in order to ensure more comparable sample sizes.

The results from the UNSW students showed no changes when comparing first year to third year students. This could be the result of either a non-impactful laboratory program or a program that meet their expectations, but is worth noting that only a small number of third year students responded to the survey ($n = 58$). Hence, it is possible that this result is simply underestimating the effect at UNSW. At Monash University, the University of Sydney, and the University of Warwick, Table 3 indicates almost no effect of the laboratory programs on the expectations of the students as only two to three questions showed a significant difference at any given university.

It is worth noting that the second-year data indicates a combination of either gradual or erratic changes (*e.g.* inconsistent changes between year levels and even bimodal data). However, most questions changed in overall gradual manner and are therefore more likely an artefact that changes with time (*e.g.* experience or maturation). An example from the University of Sydney is shown in Fig. 1.

Of the questions that do show significant change, it would appear that they represent students either experiencing significant fatigue with the teaching laboratory experience or becoming more realistic in their expectations, having lost their original assumptions over time. There was only one positive change noted at the University of Warwick where students were more likely to expect to 'make decisions about what data to collect'. This may be because a larger portion of the third year teaching laboratories at the University of Warwick are open-ended and project-based (as compared to the Australian universities), likely explaining the students' increasing expectations around making decisions. It is also worth noting that the questions that showed a significant change are not consistent between the universities. Hence, the MLLI survey is able to differentiate between the effects of different laboratory programs on student expectations. Overall, this investigation would appear to show that student expectations are not being significantly changed by their experiences.

However, if any change is to be believed, it would appear that students are generally harbouring more negative expectations as they proceed further in their undergraduate experiences.

Table 3 Questions that showed a significant difference when comparing the responses of first year students (FY) with third year students (TY) as measured by the Pearson's chi-squared test. Due to the 95% confidence interval chosen, only changes above 5% are noted

Question showing significant difference	>5% decrease in response (TY compared to FY)	>5% increase in response (TY compared to FY)	p value	ϕ_C
Monash University				
To make decisions about what data to collect.	Agree or strongly agree	Neutral or disagree	<0.0005	0.278
To use my observations to understand the behaviour of atoms and molecules.	Strongly agree	Neutral, agree or disagree	<0.0005	0.263
To think about what the molecules are doing.	Strongly agree	Neutral or agree	0.002	0.219
University of Sydney				
To be frustrated.	Strongly disagree or disagree	Agree	<0.0005	0.353
To be confident when using equipment.	Strongly agree	Neutral or disagree	0.001	0.319
The University of Warwick				
To be excited to do chemistry.	Strongly agree	Neutral	0.001	0.277
To worry about finishing on time.	Disagree or neutral	Agree or strongly agree	0.002	0.267
To make decisions about what data to collect.	Disagree or neutral	Strongly agree	0.002	0.264

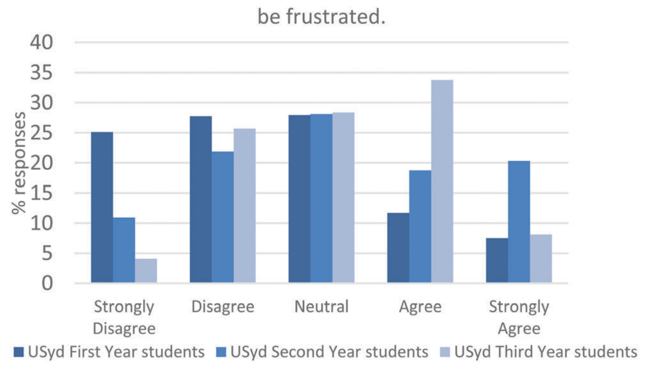


Fig. 1 The percentage of students selecting a given response at the University of Sydney to the statement 'I expect to be frustrated'.

Table 4 The demographic breakdown of the teaching associates who responded to the survey at Monash University and the University of Warwick

Monash University		
N		111
Gender (%)	M	54
	F	45
	Other	1
Age (%)	19–21	9
	22–24	61
	25+	34
Teaching experience (%)	< One year	44
	≥ One year	56
Industry experience (%)	< One year	69
	≥ One year	31
Occupation (%)	Postgraduate student	75
	Other	25
University of Warwick		
N		32
Gender (%)	M	57
	F	40
	Other	3
Age (%)	22–24	50
	25+	50
Teaching experience (%)	< One year	30
	≥ One year	70
Industry experience (%)	< One year	78
	≥ One year	22
Occupation (%)	Postgraduate student	100

Whether this is due to more challenging higher year experiences, poorly designed laboratory programs, student maturation or even student fatigue is difficult to discern at this point. That being said, some items, such as the level of student decision making, is something that can be directly impacted by the design of the laboratory program. Hence, the change in student perception is likely a combination of all these three factors. These results are consistent with the original MLLI results which showed that many students held expectations that were not being met by their current teaching laboratories (Galloway and Bretz, 2015a).

Teaching associates

As teaching associates can have a significant impact on the engagement/learning of the students, their beliefs about the behaviours of students were also considered. Hence, teaching associates were asked to answer a modified form of the survey with the parent statement changed from 'I expect to...' to 'I think the students will...'. Table 4 shows the demographic breakdown of the teaching associates. Note that there were too few responses from the UNSW and the University of Sydney teaching associates to be considered representative, hence these data sets are not shown.

Consideration of demographics by the Pearson's chi-squared test resulted in no significant differences in each demographic group (gender, age or teaching/industry experience). This may be because the demographics of the respondents had no effect on their responses, but it is also possible that there were simply not enough responses from teaching associates to measure these effects.

A comparison was then made between the teaching associates and the students. As third year students have had more laboratory experiences, it was decided to compare teaching associates to these students (as they were likely to be more closely aligned). Consequently, the responses of 111 Monash University teaching associates was compared to all 159 Monash third year student responses.

This comparison is the only example in this overall study where the F-test indicated that teaching associates responded in a significantly different manner to students overall ($p < 0.005$, $\eta^2 = 0.06$). However, it is worth noting that the effect size ($\eta^2 = 0.06$), whilst moderate (*i.e.* $0.04 < \eta^2 < 0.36$), is low and likely has no practical significance.

Of the 30 questions in the survey, 21 showed a significant difference with a large effect size. The questions that yielded these differences are shown in Table 5, which highlights many important changes.

Firstly, the effect sizes are much larger than those previously noted in this study with 16 showing effect sizes above 0.4. This alone indicates that there is a large variation in opinion between the teaching associates and the students. Secondly, many of the questions (13) saw drastic shifts from the students responses of strongly disagree/disagree to the teaching associates responses of agree/strongly agree. These questions were generally items that would provide a negative impact on students' meaningful learning (*e.g.* to be frustrated or to worry feel disorganised). This implied that teaching associates were far more likely to believe that students would experience negative emotions and undertake actions that would hinder their meaningful learning. An example shift is shown in Fig. 2.

Teaching associates also did not tend to agree to many items (eight in total) that were positive for the students' meaningful learning. Questions that focused on student confidence, interpretation of data or even just general excitement were met with more neutral responses. Overall, the results from the Monash University teaching associates indicated that teaching staff held far more negative views of the experiences that they expect students to have during teaching laboratories. Furthermore, as this comparison was made with the more negative third year cohort, this issue would likely be even more prominent for first year classes. It is also worth noting that even though a large number of the teaching associates were relatively inexperienced postgraduate students, these differences were already beginning

Table 5 Questions that showed a significant difference when comparing the responses of third year students with teaching associates at Monash University, as measured by the Pearson's chi-squared test. Due to the 95% confidence interval chosen, only changes above 5% are noted. NB: the final p value shown (0.003) is strictly above the threshold due to the Holm–Bonferroni method used instead of just a Bonferroni method

Question showing significant difference	>5% decrease in response (teaching associates compared to students)	>5% increase in response (teaching associates compared to students)	p value	ϕ_C
Be confused about how the instruments work.	Strongly disagree or disagree	Neutral	<0.0005	0.584
Be nervous about making mistakes.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.568
Be confused about the underlying concepts.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.568
Focus on procedures, not concepts.	Disagree	Agree or strongly agree	<0.0005	0.528
Be confident when using equipment.	Agree or strongly agree	Neutral or disagree	<0.0005	0.524
Feel intimidated.	Strongly disagree or disagree	Neutral, agree or strongly agree	<0.0005	0.520
Feel unsure about the purpose of the procedures.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.516
Be frustrated.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.512
Think about what the molecules are doing.	Strongly agree or agree	Disagree or neutral	<0.0005	0.510
Be nervous when handling chemicals.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.509
Feel disorganized.	Strongly disagree or disagree	Agree	<0.0005	0.500
Worry about finishing on time.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.483
Be confused about what their data means.	Disagree or neutral	Agree or strongly agree	<0.0005	0.470
Worry about getting good data.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.460
Interpret their data beyond only doing calculations.	Agree or strongly agree	Neutral	<0.0005	0.432
Use their observations to understand the behaviour of atoms and molecules.	Agree	Neutral or disagree	<0.0005	0.418
Worry about the quality of their data.	Strongly disagree or disagree	Agree or strongly agree	<0.0005	0.385
Be excited to do chemistry.	Agree or strongly agree	Neutral	<0.0005	0.343
Consider if their data makes sense.	Strongly agree	Disagree or neutral	0.001	0.296
Develop confidence in the laboratory.	Strongly agree	Neutral	0.001	0.285
Learn critical thinking skills.	Strongly agree	Neutral or disagree	0.003	0.270

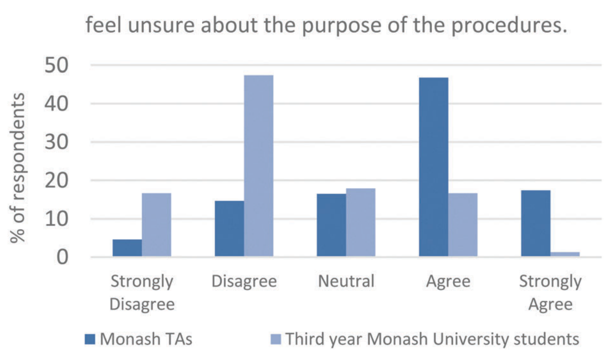


Fig. 2 The percentage of third year students or teaching associates selecting a given response to the statement 'I expect to/I think the students will: feel unsure about the purpose of the procedures'.

to surface in their expectations of students during teaching laboratories.

The uniformity of the teaching associate responses over multiple institutions was also considered. As sufficient responses were only obtained from Monash University and the University of Warwick (118 and 34, respectively) only this comparison was made. No questions were found to be answered in a significantly different manner.

With no differences noted it would appear that the teaching associates held relatively similar views between the two universities, potentially implying that these results may be somewhat generalisable to other universities.

These comparisons highlight the need to ensure that the teaching associates are adequately trained in order to either deal with potential pitfalls for the students or to recognise their own potential negativity. These results also highlight a large mismatch between student and staff expectations which, left unaccounted for, could lead to greater frustration for both students and staff. Lastly, these results indicate that whilst teaching staff can vary greatly between institutions, their overall viewpoints may be somewhat similar with regards to their perceptions of students' experiences during teaching laboratories.

Academic staff

Academic members of staff are often responsible for training teaching associates and regularly interact with students directly during teaching laboratories (although often to a lesser degree than teaching associates). Therefore, their perceptions of what

Table 6 The demographic breakdown of the 102 academic staff who responded to the survey

Gender	Male	Female	Rather not say
	74%	22%	4%

Teaching experience	1–3 years	4–6 years	7+ years
	11%	17%	71%

Industry experience	0 years	1–2 years	3–4 years	5+ years
	77%	14%	0%	7%

Professional title	Professor	Associate Professor	Senior Lecturer	Lecturer	Other
	21%	21%	15%	24%	19%

Table 7 Questions that showed a significant difference when comparing the responses of teaching associates with academic staff, as measured by the Pearson's chi-squared test. Due to the 95% confidence interval chosen, only changes above 5% are noted

Question showing significant difference	>5% decrease in response (academics compared to teaching associates)	>5% increase in response (academics compared to teaching associates)	p value	ϕ_C
Make decisions about what data to collect.	Agree or strongly agree	Disagree	<0.0005	0.377
Consider if their data makes sense.	Agree or strongly agree	Neutral, disagree or strongly disagree	<0.0005	0.304
Focus on procedures, not concepts.	Neutral	Strongly agree	<0.0005	0.298
Learn critical thinking skills.	Strongly agree	Neutral or agree	<0.0005	0.289
Be confused about how the instruments work.	Strongly agree	Neutral	0.001	0.263
Learn chemistry that be useful in their life.	Agree	Neutral or disagree	0.002	0.263
Learn problem solving skills.	Strongly agree	Neutral or agree	0.002	0.248

students would think and do during teaching experiences was also explored. The demographics of the 102 academics who responded to the survey are shown in Table 6.

Generally speaking, most of the academic respondents identified as male, had been teaching for a significant length of time, had limited industrial experience and held a range of professional titles. Due to this, no demographic analysis could be undertaken as the numbers within many of the demographic sub-groups were too small to conduct a meaningful statistical analysis. It is important to note that the academic respondents were employed at a range of institutions across Australia and the UK. Consequently, the responses from the academics were considered as a single group. The responses of the academic staff were directly compared with the teaching associates and seven of the 30 items were responded to in a significantly different manner (p = <0.0005–0.002). Six differences exhibited a large effect size (ϕ_c = 0.263–0.377) whilst one was found to have a medium effect size (ϕ_c = 0.248) (Table 7).

An example of a particularly large effect size where academic staff, students and teaching associates from all universities all answer significantly differently is shown below in Fig. 3. Generally speaking, academic staff were less sure than the teaching associates that students would experience such meaningful learning, such as making decisions, considering if their data made sense or learning critical thinking skills. Furthermore, the academics maintained the teaching associates' belief that the students would encounter experiences negative to their meaningful learning (as noted by a lack of significant differences to those prompts).

Hence, this data implies that the responses of the academic staff contrasted even more with the responses of the students. No direct comparisons between the responses of academic staff and students were made due to issues of varied year levels, sample sizes, utilised scales and institutions. As before, this data implies that there is a very large gap between student expectations and the expectations of the teaching staff.

Whilst it is possible that many of these differences could be due to student naiveté and teaching staff experience (or simply a large disconnect between academic and student viewpoints), this cannot explain every variation. For example, students expect to make decisions whilst academics believe that they would not. This lack of belief in student inquiry is more likely due to a simple lack of inquiry-based experiences which can be easily rectified through an increase in such activities. Hence, these items should be individually probed at any given institution that endeavours to enhance the student experience by better matching the laboratory program with the expectations held by the students where possible.

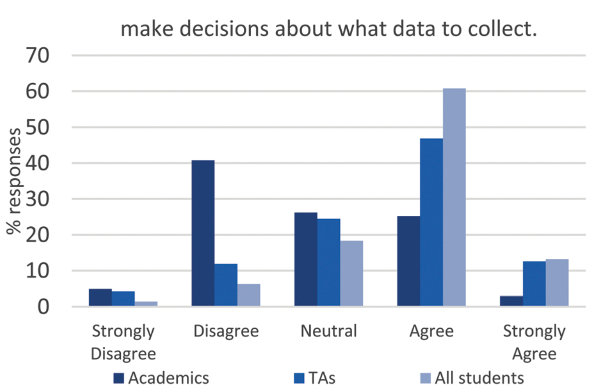


Fig. 3 The percentage of academics, teaching associates and students selecting a given response to the statement 'I expect to/I think the students will: make decisions about what data to collect'.

Conclusions

This work highlights the significantly large gap between the expectations of students and teaching staff with regards to the experiences of students in undergraduate teaching laboratories as seen through the lens of Novak's meaningful learning. Through the use of the MLLI survey, data concerning the expectations of students at three Australian universities (UNSW, The University of Sydney and Monash University) and one UK university (The University of Warwick) have been collected and analysed through the use of Pearson's chi-squared test. This survey was delivered in paper format whilst the students were either completing (or about to undertake) an experiment, during a safety induction or at a lunch provided for them.

In total, 3202 students responded to the survey across all four institutions. In general, students tended to start their university careers with positive expectations of teaching laboratories. This was noted through students agreeing to the statements on the MLLI survey that aided their meaningful learning (*e.g.* I expect to be excited). Students also tended to select either neutral, disagree or strongly disagree to statements that would have a negative impact on their meaningful learning (*e.g.* I expect to be frustrated). As students progressed through the laboratory program at any of the four universities, their responses generally did not change. However, a few minor changes appeared to indicate a small shift to a more negative outlook (albeit with a large number of the changes only showing a medium effect size). Each laboratory program elicited slightly different changes indicating that the students' expectations were likely shifting because of the experiences provided by the institutions rather than maturity or some other factor that changed with time.

The responses of 143 teaching associates from Monash University and the University of Warwick were analysed. Very few differences were found between the responses of teaching associates at either institution, implying a degree of generalisability to these results. When compared to the students, a large number of questions were found to be answered in a significantly different manner. Overall, teaching associates were far more likely to think that students would undertake actions that would lead to their negative learning (*e.g.* I think that the students will be confused about how the instruments work). Furthermore, teaching associates were also more likely to select a neutral or disagree/strongly disagree response to many of the positive items on the survey (*e.g.* I think the students will learn critical thinking skills). It would appear that the teaching associates held a far more negative, or perhaps pragmatic, view of the students during teaching laboratories. Further investigations (such as interviews or focus groups) would be required to probe further into the exact nature of this difference.

The viewpoints of academic staff were also sought. 102 academic responses were collected and were very similar to those of the teaching associates (*i.e.* they also tended to think that students would undertake actions that would negatively affect their meaningful learning). However, when considering the positive items on the survey, academic staff were even more likely to select neutral, disagree or strongly disagree to a range of statements such as 'I think the students will consider if their data makes sense'. It would appear that academic staff held an even more negative, or more pragmatic, view of the students during an undergraduate teaching laboratory.

Overall, students appeared to be generally optimistic about their laboratory experiences, which was underestimated by both the teaching associates and the academic staff. It is important to recognise these differences of opinion in order to better manage the learning experience and expectations for both students and teaching staff. This is particularly important where mismatches occur on specific items that are potentially avoidable (such as decision making or concerns around timely completion of teaching laboratories).

Conflicts of interest

There are no conflicts to declare.

Appendix

The questions asked of the students and teaching staff are shown below in Table 8.

Table 8 Questions asked on the MLLI survey. Note that students were asked 'I expect to...' whereas teaching staff were asked 'I think the students will...'

When performing experiments in a chemistry laboratory course, I expect/I think the students will...
1. Learn chemistry that will be useful in my life.
2. Worry about finishing on time.
3. Make decisions about what data to collect.
4. Feel unsure about the purpose of the procedures.
5. Experience moments of insight.
6. Be confused about how the instruments work.
7. Learn critical thinking skills.
8. Be excited to do chemistry.
9. Be nervous about making mistakes.
10. Consider if my data makes sense.
11. Think about what the molecules are doing.
12. Feel disorganized.
13. Develop confidence in the laboratory.
14. Worry about getting good data.
15. Find the procedures simple to do.
16. Be confused about the underlying concepts.
17. "get stuck" but keep trying.
18. Be nervous when handling chemicals.
19. Think about chemistry I already know.
20. Worry about the quality of my data.
21. Be frustrated.
22. Interpret my data beyond only doing calculations.
23. Please select both agree and disagree for this question.
24. Focus on procedures, not concepts.
25. Use my observations to understand the behaviour of atoms and molecules.
26. Make mistakes and try again.
27. Be intrigued by the instruments.
28. Feel intimidated.
29. Be confused about what my data mean.
30. Be confident when using equipment.
31. Learn problem solving skills.

Acknowledgements

The authors would like to acknowledge and thank, first and foremost, the students, teaching associates and academics who participated in this work, whose honest feedback will result in a far stronger learning experience for undergraduates at the participating institutions. The authors would also like to acknowledge the technical staff at Monash University, the University of New South Wales, the University of Sydney and the University of Warwick, without whom, this work would not have been possible. The authors acknowledge Monash University for funding, and hosting, the Transforming Laboratory Learning program. The authors also acknowledge the Monash–Warwick Alliance seed fund for additional funding. Ethics approval was obtained from the Monash University Human Ethics Research Committee (application number, 2016000584) and the University of Sydney Human Ethics Research Committee (application number, 2016/902). Work at the University of Warwick and the University of New South Wales was performed under the Monash University ethics application.

References

Barlett J. E., Kotrlik J. W. and Higgins C. C., (2001), Organizational research: determining appropriate sample size in survey research, *Inf. Technol., Learn., Perform. J.*, **19**(1), 43.

Bauer C. F., (2005), Beyond ''Student Attitudes'': Chemistry Self-Concept Inventory for Assessment of the Affective Component of Student Learning, *J. Chem. Educ.*, **82**(12), 1864, DOI: 10.1021/ed082p1864.

Bauer C. F., (2008), Attitude toward Chemistry: A Semantic Differential Instrument for Assessing Curriculum Impacts, *J. Chem. Educ.*, **85**(10), 1440, DOI: 10.1021/ed085p1440.

Bretz S. L., Fay M., Bruck L. B. and Towns M. H., (2013), What Faculty Interviews Reveal about Meaningful Learning in the Undergraduate Chemistry Laboratory, *J. Chem. Educ.*, **90**(3), 281–288, DOI: 10.1021/ed300384r.

Bretz S. L., Galloway K. R., Orzel J. and Gross E., (2016), Faculty Goals, Inquiry, and Meaningful Learning in the Undergraduate Chemistry Laboratory, in *Technology and Assessment Strategies for Improving Student Learning in Chemistry*, American Chemical Society, vol. 1235, pp. 101–115.

Bruck A. D. and Towns M., (2013), Development, Implementation, and Analysis of a National Survey of Faculty Goals for Undergraduate Chemistry Laboratory, *J. Chem. Educ.*, **90**(6), 685–693, DOI: 10.1021/ed300371n.

Bruck L. B., Towns M. and Bretz S. L., (2010), Faculty Perspectives of Undergraduate Chemistry Laboratory: Goals and Obstacles to Success, *J. Chem. Educ.*, **87**(12), 1416–1424, DOI: 10.1021/ed900002d.

Domin D. S., (2007), Students' perceptions of when conceptual development occurs during laboratory instruction, *Chem. Educ. Res. Pract.*, **8**(2), 140–152, DOI: 10.1039/B6RP90027E.

Gadermann A. M., Guhn M. and Zumbo B. D., (2012), Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide, *Pract. Assess., Res. Eval.*, **17**(3), 1–13.

Galloway K. R. and Bretz S. L., (2015a), Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory, *J. Chem. Educ.*, **92**(7), 1149–1158, DOI: 10.1021/ed500881y.

Galloway K. R. and Bretz S. L., (2015b), Measuring Meaningful Learning in the Undergraduate Chemistry Laboratory: A National, Cross-Sectional Study, *J. Chem. Educ.*, **92**(12), 2006–2018, DOI: 10.1021/acs.jchemed.5b00538.

Grove N. and Bretz S. L., (2007), CHEMX: An Instrument To Assess Students' Cognitive Expectations for Learning Chemistry, *J. Chem. Educ.*, **84**(9), 1524, DOI: 10.1021/ed084p1524.

Hawkes S. J., (2004), Chemistry Is Not a Laboratory Science, *J. Chem. Educ.*, **81**(9), 1257, DOI: 10.1021/ed081p1257.

Hodson D., (1990), A critical look at practical work in school science, *Sch. Sci. Rev.*, **70**(256), 33–40.

Hofstein A. and Lunetta V. N., (1982), The Role of the Laboratory in Science Teaching: Neglected Aspects of Research, *Rev. Educ. Res.*, **52**(2), 201–217, DOI: 10.3102/00346543052002201.

Hofstein A. and Lunetta V. N., (2004), The laboratory in science education: foundations for the twenty-first century, *Sci. Educ.*, **88**(1), 28–54.

Leal Filho W. and Pace P., (2016), *Teaching Education for Sustainable Development at University Level*, Springer.

Letton K. M., (1987), *A study of the factors influencing the efficiency of learning in a undergraduate chemistry laboratory, (M Phil)*, Glasgow, Scotland: Jordanhill College of Education.

Morton S. D., (2005), Response to ''Chemistry Is Not a Laboratory Science'', *J. Chem. Educ.*, **82**(7), 997, DOI: 10.1021/ed082p997.1.

Nachar N., (2008), The Mann–Whitney *U*: a test for assessing whether two independent samples come from the same distribution, *Tutor. Quant. Methods Psychol.*, **4**(1), 13–20.

Novak J. D., (1998), *Learning, creating, and using knowledge*, Mahwah, NJ: Erlbaum.

Nunnally J. and Bernstein L., (1994), *Psychometric theory*, New York: McGraw-Hill Higher, Inc.

Pearson K., (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **50**(302), 157–175, DOI: 10.1080/14786440009463897.

Sacks L. J., (2005), Reaction to ''Chemistry Is Not a Laboratory Science'', *J. Chem. Educ.*, **82**(7), 997, DOI: 10.1021/ed082p997.2.

Sheskin D. J., (2003), *Handbook of parametric and nonparametric statistical procedures*, CRC Press.

Stephens C. E., (2005), Taking Issue with ''Chemistry Is Not a Laboratory Science'', *J. Chem. Educ.*, **82**(7), 998, DOI: 10.1021/ed082p998.1.

Xu X. and Lewis J. E., (2011), Refinement of a Chemistry Attitude Measure for College Students. *J. Chem. Educ.*, **88**(5), 561–568, DOI: 10.1021/ed900071q.

3.1 Summary of findings

This part of the study provided useful information on the expectations of students and staff in undergraduate laboratory activities before the commencement of the changes through the TLL programme. The results indicated that:

- Students held relatively narrow views of the aims of teaching laboratory activities which appeared to be influenced by many prior expository experiences. This was seen to be true at three different institutions across all year levels. Students rarely raised the development of scientific theory or experimental design.
- As students progressed through their undergraduate experiences, they tended to focus less on consolidating theoretical principles and more on the development of practical and transferable skills.
- Teaching associates held similar views of the aims of teaching laboratories as the students, albeit with a greater focus on the development of skills (both practical and transferable). An additional aim (to develop an appreciation of safety practices) was also raised by the teaching associates, which is not surprising as their training and job role includes responsibility for safety during the laboratory activities.
- Academic staff held the narrowest viewpoint of teaching laboratory activities, neglecting to raise either preparation for the workforce or gaining general laboratory experience as significant aims. Academic staff tended to raise the development of practical skills, enhancement of theoretical understanding, the application of theory or the development of some transferable skills.
- With regards to their expectations on how they would act, feel and behave in a teaching laboratory, students held remarkably positive views over all three year levels at all four institutions investigated. Minor changes between year levels were noted with students becoming slightly more negative in their expectations as they proceeded through their respective undergraduate degrees.

- Whilst teaching associates tended to also believe that students would undertake actions that would result in meaningful learning, they were also very likely to believe that students would have negative experiences as well (e.g. to be frustrated).
- Like the teaching associates, academic staff tended to feel that students would undertake activities that would detract from their meaningful learning. However, academic staff were less likely than teaching associates to believe that students would even undertake actions that led to meaningful learning. These findings are consistent with the literature which states that students often have incredibly high expectations for their university experience whilst teaching staff often perceive students as lacking in motivation beyond a simple desire to obtain high marks (OECD, 2002).

Both studies appear to indicate a large gap between student and staff perceptions of how students will act, feel and think during teaching laboratories. This work could provide a solid dataset to generate discussion between staff and students to break down these opposing perceptions by either decreasing staff negativity or perhaps decreasing student naivety, or a mix of the two. Additionally, it is also worth discussing whether the laboratory activities themselves can be altered to overcome negative considerations (e.g. concerns around time constraints) or increase positive items (e.g. giving students more freedom to make decisions).

With respect to studying the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments, these student results formed a strong baseline from which to measure the impact of the new laboratory experiments generated through the TLL programme (Chapters 4 and 5) with discussion of the outcomes of the programme forming Chapter 6

Chapter 4 – Examples of specific laboratory activities generated/altered by the TLL programme.

The next research questions considered were 'What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the development of students' employability skills and their recognition of these skills?' (Page 30, Research Question 2) and 'What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' level of enjoyment of the laboratory exercises?' (Page 30, Research Question 3).

In order to measure student perceptions immediately after the completion of any laboratory activity, a survey was created through modifications to the work of Russell (2008). The original 24 question survey was designed to measure the effects of the inclusion of many research/inquiry-based laboratory activities into the undergraduate chemistry programme at Purdue University. In an early version of this survey (before its use in Chapter 5), only two major changes were made - reformatting and modification of eight question stems from positive to negative (to avoid acquiescence bias).

The survey was then used to investigate two of the three laboratory activities raised in Chapter 2 – 'Investigation into the Efficacy of a Digestive Enzyme (Beano™)' and 'Electronic Waste'. The results have been published in (or submitted to) *Monash Education Academy Digest* (Stephen R George-Williams et al., 2018a, 2018b) and are reproduced on the following pages.

Enhancing inquiry in the teaching laboratory; a consecutive expository/inquiry-based laboratory exercise.

Stephen R George-Williams,* Angela L Ziebell, Christopher D Thompson, Tina L Overton

School of Chemistry, Monash University, Clayton, Melbourne, VIC, 3168, Australia.

ABSTRACT

The pedagogy of a literature sourced laboratory exercise was changed from expository to expository/inquiry and its success measured using a literature-modified survey (particularly in terms of student engagement and student recognised skill development). Over two four-hour sessions, students (in groups of 3-4) investigated the efficacy of a digestive enzyme in breaking down polysaccharides into simple sugars, from an aqueous extract of common legumes. The first session was expository in style and utilised a commercially available glucometer to monitor the amount of glucose production throughout the reaction. The second session was inquiry-based where students utilised the technical skills developed in the first session to investigate the effects of a set of variables either from a given list or of their own choosing. The experiment was simple to perform and allowed students to a) practice analytical technical skills on a new instrument whilst b) encountering scientific methodology through method development/implementation and analysis of results. Survey results indicate high student engagement and strong student recognition of the development of scientific investigation skills.

KEYWORDS

Biochemistry, Laboratory Instruction, Hands-On Learning, Inquiry-Based Learning, Enzymes

INTRODUCTION

Typically, many teaching laboratories around the world utilise expository (or cookbook/recipe) activities, which, as Domin states (Domin, 2007), rely 'almost exclusively on laboratory manuals to create a situation where students perform the activity by following a prescribed procedure to experience a pre-determined outcome'. These experiences are often criticised for invoking little critical thought (Gallet, 1998; Hodson, 1990; Hofstein & Lunetta, 1982; Pavelich & Abraham, 1979) and often students are left with little to no understanding of the underlying science (Letton, 1987). That being said, these exercises often provide students with an experience that allows them to develop technical skills (Abraham, 2011), such as pipetting or titrating.

To address some of the issues associated with expository laboratory activities, a teaching method called inquiry-based learning (Bruner, 1961; Domin, 1999) was developed that required students to seek out the answers to scientific questions themselves rather than be given direct instruction (Cummins, Green, & Elliott, 2004). When new laboratory exercises have been devised with this inquiry-based learning in mind, the student and staff perceptions of these changes have been noted to be typically positive, albeit with teething issues arising from such a dramatic change in pedagogy (Basey, Maines, Francis, & Melbourne, 2014; Cummins et al., 2004; Schmidt, Rotgans, & Yew, 2011; Tosun & Taskesenligil, 2013).

The issue presented at Monash University was to develop a new teaching laboratory exercise for a second year Biological Chemistry course that allowed students to grasp the concepts of enzyme specificity as well as gain experience in designing their own experiment. Whilst this is a highly specialised topic area, the generation of any activity that allows for the development of specific skills or knowledge as well as cultivating an understanding of method development is of importance in any modern field that requires experimentation (e.g. computer science, psychology, IT and so on).

A journal article by Hardee *et. al* (Hardee, Montgomery, & Jones, 2000) described an experiment that monitored the production of glucose from the breakdown of polysaccharides extracted from split peas using water. The glucose level was monitored by the use of a glucometer, which utilised the redox chemistry of a specific enzyme, and a commercial enzyme product known as Beano™ catalysed the breakdown of the polysaccharides. Beano™ is a digestible tablet that contains the enzyme alpha galactosidase, which is able to covert complex polysaccharides in simple sugars such as glucose and fructose.

The ideas explored in this experiment were used as the basis of the development of the new undergraduate laboratory exercise. It was considered ideal to break the experience into two sessions. The first used an expository teaching style in order to teach the students relevant techniques. Students learnt to monitor the production of glucose from the action of a commercially available non-specific enzyme (i.e. the enzyme can break down different polysaccharides rather than just one) on an aqueous extract of split peas.

During their second session, students were allowed to design and implement their own investigation into the action of Beano™ or similar products. This session thus capitalised on the newly developed technical skills (from session one) in order to deliver an inquiry-based experience where students investigated the effects of other variables, such as pH, the presence of multivitamins or supplements, of dissolved metals, alternate soaked beans, alternate sources of alpha galactosidase and the presence of alcohol, on the rate of glucose formation.

Thus, this new experiment was designed to meet the dual requirements of a) delivering scientific content (specifically the activity of enzymes and how they can be affected by many factors) and b) allowing for the desired inquiry-based learning experience, which imparted experience of scientific methodology and developed experimental planning skills. Again, the specific details of the topic area are secondary to the dual focus of the activity, which is the true aim/challenge of this work.

METHODS

Student generated data was collected throughout the 2-week experiment and during trials. Teaching Associates and other teaching staff were also consulted throughout the teaching period.

After the completion of the activity, students were invited to fill out a survey of 24 closed Likert questions (which focused on student motivation, recognition of industrial/research context and the ease of the practical) alongside three open questions. The Likert questions were obtained from an available PhD thesis (Russel, 2008) and the following changes were made:

i) The formatting was altered to be more readable,

ii) Two questions were merged (as it was considered unlikely that undergraduate students could recognise the difference between a Masters and PhD degree),

iii) Any reference to 'chemistry course' was changed to 'lab' in order to ensure the student correctly thought about the individual laboratory experience,

iv) A distractor question ('Please select Agree and Disagree to this question') was added to enable the removal of inappropriate responses, such as selecting 'Agree' to all statements on the survey.

The three open questions were 'What skills did you develop throughout the two week exercise?', 'Was there anything that could be improved about the two week lab?' and 'Did you enjoy the overall lab? Why/why not?'. Analysis of the quantitative data was achieved through the generation of frequency graphs (i.e. the percentage of students who picked a given response) whilst qualitative data was analysed using an inductive coding approach (Thomas, 2006). Inductive coding involved student responses being read and re-read until common themes emerged, after which the number of times a theme arose throughout the data set was collated.

RESULTS AND DISCUSSION

The enzyme in this experiment was used in the form of a commercial product called Beano™, which acts in the gut to break down many polysaccharides found in commonly consumed legumes. Glucometers were purchased from a local pharmacyand used to monitor glucose production over time (in this case, the AccuCheck Performa Nano Blood Glucose Meter). It is worth noting at this point that whilst the initial cost of this experiment is relatively low (a single reusable glucometer costs ~$40-50), the costs of the strips required (~$30-40 for 100 strips) can add up as the students typically required 80-100 strips per group of 3-4 students over the two sessions. This cost was not considered prohibitively high.

Before running the experiment during the semester, two trials were undertaken with eight high school students and seven undergraduate students (who had been at university for between one and three years), respectively. Both groups completed the experiment within one day (two four hour sessions), with the high school students being given a more prescriptive manual due to their limited technical experience. Regardless of the group, the results were very similar to those gathered by the authors, indicating the experiment was achievable by students from high school up to third year undergraduate students.

After the trial runs, the laboratory exercise was administered to a second year cohort of approximately 44 students. Reports from teaching staff, technical staff and students indicated that the laboratory exercise ran smoothly with a communication error (students weren't correctly informed how to prepare for week 2) being the only notable issue. Engagement was reportedly very high, especially during the student driven practical exercise in week 2. Investigations undertaken by the students provided fruitful data that corresponded to the results seen in previous trials.

The aforementioned survey was administered to the students on completion of the exercise with 28 responses (~64% of the cohort) collected. Through the open questions, about 82% of the students who responded, indicated that they enjoyed the class with almost half directly citing the open investigation style as the reason they enjoyed it. Examples included:

‘It was interesting to design our experiments.’

‘It was fun because we answered real life questions ourselves.’

‘I can learn how to plan an experiment to investigate something new.’

Other positive responses included the ‘real-world’ connection, simple enjoyment or ease of the overall laboratory exercise. Negative responses were few and included boredom, poor group communication and poor time management. Of those that answered the question pertaining to potential improvements to the laboratory exercise, 36% of the students did not feel any adjustments needed to be made. The students who did suggest improvements stated a desire for more guidance and raised a few minor procedural issues. When asked what skills they believed they developed during the exercise, the most common responses related to investigation skills. Examples included:

‘Planning skills - Design the experiment coming in’

‘How to construct a method to investigate questions that are applicable for the real world’

‘Thought process of designing own experiment, i.e. variables, method, equipment, etc.’

These results indicate that the experiment was successful in creating an authentic experience for the students. Furthermore, the exercise avoided overloading the students with new techniques whilst teaching and imparting experimental design.

From the closed Likert questions, the student responses to three separate questions resulted in notable neutrality (>33% of the responses). This highlighted that at least a third of the students:

didn't realise or appreciate the similarity of an authentic investigation to what was undertaken; appeared unsure if their knowledge had increased (or decreased); couldn't ascertain if the investigation lab motivated them to do well and couldn't decide if they gained any useful knowledge from the experience. These issues are potentially a result of poor communication between teaching staff and students, as it became clear (through conversations with teaching staff) that unprepared students in week 2 were more frustrated and had more difficulty with the activity. It is believed that this frustration made it more difficult for these students to appreciate the activity and to undertake the desired learning outcomes. This issue could likely be addressed through better staff training. A final three questions showed bimodal data, which related to the use of the lab manual or whether a student repeated any results. This was to be expected considering the more open nature of the laboratory exercise.

CONCLUSION

The use of sequential teaching methods, expository in session one and inquiry-based in session two, allowed the students to learn technical skills in a less stressful environment whilst later gaining the positive outcomes associated with a student driven scientific investigation (namely, engagement and enhanced scientific methodology). Responses to the closed survey questions showed that a lack of clear communication potentially undermined the desired uptake of scientific knowledge and an appreciation of the authentic nature of the exercise. Regardless, the responses to the open questions indicated an overall positive experience, such as:

'[The exercise] is different to other repetitive pracs … gives you an idea of applications'

'[The exercise] made me realise that the science we are conducting is used in the real world'

'We got to investigate our own hypothesis and were allowed to come up with our own experiment.'

Overall, the generation of the dual-focused experiment appears to be a success and highlights the use of the consecutive model (expository then investigation) in both delivering new content *and* allowing for a deepening understanding and appreciation of experimental design and development.

AUTHOR INFORMATION

Corresponding Author

*E-mail: [redacted]

Co-Authors

E-mail: [redacted]

E-mail: [redacted]

E-mail: [redacted]

ACKNOWLEDGMENTS

Ethics approval was obtained from the Monash University Human Ethics Research Committee, application number 2016000584. The authors would like to acknowledge and thank, first and foremost, the students who participated in the trials, whose honest feedback resulted in a far stronger learning experience. The authors would also like to express their gratitude to the technical staff at Monash University for their constant support and advice. Lastly, the authors acknowledge Monash University for funding, and hosting, the Transforming Laboratory Learning program.

REFERENCES

Abraham, M. R. (2011). What Can Be Learned from Laboratory Activities? Revisiting 32 Years of Research. *Journal of Chemical Education, 88*(8), 1020-1025. doi: 10.1021/ed100774d

Basey, J. M., Maines, A. P., Francis, C. D., & Melbourne, B. (2014). An Evaluation of Two Hands-On Lab Styles for Plant Biodiversity in Undergraduate Biology. *CBE-life sciences education, 13*(3), 493-503. doi: 10.1187/cbe.14-03-0062

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*, 21-32.

Cummins, R. H., Green, W. J., & Elliott, C. (2004). "Prompted" Inquiry-Based Learning in the Introductory Chemistry Laboratory. *Journal of Chemical Education, 81*(2), 239. doi: 10.1021/ed081p239

Domin, D. S. (1999). A Review of Laboratory Instruction Styles. *Journal of Chemical Education, 76*(4), 543. doi: 10.1021/ed076p543

Domin, D. S. (2007). Students' perceptions of when conceptual development occurs during laboratory instruction. *Chemistry Education Research and Practice, 8*(2), 140-152. doi: 10.1039/B6RP90027E

Gallet, C. (1998). Problem-Solving Teaching in the Chemistry Laboratory: Leaving the Cooks. *Journal of Chemical Education, 75*(1), 72. doi: 10.1021/ed075p72

Hardee, J. R., Montgomery, T. M., & Jones, W. H. (2000). Chemistry and Flatulence: An Introductory Enzyme Experiment. *Journal of Chemical Education, 77*(4), 498. doi: 10.1021/ed077p498

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review, 70*(256), 33-40.

Hofstein, A., & Lunetta, V. N. (1982). The Role of the Laboratory in Science Teaching: Neglected Aspects of Research. *Review of Educational Research, 52*(2), 201-217. doi: 10.2307/1170311

Letton, K. M. (1987). *A study of the factors influencing the efficiency of learning in a undergraduate chemistry laboratory,.* (M Phil), Jordanhill College of Education,, Glasgow, Scotland.

Pavelich, M. J., & Abraham, M. R. (1979). An inquiry format laboratory program for general chemistry. *Journal of Chemical Education, 56*(2), 100. doi: 10.1021/ed056p100

Russel, C. B. (2008). *Development and Evaluation of a Research-Based Undergraduate Laboratory Curriculum.* (PhD), Purdue University, West Lafayette, Indiana.

Schmidt, H. G., Rotgans, J. I., & Yew, E. H. J. (2011). The process of problem-based learning: what works and why. *Medical Education, 45*(8), 792-806. doi: 10.1111/j.1365-2923.2011.04035.x

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation, 27*(2), 237-246.

Tosun, C., & Taskesenligil, Y. (2013). The effect of problem-based learning on undergraduate students' learning about solutions and their physical properties and scientific processing skills. *Chemistry Education Research and Practice, 14*(1), 36-50. doi: 10.1039/C2RP20060K

Electronic Waste – A case study in attempting to reduce cognitive overload whilst increasing context and student control in an undergraduate laboratory.

Stephen R George-Williams,* Angela L Ziebell, Christopher D Thompson, Tina L Overton

School of Chemistry, Monash University, Clayton, Melbourne, VIC, 3168, Australia.

ABSTRACT

During a three hour session, students individually investigated the reaction of colourful aqueous metal salts with various bases and acids, the total number of which was drastically reduced from previous years. Following this, students then identified the composition of an unknown solution of two metal salts. A partially filled flowchart was used, rather than a step-by-step list of instructions, designed to allow for more student choice and a more genuine scientific environment. The experiment was simple to perform and was contextualised in metal wastes produced by modern electronic devices, although this context was not readily noted by the students. Student survey results and Teaching Associate interviews highlighted the success of the non-traditional manual and handout format whilst the contextualisation of the experiment will require further work to become truly effective.

KEYWORDS

Non-traditional manual, Student driven, Undergraduate chemistry

INTRODUCTION

The formation of metal compounds and their precipitation are possibly some of the most ubiquitous laboratory activities in first year chemistry undergraduate courses, with many even freely available online ("Chem lab 7: Precipitation and Complex Formation," ; "Experiment 2-3 Qualitative Analysis of Metal Ions in Solution," 2017; "Lab 4 - Qualitative Analysis," 2017; "Qualitative Analysis of Cations," ; "A Study of Transition Metal Ions,"). Prior to 2016, Monash

University asked students to undertake one of these experiments as part of an introductory chemistry course. Many issues were noted by teaching staff, with the most common being the lack of student engagement and lack of understanding of the scientific content. It was also highlighted by the technical staff that the number of chemicals required (22) was incredibly difficult to manage, with several being toxic (namely mercury(II) nitrate and lead(II) nitrate) or likely to stain skin (silver nitrate).

With this in mind, a new laboratory activity was needed that tackled the issues mentioned above whilst delivering the same learning to students. The first issue to be addressed was the lack of student engagement with the scientific content, which was potentially due the large amount of concepts that students were expected to learn and utilise, which could be causing cognitive overload.

Cognitive load refers to the number of discrete pieces of information that one can have in their working (or short-term) memory at any given moment (Kirschner, Sweller, & Clark, 2006). Typically, this number is around seven ± two (Miller, 1956) and if cognitive overload occurs during instruction, the learner experiences reduced performance (and subsequently, reduced engagement) on the given task (Mayer & Moreno, 2003). Considering the original activity required students to consider at least eleven different concepts, alongside the normal pressures of any laboratory experience, it is likely that at least some students encountered this issue. Hence, many of these concepts were removed in the new experiment, which simultaneously addressed the issue of the number of chemicals required and their toxicity (as mercury, silver and lead compounds were simply removed). Whilst this does result in the students encountering fewer concepts overall, it was considered that a few concepts learnt well was far better than many concepts learnt poorly or, at worst, none learnt at all.

Additionally, the original layout of the laboratory exercise had no obvious connection to the real life experiences of the students, it was simply embedded in the theory of the course. This was considered an issue as ‘Learning is essentially a matter of creating meaning from the real activities of daily living’ (Stein, 1998). As the original experiment had no connection to real life, it was thought that this may be what was causing the students to disengage from the material or, at least, compounding the previously discussed cognitive load issue.

Therefore, a real world context was sought that would allow the students to encounter the relevant scientific material but simultaneously connect it to their daily lives. In this case, the context was the electronic waste generated from their mobile phones after disposal, which have large negative impacts on the environment if left in landfill. Whilst current methods of extracting metals from electronic waste generally require pyrometallurgical processes (Kang & Schoenung, 2005) (i.e. separation through heating) to recycle metals from electronic waste, processes do exist for separation based on metal reactivity and solubilisation (Cui & Zhang, 2008). Whilst the experiment the students performed was not a perfect comparison to these processes, the link to real world processes was believed to be a strong one.

It was also considered prudent to consider the design of the manual (and handout) which, in this case, were highly prescriptive and detailed each individual step to be performed. Whilst this style of laboratory, known as expository or ‘recipe-based’, is relatively good at developing practical skills (Abraham, 2011), it generally doesn’t allow students to think critically (Gallet, 1998; Hodson, 1990; Hofstein & Lunetta, 1982; Pavelich & Abraham, 1979) or fully grasp the underlying science (Letton, 1987). To combat this, the learning theory known as Constructivism which ‘underlies the assumption that learning is an active process where knowledge is constructed based on personal experiences and the continual testing of hypotheses’ (Leal Filho & Pace, 2016) was utilised. It was desired that students be given a chance to form their own hypotheses (and to test them) by providing a manual (and handout) design to allowd students to choose which reaction to investigate at any given time, through the use of partially filled flowcharts. It was further hoped that by giving the students more control of the experiment, their engagement in the activity would also increase. Additionally, by slowly removing the expository nature of the instructions and allowing for easy repetition and correction of minor mistakes, it was hoped that students would be less challenged by completely non-expository activities delivered in later years of their undergraduate experience.

Overall, this experiment provides a reasonable case study for removing excessive amounts of content whilst simultaneously increasing student control, critical thinking and introducing motivating context – which is a worthy goal for any experience taught at university, rather than just science taught in undergraduate laboratories.

RESULTS AND DISCUSSION

Before the first laboratory session was even run, it was already noted by the technical staff that the experimental setup for this new exercise was significantly easier than previous years with the number of reagents needed to be generated being noticeably lower and the more toxic chemicals (mercury and lead nitrates) no longer present. Hence, the new laboratory exercise was, at the very least, succeeding from a safety and green chemistry perspective.

After being rolled out to first year students, feedback from Teaching Associates (TAs) and other teaching staff about the outcome of the laboratory activity were remarkably positive. Six TAs participated in recorded interviews about the new laboratory activity and stated many positive outcomes including, but not limited to:

> ‘they’ve [the students] got more time [compared to previous years] for them to actually sit down and think about the equations’

> ‘I think normally my students, they struggle for time … but I think it was a better alternative [to previous years] because it let them sit down and really think about it’

> ‘I just felt that, for the student, there is a lot to learn for them I think. I know it’s difficult in the beginning to, especially when they don’t have the background, but concept wise I think they picked it up a lot faster [compared to previous years]’

Overall, these responses indicate that the students seemed to handle the concepts well and the purposeful removal of a large amount of the content lead to an environment of reduced cognitive load. Student assessment marks (average of 82% ± 10%) were not significantly different from the prior year (average of 75% ± 14%), but do indicate a potential trend of a slightly higher average with a lower spread of marks, which may also indicate greater student understanding.

One of the other aims of this new laboratory activity was to provide the students with a real world context to their investigations – specifically the concept of electronic waste and identifying the metal salts that leach from them into landfills. When discussing the laboratory exercise in recorded interviews, none of the TAs mentioned the new context whatsoever, either indicating that they were mostly unaware of it or simply didn’t feel it was an important component to the success of the lab. Student responses to a previously utilised survey (George-Williams, Soo, Ziebell, Thompson, & Overton, 2018) indicated that students didn’t appear to attribute importance to the context with only 12 out of 118 responses directly citing the real life context as a reason for their enjoyment of the lab. Furthermore, one student even stated ‘I don't think it was related back to the theory of ‘ewaste’ well - I still don't understand the concept’.

Whether this is a purely isolated view is hard to determine, but it would appear that the electronic waste context was lost on the students and TAs. Future changes could include more context based questions in the assessment or pre-laboratory activities and encouraging TAs to strengthen the context for the students throughout the experiment.

The other major change made throughout the experiment was the non-traditional instructions in the manual and handout, namely partially filled flowcharts in place of a step by step instruction list. This was mentioned many times by both students and TAs in the surveys and recorded interviews respectively. When asked about potential improvements to the laboratory exercise, ~45% of the student responses were in relation to a desired improvement to the instructions in the manual and handout. Whilst this appears negative, it should be noted that it was expected that students would be challenged by the non-traditional template in order to develop skills beyond simply following a list. TAs clearly encountered this mindset, which was evident in their responses:

> ‘I think they read the lab manual and they start to think - oh I don’t understand what the lab manual is saying without looking at the [handout] - Yeah, so then you have to guide them’
>
> ‘I started them off with ‘you’re confused just do the experiment first, forget about the theory’.

However, even noting these issues, they also mentioned the benefits of the new style, such as:

‘I think the first thing is more questions [from the students]’

‘I was showing them how chemistry works, how tests work’.

Overall, whilst the design of the manual and handout could be enhanced (e.g. by further clarifying how to use the flowchart), it would seem the students benefit from the new style, even if they didn’t fully appreciate it.

Finally, in the survey the students were asked about the skills they developed in the laboratory exercise. Typically, in a normal expository laboratory exercise, students focus on the practical skills that have developed (e.g. use of a burette (George-Williams et al., 2018)) whereas, in this case, the most common answers were the use and development of observational skills and note taking skills. Hence, students were recognizing the development of skills beyond the simple technical skills developed during most laboratory activities. Importantly, considering the chemical concepts that were not removed, this seemed to be achieved without sacrificing the theory being taught, as in earlier discussions.

CONCLUSION

A one-session (three hours) laboratory exercise investigating the formation of complexes and precipitates of metal salts was developed to replace a content heavy experience. The large decrease in content from previous years reduced the overall stress of the learning activity and allowed students the time they needed to actually comprehend the science involved in the reactions of the metal salts. Additionally, the removal of a large number of reagents eased the difficulty in setting up the laboratory exercise as well as increasing the safety and 'green' nature of the experiment. Unfortunately, the addition of a real-world context (electronic waste) does not seem to have had a large impact on the students or the Teaching Associates, at least from the data collected through the survey and interviews. Further work will need to be undertaken to maximize the effect of context on the students' overall learning and engagement. Lastly, the new non-prescriptive manual and handout style challenged the students but, according to the TAs, resulted in a more scientific, inquiry driven environment.

Overall, this case study indicates the importance of varying activities (to achieve broader skill development) whilst regulating the amount of content being delivered (to avoid cognitive overload). The failure of the context serves as a reminder that Teaching Associates (or similar teaching staff) must be significantly convinced of teaching developments, lest they fail short due to simple disengagement.

AUTHOR INFORMATION

Corresponding Author

*E-mail: [redacted]

Co-Authors

E-mail: [redacted]

E-mail: [redacted]

E-mail: [redacted]

ACKNOWLEDGMENTS

The authors would like to acknowledge the Monash University Human Research Ethics (MUHREC) for approving the surveys and interviews, application number 2016000584. The authors would also like to express their gratitude to the technical staff at Monash University for their constant support and advice. The authors also acknowledge Monash University for funding, and hosting, the Transforming Laboratory Learning program. Lastly, a big thankyou to the Teaching Associates and students who participated in the surveys and interviews – without whom this study would have been impossible.

REFERENCES

Abraham, M. R. (2011). What Can Be Learned from Laboratory Activities? Revisiting 32 Years of Research. *Journal of Chemical Education, 88*(8), 1020-1025. doi: 10.1021/ed100774d

Chem lab 7: Precipitation and Complex Formation. Retrieved January 2017, from https://www.coursehero.com/file/10161724/Chem-lab-7-Precipitation-and-Complex-Formation/

Cui, J., & Zhang, L. (2008). Metallurgical recovery of metals from electronic waste: A review. *Journal of Hazardous Materials, 158*(2–3), 228-256. doi: http://dx.doi.org/10.1016/j.jhazmat.2008.02.001

Experiment 2-3 Qualitative Analysis of Metal Ions in Solution. (2017). Retrieved January 2017, 2017, from http://src.gov.jm/wp-content/uploads/2013/02/2-3-qualanalysis.pdf

Gallet, C. (1998). Problem-Solving Teaching in the Chemistry Laboratory: Leaving the Cooks. *Journal of Chemical Education, 75*(1), 72. doi: 10.1021/ed075p72

George-Williams, S. R., Soo, J. T., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2018). Inquiry and industry inspired laboratories: The impact on students' perceptions of skill development and engagements. *Chemistry Education Research and Practice*. doi: 10.1039/C7RP00233E

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review, 70*(256), 33-40.

Hofstein, A., & Lunetta, V. N. (1982). The Role of the Laboratory in Science Teaching: Neglected Aspects of Research. *Review of Educational Research, 52*(2), 201-217. doi: 10.2307/1170311

Kang, H.-Y., & Schoenung, J. M. (2005). Electronic waste recycling: A review of U.S. infrastructure and technology options. *Resources, Conservation and Recycling, 45*(4), 368-400. doi: http://dx.doi.org/10.1016/j.resconrec.2005.06.001

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi: 10.1207/s15326985ep4102_1

Lab 4 - Qualitative Analysis. (2017). *Lab 4 - Qualitative Analysis.* Retrieved January 2017, 2017, from http://www.webassign.net/question_assets/ncsugenchem102labv1/lab_4/manual.html

Leal Filho, W., & Pace, P. (2016). Teaching Education for Sustainable Development at University Level: Springer.

Letton, K. M. (1987). *A study of the factors influencing the efficiency of learning in a undergraduate chemistry laboratory,.* (M Phil), Jordanhill College of Education,, Glasgow, Scotland.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 43-52. doi: 10.1207/S15326985EP3801_6

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. doi: 10.1037/h0043158

Pavelich, M. J., & Abraham, M. R. (1979). An inquiry format laboratory program for general chemistry. *Journal of Chemical Education, 56*(2), 100. doi: 10.1021/ed056p100

Qualitative Analysis of Cations. Retrieved January 2017, from https://www.dartmouth.edu/~chemlab/chem3-5/qual_cat/full_text/procedure.html

Stein, D. (1998). Situated Learning in Adult Education. ERIC Digest No. 195.

A Study of Transition Metal Ions. Retrieved January 2017, from https://www.deltacollege.edu/emp/ckim/Labs_PDF/Lab%2043%20-%20Coordination%20Chemistry.pdf

4.1 Summary of findings

In these two publications, 28 (or 47% of the cohort) and 160 (or 59% of the cohort) responded to the survey for the Beano and Electronic Waste laboratories, respectively. The results for these two teaching laboratories indicated that the survey showed clear utility in being able to measure how students perceived new laboratory activities with regards to their enjoyment, their perceived usefulness and what skills they felt they had developed.

For the Beano™ investigation, results directly correlated with the inclusion of an inquiry-based activity i.e. students directly reference inquiry/investigation both regarding skill development and enjoyment. For both the Beano™ investigation and the Electronic Waste activity, context was also detected as a major theme with students raising it as a significant reason for enjoying in laboratory. Furthermore, the survey allowed for a measurement of the students' perceptions of the skills they felt they had developed even in the Electronic Waste laboratory, showing a large diversity of transferable skills, even when inquiry-based activities were not fully utilised and practical skills were very straightforward i.e. just simple dropwise addition and observation.

The closed Likert questions were generally met with positive responses, making detailed analysis difficult. The only deviations from this was students responding in a more neutral manner (>33% of the cohort) when asked about the similarity of the experiment to an authentic investigation, whether their knowledge level had increased or decreased, if the laboratory motivated them to do well or, finally, if they had gained any useful knowledge from the teaching laboratory. However, it was often difficult to ascertain whether students found the tasks easy/challenging, open/closed or even interesting or worthwhile with many responses to the open questions often only very brief in their descriptions. Hence, new closed questions were considered before using the survey after other laboratory activities and this is further discussed in Chapter 5.

Chapter 5 - Student perceptions of their enjoyment and skill development after individual redesigned laboratory activities.

Whilst the previous chapter touched on the first investigation of the research questions 'What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the development of students' employability skills and their recognition of these skills?' (Page 30, Research Question 2) and 'What was the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' level of enjoyment of the laboratory exercises?' (Page 30, Research Question 3), the scope was limited to only two new activities and utilised an early version of the survey ultimately utilised. The final changes made to the survey were the inclusion of six new closed questions (to cover student perceptions of enjoyment, difficultly, contextualisation and the openness of the experiments to student decisions), the removal of questions that held little relevance (e.g. The lab experience in this chemistry course has made me more interested in earning a Doctoral degree (Ph.D.) in a science field) and, lastly, the inclusion of three open questions to allow for greater depth in the students' responses.

This survey (Appendix 4) was delivered to students at the completion of 14 different laboratory activities (seven redeveloped by the TLL programme and seven prior to the TLL programme) and the results published in *Chemistry Education Research and Practice* (Stephen Robert George-Williams, Soo, Ziebell, Thompson, & Overton, 2018) as shown on the following pages.

Chemistry Education Research and Practice

PAPER

Cite this: *Chem. Educ. Res. Pract.*, 2018, **19**, 583

Inquiry and industry inspired laboratories: the impact on students' perceptions of skill development and engagements

Stephen R. George-Williams, * Jue T. Soo, Angela L. Ziebell, Christopher D. Thompson and Tina L. Overton

Many examples exist in the chemical education literature of individual experiments, whole courses or even entire year levels that have been completely renewed under the tenets of context-based, inquiry-based or problem-based learning. The benefits of these changes are well documented and include higher student engagement, broader skill development and better perceived preparation for the workforce. However, no examples appear to have been reported in which an entire school's teaching laboratory programme has been significantly redesigned with these concepts in mind. Transforming Laboratory Learning (TLL) is a programme at Monash University that sought to incorporate industry inspired context-based, inquiry-based and problem-based learning into all the laboratory components of the School of Chemistry. One of the ways in which the effect of the programme was evaluated was through the use of an exit survey delivered to students at the completion of seven experiments that existed before the TLL programme as well as seven that were generated directly by the TLL programme. The survey consisted of 27 closed questions alongside three open questions. Overall, students found the new experiments more challenging but recognised that they were more contextualised and that they allowed students to make decisions. The students noted the lack of detailed guidance in the new laboratory manuals but raised the challenge, context and opportunity to undertake experimental design as reasons for enjoying the new experiments. Students' perceptions of their skill development shifted to reflect skills associated with experimental design when undertaking the more investigation driven experiments. These results are consistent with other literature and indicate the large scale potential success of the TLL programme, which is potentially developing graduates who are better prepared for the modern workforce.

Received 29th November 2017,
Accepted 1st March 2018

DOI: 10.1039/c7rp00233e

rsc.li/cerp

Introduction

Transforming Laboratory Learning (TLL) is a programme at Monash University that was designed to significantly modernise the entire teaching laboratory programme in the School of Chemistry. This programme included 17 chemistry courses – four in first year, five in second year and eight in the third year, with the most significant changes focused towards the second and third years of the programme. Monash University is a large Australian university and the School of Chemistry has over 2000 enrolled students.

Several studies have suggested that chemistry graduates lack (or are unable to articulate) many transferable skills that are desired by employers, such as time management, independent learning and team-working (Hanson and Overton, 2010; Sarkar *et al.*, 2016). Even students who continue into research positions have been found to be lacking an appreciation of scientific methodology or experimental design as 'virtually no attention is given to the planning of the investigation or to interpreting the results' (Domin, 1999).

The skills agenda has gained prominence within Australia and is well exemplified by the 2016 governmental report (Norton and Cakitaki, 2016) which found that many undergraduates struggled to find work within four months of graduation, with science graduates faring less well than arts graduates. Monash University offers many internal programmes designed to enhance the employability of undergraduate students either through attempts to broaden skill development or work placements. However, until recently these have been largely extracurricular.

The TLL programme sought to enable undergraduate students to develop the skills they needed to obtain employment through a redesigned laboratory programme. In common with many other institutions, the original Monash University

School of Chemistry, Monash University, Victoria, 3800, Australia.
E-mail: [redacted]

laboratory programme relied heavily on traditional expository (or recipe-based) experiments. These types of laboratory activities (*i.e.* heavily prescriptive ones) are generally utilised to consume minimal resources whether these be time, space, equipment, or personnel (Lagowski, 1990). Whilst consumption is a major issue, it has been noted that students will often proceed through these experiences with little to no thought about the reasoning behind the procedures (Woolfolk, 2005). Furthermore, these experiences achieve little in the way of developing a wider range of transferable skills and usually lack any real-world context (Johnstone and Al-Shuaili, 2001; McDonnell *et al.*, 2007). To address this, a variety of different delivery methods have been attempted.

Industry inspired context and assessment

The use of industrial context in the design of new laboratory programmes has been achieved in many modern examples (Bingham *et al.*, 2015; Pilcher *et al.*, 2015; Erhart *et al.*, 2016). Typically, either the issues faced in an industrial problem or the actual methodology used in industry is used to guide the experimental design. Students generally respond positively to the experiences and respond in a manner consistent with context-based learning, *i.e.* they become more engaged with the material and tend to achieve higher levels of learning (Pringle and Henderleiter, 1999).

It should be noted that these same outcomes, *i.e.* high engagement with real industry inspired examples, could be achieved through the use of industrial placements or work-integrated learning (Cooper *et al.*, 2010). However, when cohort sizes are very large industrial placements are simply not a practical means to achieve this contextualisation. Hence, the inclusion of some work-integrated learning into the undergraduate teaching laboratories may bridge this gap.

Increased inquiry and student control

The use of inquiry-based learning is another common experimental design used to address the issues associated with expository experiments (Domin, 1999; Cummins *et al.*, 2004). In these cases, students are given a greater amount of freedom to either discover results for themselves (as opposed to simply confirming theory) or to choose how to undertake a given investigation (*i.e.* what methods to use) (Domin, 1999). These experiences are known to diversify the skills developed by students, particularly developing a greater range of transferable skills and a deeper understanding of scientific methodology (Weaver *et al.*, 2008). These experiences do however require significant scaffolding to support the students (Bruck and Towns, 2009). In fact, it has been noted that:

'*Inquiry lab students valued more authentic science exposure but acknowledged that experiencing the complexity and frustrations faced by practicing scientists was challenging, and may explain the widespread reported student resistance to inquiry curricula*'. (Gormally *et al.*, 2009)

Solving real-world problems

Problem-based learning can be described as a composite of inquiry-based learning and context-based learning, wherein students are given control to investigate a scenario that has been contextualised to the real-world before all appropriate content has been delivered to them (Duch *et al.*, 2001). As such, the benefits of both inquiry-based learning and context-based learning are achieved, with students reporting high levels of engagement whilst developing a wide range of scientific and transferable skills (Ram, 1999). The specific development of problem solving skills and strategies was shown in the work of Sandi-Urena *et al.* (2012) who used both quantitative and qualitative means to indicate a notable increase in these skills, even without explicit instruction. There are a large number of examples of problem-based learning in chemistry education in the literature which include individual experiments (Chuck, 2011; Mc Ilrath *et al.*, 2012; Dopico *et al.*, 2014), whole courses (Jansson *et al.*, 2015; Pilcher *et al.*, 2015) or even entire year levels (Kelly and Finlayson, 2007). However, there are no apparent examples of an entire school wide laboratory programme being reformed.

Staff involvement and training

It is also worth noting that throughout the TLL program, both Teaching Associates and technical staff were routinely involved in the generation of the new experiments. This approach was used in an attempt to ensure buy-in from all teaching and technical staff. Teaching Associate notes and guidelines were also generated throughout the program in an attempt to ensure the legacy of the TLL program.

Measuring the effect of TLL

The focus of this study was to identify a means by which to monitor and measure the overall effect of the TLL programme on a very large number of undergraduate students. Considering the numerous changes being implemented, it was decided that a survey would be an appropriate means to probe the effect on such a large cohort.

Through this survey, the effect of the TLL programme on a range of areas could be monitored. This survey was used in order to measure whether the programme was truly better at preparing the students for the workforce by creating a more engaging, industry-focused laboratory programme that allowed students to develop a wider range of transferable skills. This survey was designed to measure:

(1) The level of inquiry and contextualisation of the experiments, as noted by the students.

(2) The reported underlying motivations of the students.

(3) The overall perception of learning.

(4) Student perception of the skills developed in a given experiment.

(5) Student identified issues associated with the experiments.

(6) The level of (and reason for) the student enjoyment of the experiments.

Method

The aim of this study was to compare students' perceptions of a range of teaching laboratories that either existed prior to the

TLL programme or were produced as a result of the TLL programme. Industrial partners were sought to consult on the new implemented experiments. In many cases, the learning material was branded with their respective logos and the instructions were often written as though the industrial partners themselves were directing the students (*e.g.* 'Rationale, a Melbourne based skin care company, has asked Monash Consulting to investigate the use of a new range of compounds to be used as active ingredients in sunscreen'). Assessment was made more varied, often tailoring it to this new context. For example, students were asked to provide an executive summary to the company. This is an example of authentic assessment, wherein the assigned task matches real world procedures (Wiggins, 1990) and was used to further embed the experiments in the real world.

Alongside the inclusion of industrially relevant experiments, an increase in the level of inquiry was achieved by removing excessive guidance in the laboratory manual (which was replaced by prompting questions or multi-directional flowcharts), obscuring the experimental outcome that the students might achieve or through simply instructing the students to devise their own means by which to complete the experiment. All experiments generated from the programme were designed with the principles of problem-based learning in mind, either by providing a context or through ensuring at least a minimal level of inquiry.

Survey development

There are many surveys in the chemical education literature designed to measure students' attitudes, self-efficacy and overall learning in undergraduate chemistry courses (Bauer, 2005, 2008; Grove and Bretz, 2007; Cooper and Sandi-Urena, 2009). However, there are few that focus specifically on the undergraduate laboratory experience. Those that do, such as the MLLI (Meaningful Learning in the Laboratory Instrument) survey (Galloway and Bretz, 2015), focus on the students' expectations, thoughts and feelings rather than specific skills being developed. The MLLI survey also contains questions relating to perceived student control or the contextualisation of the lab, but in a non-direct fashion (*e.g.* 'I expect to learn chemistry that will be useful in my life' or 'I expect to make decisions about what data to collect').

A survey more suited to our purpose was found in the work of Russel (2008), which contained closed questions designed to monitor the effect of the inclusion of more research-based experiments in their undergraduate teaching laboratories. Many questions overlapped with the aims of TLL, especially those around context, the level of guidance within the lab manual or the underlying motivation or engagement of the students. This survey had been validated and was considered an ideal starting point for the final survey used in this study.

The Russel (2008) survey was modified over several iterations. The formatting was changed to be consistent with other studies being undertaken and any mention of 'chemistry course' was changed to 'lab'. Three questions were removed as they were considered to unlikely to be altered by a single laboratory experience (*e.g.* This lab experience has made me more interested in earning a Doctoral degree (PhD) or Master degree in a science field). Eight items were altered to their negative version (*e.g.* 'This lab experience made me learn' became 'This lab experience did not make me learn') to avoid students agreeing to every item, *i.e.* to avoid acquiesce bias (Watson, 1992). Six new closed questions were also added to capture the students' perceptions of the ease, challenge, contextualisation, openness or level of interest in the experiments. Finally, three open questions were added to further probe the students' perceptions. These questions were 'What skills did you develop throughout today's experience?', 'Was there anything that could be improved about today's lab?' and 'Did you enjoy today's lab? Why/why not?' The final version of the survey consisted of 27 closed questions, one distractor question ('Please select agree and disagree to this question') and three open questions.

Data collection

The survey was administered to students either in the teaching laboratory at the completion of an experiment or immediately after their laboratory session during a free lunch. In many cases, the same students filled out multiple surveys for multiple experiments. The seven experiments selected for the 'Pre-TLL' sample were chosen to represent traditional expository experiments, as opposed to any that already utilised inquiry-based learning. These seven experiments were chosen predominately due to convenience and to avoid overlap with other research programs (*i.e.* to avoid students filling out too many surveys in a given course). The seven 'Post-TLL' experiments were the first new experiments to be generated by the TLL programme. The surveys were disseminated in 2016 and 2017 – the first two years of the TLL programme. Consequently, students enrolled in chemistry courses in 2016 and 2017 could have seen the survey multiple times if they happened to be enrolled in the chemistry courses of interest. No repeat measurements of any experiment was performed.

The number of students varied between the courses, from approximately 100 to 1200+. In many cases where the student cohort was large, only a subset of the students were surveyed. For example, for the analysis of a new experiment in first year that ran twice a day for five days of the week, students were invited to complete the survey at the completion of the experiment on Tuesday and Thursday morning. Hence, not all students were surveyed but the sample was large and representative of the cohort.

Research theoretical framework

Through the use of both quantitative and qualitative questions, this research adopts a concurrent mixed-methods approach, *i.e.* both types of data were collected simultaneously (Creswell, 2009). It is believed that the students responded to the questions in a manner consistent with Constructivism which 'underlies the assumption that learning is an active process where knowledge is constructed based on personal experiences and the continual testing of hypotheses' (Leal Filho and Pace, 2016). Hence, the experiences that students have just had will directly impact their responses to the questions. The students were neither discouraged nor encouraged to discuss the survey with one another, so occasional collusion cannot be discounted and may lead to a small proportion of group responses rather than individual. The laboratory teaching staff were unaware of the content of the survey and were therefore unlikely to bias the viewpoints of the students through in-class conversations. The open questions allowed the students to respond in a less directed manner ensuring a greater depth to their responses.

Data analysis

The total number of responses are shown in Table 1 and completion rates range from 38% to 76% of the cohort surveyed. Further details regarding the individual experiments can be found in Table 2. The quantitative data was transcribed into Excel after recoding (*e.g.* Strongly Disagree = 1, Disagree = 2, Neutral = 3, Agree = 4 and Strongly Agree = 5) and then analysed (as frequencies, not averages) with the SPSS data analysis software. The responses to the open questions were transcribed verbatim into Excel and were analysed in Excel as well. Overall, there were 525 responses collected about experiments that existed prior to the TLL programme and 609 responds to those generated by the TLL programme.

For the quantitative analysis, the responses were combined into two major groups, the largely expository, unchanged experiments (Pre-TLL) and those generated or revised during the TLL programme (Post-TLL). Significant differences between student responses to each of the 27 closed questions where measured through the Wilcoxon Signed Rank test, which presumed dependence between students responding to both the Pre-TLL and Post-TLL surveys. For questions that showed a statistical difference to a 95% confidence interval, the resulting Z values (obtained from SPSS) were divided by the square root of the respective sample size of respondents to generate a measure of the effect size (r). The cut-offs for the 'size' of the effects were determined through the work of Hattie (2008), which was later extended (Fritz *et al.*, 2012; Lenhard and Lenhard, 2016) to r values. The ranges were defined by:

(1) $0 \leq r \leq 0.1$. 'Student effect size'. This refers to the natural variation in any group of students. For example, a more motivated student may respond more positively than a less motivated student.

(2) $0.1 < r \leq 0.2$. 'Teacher effect size'. This refers to the effect of a particularly motivated teacher over the course of a single year (*i.e.* this effect size could be achieved given time/motivation).

(3) $r > 0.2$. 'Zone of desired effect'. This refers to interventions that have an immediate impact and are where educators should typically focus their efforts.

The qualitative data was analysed for emerging themes in an inductive manner. Themes were generated through several rereads of the responses during which recurring themes were identified. These themes were studied in order to identify any redundancies and each theme was given a code. Themes extracted from the data and their codes were given to two other researchers who attempted to code a portion of the qualitative data for each of the three open questions. If needed, themes and codes were refined and these final themes were then used to recode the rest of the responses. This data was then expressed as a percentage of participants who raised a particular theme.

Table 1 The number of responses to the surveys and as a the percentage of students who had just completed the respective experiment

	Experiment name	Number of responses	% of students
Pre TLL	Rearrangement	51	49
	EAS	40	38
	Anthracene oxidation	91	76
	Macrocycles	37	46
	Isomerisation	49	61
	Proteins (2016)	70	58
	Panacetin	184	74
Post TLL	Proteins (2017)	138	69
	Sunscreen	51	49
	Pseudenol	65	63
	Electronic waste	160	59
	Nylon	69	58
	Food project	97	49
	Enzymes	28	47

Limitations

The first limitation of this study is the percentage of students responding (38–76%) and the overall sample sizes (28–184). The percentages of a given cohort responding varied from somewhat representative (38%) to highly representative (76%). In some cases, the cohort itself was relatively small (approximately 60) whilst others were impractically large (approximately 1200+). Therefore, this situation either resulted in small data sets (Enzyme experiment) or required the use of subsets (*e.g.* only surveying two out of eight lab classes), a process known as convenience sampling (Henry, 1990). Potentially, the use of subsets or small cohorts could lead to less statistically valid data or non-representative data.

This can be further considered through comparison to the work of Barlett *et al.* (2001) in which they provided acceptable minimal sample sizes for varying populations. However, this is complicated by the fact that the researchers only provided acceptable sample sizes for either continuous or categorical data, whilst this study utilises ordinal data. Even presuming the lower acceptable sample sizes for the continuous data set apply for an alpha of 0.05 (*i.e.* a 5% error), six of the seven sample sizes are considered below the required levels. This is simply due to the smaller cohorts enrolled in those courses. However, when the datasets are combined into Pre-TLL and Post-TLL, the number of responses (522 and 608 respectively) are considerably above the required number (106). Hence, whilst some individual experiments may not be fully represented through this analysis, the overall Pre-TLL *vs.* Post-TLL comparison has sufficient statistical power.

Additionally, as students may have filled out the survey on multiple occasions, it is possible that they simply became accustomed to the survey and simply answered the questions

Table 2 A brief description of the 14 experiments surveyed

	Experiment name	Year level – course focus	Method type	Context (real life or industry)	Scientific content	Additional notes
Pre TLL	Rearrangement	2nd – Inorganic and organic chemistry	Expository	None noted	Carbocation rearrangement	One four hour session. Historical methods used (*e.g.* hydrazone wet test). Students worked in pairs.
	EAS	2nd – Inorganic and organic chemistry	Expository	None noted	Electrophilic aromatic substitution	One four hour session. Typically completed within 2 hours. Students worked in pairs.
	Anthracene oxidation	3rd – Medicinal chemistry	Expository	Enzyme mimics	Oxidation using vanadium catalysts	One four hour session. Contains a long wait time of 2 hours. Underutilised context. Students worked in pairs.
	Macrocycles	3rd – Advanced inorganic chemistry	Expository (mimics literature)	None noted	Synthesis of macrocyclic cage complexes	One four hour session. Required students to obtain method from literature sources. Students worked in pairs.
	Isomerisation	3rd – Advanced inorganic chemistry	Expository	None noted	Kinetics of ligand isomerisation	One four hour session. Utilises kinetics/physical chemistry in student perceived synthetically focused course. Students worked in pairs.
	Proteins (2016)	2nd – Food chemistry	Expository	Food proteins	Protein detection and measurement	One four hour session. Context limited by use of non-commercially available defatted soy protein. Students worked in groups of 4.
	Panacetin	2nd – Inorganic and organic chemistry	Expository	Black market pharmaceuticals	Solute/solvent portioning	One four hour session. Typically completed within 2 hours (out of the 4 assigned). Students worked in pairs.
Post TLL	Proteins (2017)	2nd – Food chemistry	Expository	Food proteins	Protein detection and measurement	One four hour session. Commercially available milks used. Required students to obtain method from literature sources. Students worked in groups of 4.
	Sunscreen	2nd – Inorganic and organic chemistry	Expository (mimics literature)	Sunscreens and UV-active materials	Aldol condensation	One four hour session. Traditional synthesis with students aiming to make the best sunscreen. Students worked in groups of 3.
	Pseudenol (previously Panacetin)	2nd – Inorganic and organic chemistry	Flowchart/student directed	Black market pharmaceuticals	Solute/solvent portioning	One four hour session. Non-stepwise method, prompting questions used. Context strengthened. Students worked in pairs.
	Electronic waste	1st – Introductory chemistry	Flowchart/student directed	Metal wastes from electronic goods	Transition metals, complexes and colour	One three hour session. Non-stepwise method, students follow multi-directional flowchart. Students worked individually.
	Nylon	3rd – Materials chemistry	Inquiry/investigation	Production of Nylon	Step-growth polymerisation	One four hour session. Very simple method. Students allowed to investigate anything available. Students worked individually.
	Food project	2nd – Food chemistry	Inquiry/investigation	Nutritional components of foods	Methods used for non-ideal food samples.	2 four hour sessions, 1 week apart. Methods utilised in prior experiments of same unit. Students worked in groups of 4.
	Enzymes	2nd – Biological chemistry	Expository then Inquiry/investigation	Commercially available digestive supplements	Enzyme degradation of complex sugars	2 four hour sessions, 1 week apart. First week traditional method followed by inquiry-based second week. Students worked in pairs.

in a repetitive manner. The inclusion of negative stems and a distractor question hopefully forced students to stop and reread questions but this cannot be confirmed at this time.

Another potential limitation was the experiments that were chosen to represent the Pre-TLL dataset. These experiments were selected for practical concerns with timetabling and the delivery of other questionnaires not related to this study. Hence, some experiments that did not form part of this study may have affected the results obtained (*i.e.* if all teaching experiments in all chemistry courses were surveyed, different results may have been obtained).

It is also possible that the previous background of the students, in terms of encounters with other teaching laboratories could influence their responses. This cannot be completely discounted but, during the first year at Monash University, all students undertake multiple inquiry-based laboratories (called IDEA experiments) during semester one and two. Consequently, all students in higher years have at least some experience with inquiry experiments which would potentially mitigate this issue.

Beyond the performance of a Cronbach's alpha test, no further measurements of reliability or validity were performed. Hence, there may be issues regarding the validity or reliability of the instrument. However, the inclusion of qualitative data through the open questions and sourcing the bulk of the survey from a pre-validated and thoroughly tested source was believed to be sufficient to counter this concern.

Finally, the method of interpretation could be a source of error. However, it is believed that the iterative nature of the theme generation negated this issue to a significant degree, particularly through the use of multiple coders.

Results and discussion

As many changes were made to the literature version of the survey, a test of the internal reliability of the modified version was carried out using the SPSS data analysis software. After correcting for the negative items, the Cronbach's alpha value of the entire Pre-TLL or Post-TLL dataset was found to be 0.846 and 0.866 respectively. As the common literature value for an internally consistent survey is ≥ 0.7 (Santos, 1999), it appeared that the numerous changes to the survey did not effect this value. As such, the items in this test were all considered focused on the same item/concept (in this case, the single laboratory experience) and indicated the use of a reliable scale.

Quantitative analysis

The occurrence of any significant differences between students responding to the survey after a Pre-TLL or Post-TLL experience was investigated. It should be noted here that even though the sample sizes for individual experiments were very different (28–184), all data for all the experiments was combined together in order to measure the Pre-TLL and Post-TLL effects. Furthermore, the six new closed questions (focusing on the students' perceptions of the ease, challenge, contextualisation, openness or level of interest in the experiments) were added after analysis of some of the original experiments. Hence, the sample sizes are notably lower for those 6 new questions.

The Wilcoxon Signed Rank test showed that 18 of the 27 closed questions were answered in a significantly different way. All questions alongside the p, Z and r values are shown in Table 3. Of this, only eight showed an effect size within the 'zone of desired effect' (Hattie, 2008; Lenhard and Lenhard, 2016). Fig. 1 shows the responses to these eight questions.

It is worth noting that this analysis does not take into account the variation in sample sizes for the different experiments, which range from 28–184. To investigate the effect of this, 28 random responses were chosen from each data set (*e.g.* only 28 of the responses to the electronic waste experiment and so on for all of the other experiments). Following a Wilcoxon Signed Rank test, only one new question showed a significant difference (which focused on general chemistry understanding) but still exhibited a small effect size. Six other questions no longer showed a significant difference, although they previously exhibited effect sizes $r < 0.14$ and were considered irrelevant. One new question rose above the $r = 0.2$ threshold (which focused on student perception of the organisation of the experiment) whilst the effect size of the original eight questions increased by 0.02–0.08. Hence, at worst, combining the data for analysis causes the overall effect sizes to be underestimated for the original eight questions and these were considered to be the items most affected by the TLL programme. The full data for this analysis is present in the Appendix.

Fig. 1 indicates the direction of the change in the students' responses to any of the eight questions that showed a change

Table 3 The results of the Wilcoxon Signed Ranked test on the 27 closed questions showing the number of responses (N), the p result, the calculated Z value and the r effect size

Question	N	p	Z	r
This lab experience was worthwhile.	680	0.95	−0.1	—
This lab experience was interesting.	680	0.79	−0.3	—
This lab experience helped me better understand chemistry, in general, as a result of completing the chemistry lab.	1122	0.55	−0.6	—
In my life, I will *not* use skills I've learned in this chemistry lab.	1098	0.44	−0.8	—
This lab experience did *not* make me learn.	1118	0.43	−0.8	—
Having the opportunity to use chemistry instruments helped me learn course topics.	1115	0.40	−0.8	—
Even if I don't end up working in a science related job, the laboratory experience will still benefit me.	1111	0.33	−1.0	—
This lab experience has made me more interested in a science career.	1108	0.28	−1.1	—
This lab experience made me realize I could do science research in a real science laboratory (for instance at a university, or with a pharmaceutical company).	1114	0.076	−1.8	—
Having the opportunity to use chemistry instruments made this course *less* interesting for me.	1114	0.051	−1.9	—
This lab experience has made me less interested in science.	1117	0.045	−2.0	−0.06
This lab experience has made me more interested in chemistry.	1118	0.008	−2.7	−0.08
This lab experience helped me understand how the topics that are covered in chemistry lecture are connected to real research.	1119	0.003	−3.0	−0.09
This lab experience gave me a better understanding of the process of scientific research as a result of this experiment.	1106	0.002	−3.1	−0.09
Finding answers to real research questions motivated me to do well in the chemistry lab.	1115	0.001	−3.3	−0.10
Finding answers to real world questions motivated me to do well in the chemistry lab.	1106	$<5 \times 10^{-4}$	−3.5	−0.10
This lab experience presented real science to students, similar to what scientists do in real research labs.	1123	$<5 \times 10^{-4}$	−3.9	−0.12
This lab experience was *not* very similar to real research.	1120	$<5 \times 10^{-4}$	−4.3	−0.13
This lab experience was *not* well organized.	1112	$<5 \times 10^{-4}$	−5.1	−0.20
This lab experience was open enough to allow me to make decisions.	678	$<5 \times 10^{-4}$	−5.4	−0.16
This lab experience was easy.	594	$<5 \times 10^{-4}$	−6.0	−0.25
In this lab, the instructional materials did *not* provide me with explicit instructions about my experiment.	679	$<5 \times 10^{-4}$	−6.5	−0.25
This lab experience was well contextualised to real life or the workforce.	681	$<5 \times 10^{-4}$	−6.5	−0.25
This lab experience was challenging.	679	$<5 \times 10^{-4}$	−8.6	−0.26
In this lab, I did *not* repeat experiments to check results.	1121	$<5 \times 10^{-4}$	−8.6	−0.25
In this lab, the instructional materials provided me with sufficient guidance for me to carry out the experiments.	1117	$<5 \times 10^{-4}$	−9.3	−0.28
In this lab, I can be successful by simply following the procedures in the lab manual.	1118	$<5 \times 10^{-4}$	−9.3	−0.28

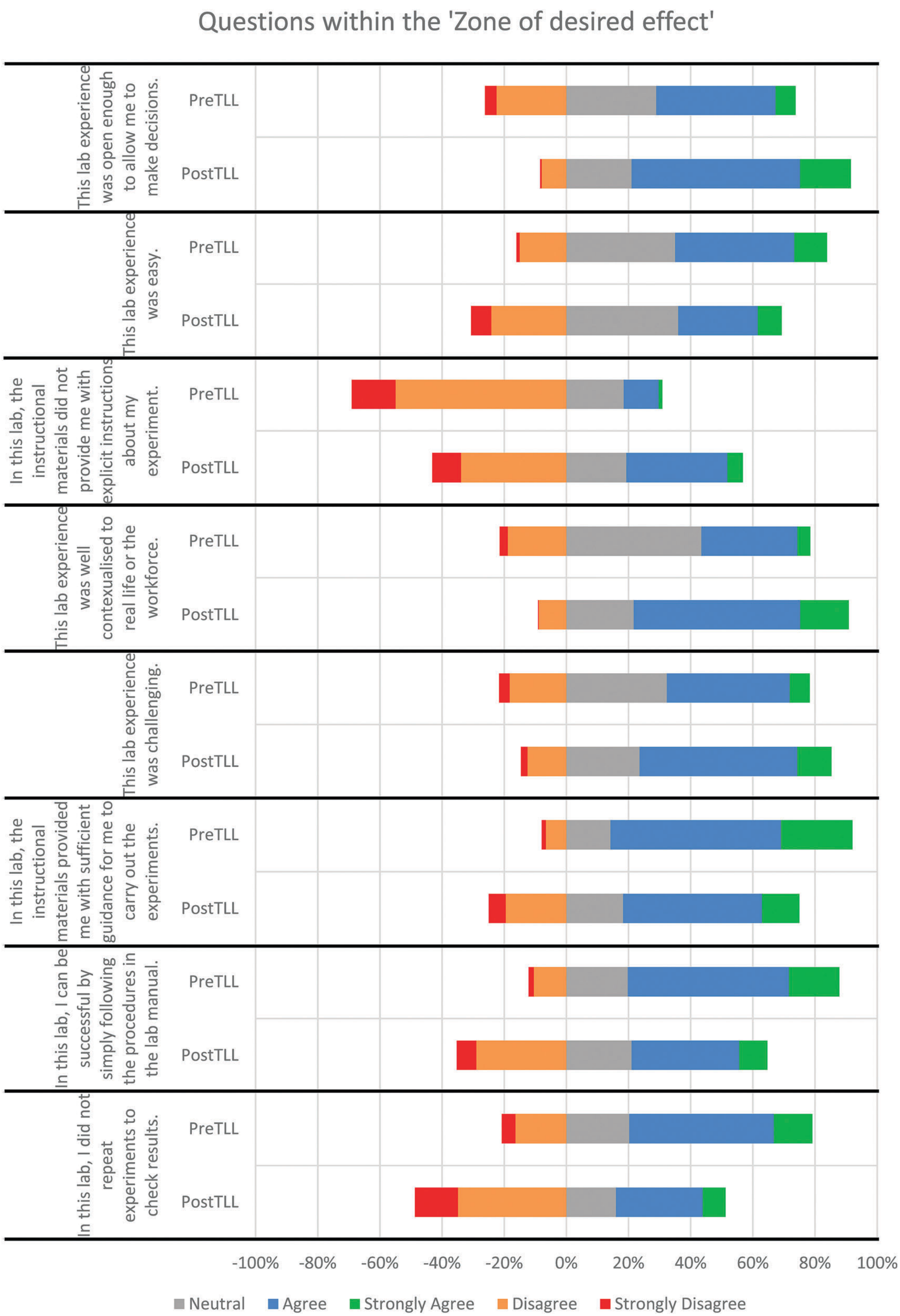


Fig. 1 Horizontal stacked bar charts for the eight questions showing an effect size within the 'zone of desired effect' Pre and Post TLL.

within the 'zone of desired effect'. For example, the top two horizontal stacked bars showed that the students' responses shifted right for the Post-TLL experiments compared to the Pre-TLL experiments. Hence, students were more likely to select neutral, agree or strongly agree to the concept that the experiment was open enough for them to make decisions in the

Post-TLL experiments. With the exception of the responses to the level of the contextualisation within the experiment (which saw a large decrease in neutral and a rise in agree and strongly agree), the shift of the horizontal stacked bars provides a simple approximation of the shift in the students' responses.

Overall, students responded in such a manner to imply that they found the Post-TLL experiments less easy whilst more challenging, contextualised, and open (*i.e.* they could make more decisions themselves). They were also more likely to repeat results and found that the laboratory manual offered less guidance and could not be relied upon in order to complete the experiment with no additional materials or aids. Overall, these results are considered very positive outcomes for the TLL programme as removing dependence on the laboratory manual was a key goal (as it was perceived that students over-relied upon it for guidance). This indicates a move away from expository recipe-style manuals and an increase in inquiry. Furthermore, the recognition of increased openness, challenge and contextualisation were also considered positive results.

It is also of interest to note the questions that showed little to no significant difference after the modified experiments. Questions referring to overall motivation (*e.g.* finding answers to real research questions/real world questions motivated me to do well), interest (*e.g.* This lab experience made me more interested in a science career or This lab experience was interesting/worthwhile) or even overall learning (*e.g.* This lab experience did not make me learn) showed little to no significant difference. It would seem that either even more significant changes to the laboratory programme would be required to effect these items or that the students were unlikely to become more positive. In this case, approximately 80% of students always stated that the lab was interesting or worthwhile so there was very little room for improvement no matter what changes were made. It is also possible that these inherently intrinsic factors (such as interest and motivation) are simply too innate to a given student and are unlikely to be influenced by external factors such as the type of laboratory experienced.

That being said, the closed questions, and the subsequent quantitative analysis, can only provide a surface analysis of the impact of the new experiments. Hence, discussion of the responses to the 3 open questions is required.

Perceived skill development

The first open question asked the students 'What skills did you develop throughout today's experience?'. It should be noted that the Panacetin experiment is absent from the open questions as the survey did not include these prompts at that time. For the other experiments, students raised between one and four different skills. Tables 4 and 5 show the top three skills (raised by ≥10% of the students, which was believed to result in meaningful data) for any given experiment.

The Pre-TLL results in Table 4 show several notable features. Firstly, the development of practical skills (student examples include '*Use of the glassware*' and '*Liquid–liquid extraction. TLC and how to interpret it. How to use pKA and separate solution mixture. Saw HNMR machine being used*') was a major theme for all six Pre-TLL experiments raised by 69–88% of the respondents. Secondly, many students stated a greater understanding of particular theoretical concepts (student examples include '*Improved my mechanism understanding*' and '*Understanding of how a catalyst works*') as a skill that they had developed, ranging from 13 to 32% of the responses. Thirdly, even though these

Table 4 The top three skills raised by ≥10% of the students to any Pre-TLL experiment

Experiment	Top three skills developed	Percentage of respondents (%)
Rearrangement	Practical skills	88
	Transferable skills	25
	Theory	13
EAS	Practical skills	76
	Theory	32
	Transferable skills	10
Anthracene oxidation	Practical skills	76
	Patience	17
	Theory	16
Macrocycles	Practical skills	69
	Following a literature method	16
	None	11
Isomerisation	Practical skills	84
	Transferable skills	49
	Theory	19
Proteins (2016)	Practical skills	84
	Transferable skills	35
	Teamwork	16

Table 5 The top three skills raised by ≥10% of the students to any Post-TLL experiment

Experiment	Top three skills developed	Percentage of respondents (%)
Proteins (2017)	Practical skills	80
	Transferable skills	35
	Teamwork	25
Sunscreen	Practical skills	76
	Transferable skills	43
	Time management	29
Pseudenol	Practical skills	71
	Transferable skills	36
	Critical thinking	10
Electronic waste	Transferable skills	45
	Observational skills	40
	Note taking	22
Nylon	Transferable skills	42
	Experimental design	39
	Practical skills	35
Food project	Transferable skills	76
	Teamwork	49
	Experimental design/practical skills	29
Enzymes	Experimental design	56
	Practical skills/transferable skills	33
	Teamwork	19

experiments were predominately expository, students often raised a range of transferable skills, including, but not limited to, time management, teamwork, critical thinking and communication (student examples include '*Teamwork skills and communication skills*' and '*Critically thinking about instructions*'). Individual transferable skills were generally not raised by more than 10% of the cohort, hence an overarching theme was generated that subsumed all transferable skills and was raised in four of the six Pre-TLL cases by 16–49% of respondents.

Deviations from the above observations can be explained by the nature of the experiments themselves – patience was raised in the anthracene oxidation experiment that involved a two hour reflux whilst following a literature method was raised in the macrocycles experiment as students were expected to find the method in the literature before arriving to laboratory session. Overall, the Pre-TLL results appear focused on the development of practical skills, limited transferable skills and regularly focus on theoretical understanding. As already noted, these are not unique but highlight the success of the survey at detecting these students' perceptions.

Analysis of the Post-TLL responses in Table 5 indicates a range of similarities and differences when compared to the Pre-TLL results. Three of the Post-TLL experiments (Proteins (2017), Sunscreen and Pseudenol) show very similar results to the Pre-TLL experiments. This is to be expected as these new experiments, whilst contextualised, were still focused on the development of new practical techniques and could be considered largely expository, albeit with more inquiry focus than many Pre-TLL experiments as the overall outcome of the experiments were discovered by the students. However, it is worth noting that in all three cases an individual transferable skill (teamwork, time management and critical thinking respectively) was now raised enough by the students to become one of the top three aims. Hence, these Post-TLL experiments were still being recognised as opportunities to develop practical skills whilst raising recognition of several transferable skills. This is likely to be a result of the increased connection to the students' daily lives and/or potential career paths when experiencing a conceptualised laboratory.

The Post-TLL experiments that allowed students to undertake experiments of their own design (Nylon, Food Project and Enzymes) showed common responses to one another. In all three cases, the development of experimental design skills (student examples include '*Creating methods*' and '*How to develop experiments to achieve certain aims or outcome*') were recognised by 29–56% of the students. The development of transferable skills also become much more prominent in the responses, with the extreme result of 76% of students raising them in the Food Project experiment. The development of practical skills was still raised in all of these experiments, but to a much lower degree (29–35%). This is likely an artefact of the students now raising a much larger breadth of developed skills.

The only case in which the development of practical skills was not a significant theme was the electronic waste experiment. This experiment involved the dropwise addition of metal ion solutions to a range of reagents and was, therefore, practically quite simple. Hence, other skills were raised by the students such as taking observations and making detailed notes.

Lastly, an increase in theoretical understanding was no longer raised as a common theme in the Post-TLL experiments. This would appear to contradict the quantitative results in which there was no notable difference in the students reported level of chemistry understanding or overall learning. However, it is possible that the new experiences simply provided a richer environment for skill development, which resulted in students identifying a broader range of skills that reduced the extent to which they viewed developing a deeper theoretical understanding as a skill that had been developed. Hence, it appeared to a lesser extent in the open answers but remained unchanged in the closed responses. This is particularly positive as the TLL experiments were designed to incorporate a larger diversity of learning experiences, rather than simply providing a chance to study a given theoretical principle. Overall, the Post-TLL results showed a larger range of skill development, particularly incorporating more transferable skills and experimental design. This was achieved without sacrificing the development of practical skills in the more expository experiences.

Improvements suggested by students

The second open question asked 'Was there anything that could be improved about today's lab?'. The top three issues (raised by $\geq 10\%$ of the students) for any experiment are shown in Tables 6 and 7.

Table 6 shows that the improvements asked for with the Pre-TLL experiments were quite varied. The themes ranged from better guidance (student examples include '*More guidance*' and '*The instructions in the lab manual were vague*'), better use of time (student examples include '*Ran very close to time, organisation could be better*' and '*Could maybe find something to make it last*

Table 6 The top three improvements raised by $\geq$10% of the students to any Pre-TLL experiment

Experiment	Top three improvements raised by $\geq$10% of respondents	Percentage of respondents (%)
Rearrangement	Greater guidance	42
	No changes required	21
	Better time management	19
EAS	No changes required	43
	Procedural issues	14
	Greater guidance	14
Anthracene oxidation	Less waiting time	48
	Greater guidance	35
	No changes required	12
Macrocycles	No changes required	31
	Greater guidance	26
	More explanation of theory/ better time management	14
Isomerisation	Greater guidance	55
	Better teaching associates/ greater context required	15
	Procedural issues	10
Proteins (2016)	No changes required	48
	Greater guidance	25
	Better time management	15

Table 7 The top three improvements raised by ≥10% of the students to any Post-TLL experiment

Experiment	Top three improvements	Percentage of respondents (%)
Proteins (2017)	Greater guidance	47
	No changes required	28
	Procedural issues	17
Sunscreen	Greater guidance	40
	Better time management	33
	No changes required	13
Pseudenol	Greater guidance	64
	Better time management	20
	No changes required	11
Electronic waste	Greater guidance	50
	No changes required	15
	Introduce group or team work	11
Nylon	Greater guidance	62
	Better time management	22
	No changes required	15
Food project	Greater guidance	52
	Pre-assignment to groups	21
	—	
Enzymes	No changes required	29
	Procedural issues	21
	Greater guidance	18

closer for the 4 hours, as opposed to finishing at 4:30 [a 90 minute early leaving time]'), better Teaching Associates (student examples include '*Lab demonstrators gave barely any info as to the theory, explaining kinetics (4102) or any of my data*' and '*Having better TA*'), fixed procedural issues (student examples include '*Flask was not sufficient to capture all solid after recrystallisation*' and '*Filtration, most solid fell through into the flask*') or even calls for no changes at all (student examples include '*Not that I can think of*' and '*Not really, exercise 3 is pretty well organised and went smoothly*'). The only universal issue noted was that students routinely called for a greater amount of guidance in every Pre-TLL experiment.

The Post-TLL results (Table 7) show a similar range of themes to the Pre-TLL results, albeit with a much different focus. The desire for greater guidance was now the main issue raised in six of the seven Post-TLL experiments. This shift is in good agreement with the quantitative data that highlighted that the students no longer felt that the laboratory manual provided sufficient guidance to complete the experiment. The strong desire for guidance is also a logical extension of the Pre-TLL results, as guidance was already a perceived issue for the students and the TLL programme deliberately sought to remove the recipe-like approach. This backlash is a likely result of students already being accustomed to expository experiments and the stepwise instructions normally provided. Hence, this result is considered positive as students will need to learn to deal with limited guidance throughout their future careers. This is the first step in acclimatising students to the uncertainties of a real workplace.

Calls for better time management in three of the cases (20–33%) were increased compared to pre-TLL (14–19%). This is most likely due to the longer, and more challenging, experiments generated through the TLL programme. Additionally, the call for group/team work in the electronic waste experiment was simply due to the requirement for students to work individually in a course where they typically worked in groups. Finally, the desire for pre-assignment to groups was a response to the fact that students were given a topic to investigate, rather than being allowed to choose from a list. Overall, many of the issues raised appeared to be responses from students speaking out against the new, more challenging, less prescriptive programme. This situation could possibly be ameliorated through conversations with students about the aims of the programme. That being said, these teething problems are common in cases where inquiry or problem-based learning has been implemented (Bruck and Towns, 2009; Gormally *et al.*, 2009) and may subside over time.

Perceived enjoyment

The third open question asked 'Did you enjoy today's lab? Why/why not?'. As the question was composed of two sections, the analysis of the responses was also split into two. The first reading of the responses was simply whether the student enjoyed the laboratory or not. These results are shown in Tables 8 and 9.

In both sets the average percentage of students stating that they enjoyed the experience was quite high (≥80%) and the average values were the same within one standard deviation. Hence, the new laboratory experiments had no measurable impact (at least by the survey utilised) on the reported enjoyment by the students. This would appear to be in contrast to the results of many others (Gormally *et al.*, 2009) who noted that students were 'resistant' to such changes in the curriculum.

Table 8 The percentage of students indicating that they enjoyed the Pre-TLL experiments

Experiment	Respondents who enjoyed the laboratory (%)
Rearrangement	95
EAS	87
Anthracene oxidation	57
Macrocycles	70
Isomerisation	74
Proteins (2016)	84
Average	81
Standard deviation	14

Table 9 The percentage of students indicating that they enjoyed the Post-TLL experiments

Experiment	Respondents who enjoyed the laboratory (%)
Proteins (2017)	82
Sunscreen	86
Pseudenol	78
Electronic waste	92
Nylon	78
Food project	70
Enzymes	78
Average	80
Standard deviation	7

Potentially, this could imply that the new experiments were better received than originally anticipated. However, another reading of the data is that the students enjoyed the old expository laboratory experiments just as much as the new ones. This implies that enjoyment may not be the best measure by which to judge the effectiveness of any particular teaching intervention.

The reasons behind their enjoyment were very informative. The top three reasons (raised by ≥10% of the students) for enjoying any experiment are shown in Tables 10 and 11.

Throughout the Pre-TLL experiments (Table 10), the most common reason raised (20–58%) for enjoying an experiment was that they were either interesting, worthwhile or fun (student examples include '*Yes, it was interesting*' and '*It was fun and pretty*'). Typically, students did not state why the experiment was any of these particular descriptions. The importance of a good Teaching Associate was another major theme (student examples include '*TA is nice and helpful. She makes the practical go very smoothly*' and '*engaging demos and interesting end results*'), appearing in the responses to four of the six experiments (17–32%). Whilst it is pleasing to hear that those particular Teaching Associates were well received, it is concerning that the students associated a large amount of the success of the experience to a small number of staff. This dependence on the Teaching Associate is a significant area of research (*e.g.* note the work of Velasco *et al.* (2016)), particularly in their training and development (Flaherty *et al.*, 2017), and not overly surprising to see come through in this case.

Outside of these main themes, students reported enjoying easy (student examples include '*It was simple and instructions are clear*' and '*b/c it was easier than previous labs*') or short experiments (student examples include '*it was only 2 hours*' and '*it was quick*'). This result is in good agreement with the work of DeKorver and Towns (2015), which showed that students often focus on simply completing the experiment as quickly as possible in order to achieve the highest mark possible.

Table 10 The top three reasons for enjoying the laboratory by ≥10% of the students to any Pre-TLL experiment

Experiment	Top three reasons for enjoying the laboratory	Percentage of respondents (%)
Rearrangement	Good teaching associate	32
	Good practical skills	22
	Interesting, worthwhile or fun	20
EAS	Short experiment	53
	Easy experiment	30
	Good teaching associate	20
	Interesting, worthwhile or fun	20
Anthracene oxidation	Interesting, worthwhile or fun	35
	Easy experiment	15
	Interesting theory	13
Macrocycles	Interesting, worthwhile or fun	51
	Good time management	14
	Good practical skills	11
Isomerisation	Interesting, worthwhile or fun	58
	Good time management	17
	Good teaching associate	17
	—	—
Proteins (2016)	Interesting, worthwhile or fun	32
	Good guidance	23
	Good teaching associate	17

Table 11 The top three reasons for enjoying the laboratory by ≥10% of the students to any Post-TLL experiment

Experiment	Top three reasons for enjoying the laboratory	Percentage of respondents (%)
Proteins (2017)	Strong context	36
	Interesting, worthwhile or fun	31
	Good time management	14
Sunscreen	Interesting, worthwhile or fun	26
	Strong context	23
	Good challenge	21
Pseudenol	Strong context	31
	Interesting, worthwhile or fun	28
	Good challenge	15
Electronic waste	Interesting, worthwhile or fun	48
	Easy experiment	15
	Strong context	10
Nylon	Interesting, worthwhile or fun	51
	Strong context	34
	Chance to development method or undertake investigation	26
Food project	Chance to development method or undertake investigation	24
	Good teamwork or team	23
	Strong context	14
Enzymes	Chance to develop method or undertake investigation	39
	Strong context	10
	—	—

Interesting practical skills (student examples include '*Learning/practicing interesting techniques*' and '*The techniques were consistent and satisfying*') and significant guidance (student examples include '*because clear instructions were given*' and '*very clear instructions*') were also raised, but only in response to single experiments. Overall, no mention was made of context or inquiry, which is to be expected, both from the nature of the experiments and the quantitative data discussed earlier.

The Post-TLL responses shown in Table 11 indicate a very different set of responses to the Pre-TLL responses. Firstly, whilst the students still routinely raised that the experiments were either interesting, worthwhile or fun, they were also far more likely to raise the context of the experiment as a reason for this (student examples include '*the context was interesting*' and '*interesting as an investigative exercise similar to industry processes*'). In fact, enjoyment due to the context of the experiment was a notable theme in all seven Post-TLL experiments (10–36%). This effect, *i.e.* the raising of context as a reason for enjoyment, is common in other implementations of context-based learning (Pringle and Henderleiter, 1999).

It is also interesting that whilst the Sunscreen and Pseudenol experiments were known to be difficult, a number of students (15–23%) raised the challenge as a reason for their enjoyment of the lab (student examples include '*Yes, quite challenging*' and '*Was a good thinking and practical challenge in chem and science principles*').

Themes relating to the ease of the experiment were only noted in the first year experiment, electronic waste, which is reasonable considering the year level involved and the simple practical skills utilised.

For the experiments that included a significant component of inquiry or experimental design (nylon, food project and enzymes), this was directly stated by the students (24–39%) as a reason they enjoyed the experience (student examples include '*make some polymers and test the properties within own design*' and '*It was interesting to design our experiments*'). Only one theme was unique to an experiment, which was good teamwork or tea' in the Food Project experiment (student examples include '*team was good*' and '*I really enjoyed the team dynamic*'). Overall, this increase in enjoyment as a direct result of increased inquiry or problem based learning is a well-known artefact of these types of teaching laboratories that was raised earlier in this article (Weaver *et al.*, 2008).

It is worth noting that whilst the post TLL experiments were never considered short and only rarely easy, there appeared to be no notable negative impact on student enjoyment. Overall, these results are promising for the TLL programme. The students appear to enjoy the experiments for the same reasons that they were created – to incorporate more industry contexualised, inquiry/problem based experiments. This is in spite of the reported issues with guidance as shown in the quantitative data and the desired improvements raised by the students.

Conclusions

Overall, the individual experiments generated from the Transforming Laboratory Learning (TLL) programme at Monash match the expected literature outcomes of context and inquiry-based learning. Through the use of a single survey (consisting of 27 closed questions and 3 open), the new undergraduate experiments result in students who are clearly more aware of (and more able to articulate) a larger range of skills that they have developed. Whilst the students recognised that the experiences were more challenging and contained far less guidance, this did not appear to impact their level of interest, enjoyment or overall appreciation of the experiments.

Furthermore, students routinely recognised that the experiments were more contextualised and more open (*i.e.* they were more able to make decisions). A large amount of effort was undertaken to incorporate more student control and greater real world context so this is a welcome result.

The students were also more likely to state that they repeated experiments, indicating an increase in a simple scientific practice – that of reproducibility. The students regularly reported (in both the closed and open questions) that the laboratory manual no longer provided enough information on its own to guide them through the experiment. As a central aim of TLL was to remove excessive guidance and encourage scientific practices (*e.g.* repetition), these were seen as favourable outcomes. However, it worth noting that there is always room for improvement with regards to student guidance and it is likely that the clarity of the student instructions could be further improved.

Students were more likely in the Post-TLL experiments to raise the development of transferable skills and skills associated with experimental design. This was accompanied by a decreased focus on development of theoretical understanding.

The proportion of students stating that they enjoyed the experiment did not change after the TLL programme. However, in the new experiments, students raised the strong context and open design of the experiments as reasons for their enjoyment – themes that were absent from the Pre-TLL data.

Overall, this research shows that the advantages gained by both contextualisation and inquiry/problem based learning persist when incorporated into a large, complex, multi-year undergraduate program. Whether or not these changes will have persistent, long term effects on the students understanding and articulation of their transferable skills will be determined through future research.

Implications for practice

There are two main outcomes of this research that could potentially guide staff involved in delivering teaching laboratories. The first is that one may not need to completely overhaul all undergraduate laboratories to obtain the benefits of inquiry/problem-based learning and context-based learning. Indeed, a range of laboratories can be generated that adhere to either increased context or enhanced inquiry (individually or together) and their global benefit may still prove fruitful. Furthermore, these changes can be implemented in many different chemistry courses and still provide the same apparent benefit. The second major practical outcome is the generation of the modified survey itself. Whilst further measurements of validity and reliability may be required, the use of this tool would appear to provide a powerful measure of the students' perceptions of any new experiments that may be generated. Furthermore, the data in this article may provide a useful comparison for future users of this particular instrument.

Future work

This work primarily focuses on the students' changing responses to a range of individual experiments generated through the TLL programme. More global investigations of the TLL programme are also being undertaken. These include, but are not limited to, focus groups of individual chemistry courses before and after the TLL programme and focus groups of students undertaking their final year project. Annual surveys (including the MLLI survey) are also being undertaken which are tracking the students' perceptions of (a) laboratory aims as well as their expectations of their actions and feelings throughout teaching laboratories and (b) their perceived level of employability and overall skill development over the three-year chemistry programme. It is also worth noting that this survey will continue to be used throughout the remainder of the TLL program. The results will be used to further guide the researchers, forming the basis of an action research approach.

Conflicts of interest

There are no conflicts to declare.

Appendix

The Wilcoxon results after correcting for sample size are shown below (Table 12).

Table 12 The results of the Wilcoxon Signed Ranked test on the 27 closed questions (after correcting for sample size) showing the number of responses (*N*), the *p* result, the calculated *Z* value and the *r* effect size

Question	*N*	*p*	*Z*	*r*
This lab experience was worthwhile.	652	1.00	0.00	—
This lab experience was interesting.	665	0.95	−0.06	—
This lab experience helped me better understand chemistry, in general, as a result of completing the chemistry lab.	660	0.70	−0.39	—
In my life, I will *not* use skills I've learned in this chemistry lab.	662	0.62	−0.49	—
This lab experience did *not* make me learn.	666	0.57	−0.57	—
Having the opportunity to use chemistry instruments helped me learn course topics.	663	0.49	−0.68	—
Even if I don't end up working in a science related job, the laboratory experience will still benefit me.	500	0.39	−0.86	—
This lab experience has made me more interested in a science career.	662	0.32	−0.99	—
This lab experience made me realize I could do science research in a real science laboratory (for instance at a university, or with a pharmaceutical company).	663	0.32	−0.99	—
Having the opportunity to use chemistry instruments made this course *less* interesting for me.	664	0.31	−1.01	—
This lab experience has made me less interested in science.	659	0.30	−1.03	—
This lab experience has made me more interested in chemistry.	661	0.078	−1.76	—
This lab experience helped me understand how the topics that are covered in chemistry lecture are connected to real research.	655	0.068	−1.83	—
This lab experience gave me a better understanding of the process of scientific research as a result of this experiment.	667	0.064	−1.85	—
Finding answers to real research questions motivated me to do well in the chemistry lab.	500	0.055	−1.92	—
Finding answers to real world questions motivated me to do well in the chemistry lab.	667	0.034	−2.12	−0.12
This lab experience presented real science to students, similar to what scientists do in real research labs.	665	0.018	−2.37	−0.17
This lab experience was *not* very similar to real research.	498	0.016	−2.42	−0.23
This lab experience was *not* well organized.	665	0.008	−2.67	−0.10
This lab experience was open enough to allow me to make decisions.	661	< 0.0005	−4.45	−0.34
This lab experience was easy.	665	< 0.0005	−4.83	−0.36
In this lab, the instructional materials did *not* provide me with explicit instructions about my experiment.	662	< 0.0005	−5.00	−0.21
This lab experience was well contextualised to real life or the workforce.	502	< 0.0005	−5.05	−0.29
This lab experience was challenging.	415	< 0.0005	−5.05	−0.30
In this lab, I did *not* repeat experiments to check results.	499	< 0.0005	−5.08	−0.29
In this lab, the instructional materials provided me with sufficient guidance for me to carry out the experiments.	664	< 0.0005	−5.33	−0.36
In this lab, I can be successful by simply following the procedures in the lab manual.	666	< 0.0005	−5.75	−0.29

Acknowledgements

The authors would like to acknowledge and thank, first and foremost, the students who participated in this study, whose honest feedback will result in a far stronger learning experience for students at Monash University. The authors would also like to express their gratitude to the technical staff at Monash University for their constant support and advice. The authors acknowledge Monash University for funding, and hosting, the Transforming Laboratory Learning programme. Ethics approval was obtained from the Monash University Human Ethics Research Committee, application number 2016000584.

References

Barlett J. E., Kotrlik J. W. and Higgins C. C. (2001), Organizational research: determining appropriate sample size in survey research, *Inf. Technol., Learn., Perform. J.*, **19**(1), 43.

Bauer C. F., (2005), Beyond ''Student Attitudes'': Chemistry Self-Concept Inventory for Assessment of the Affective Component of Student Learning, *J. Chem. Educ.*, **82**(12), 1864, DOI: 10.1021/ed082p1864.

Bauer C. F., (2008), Attitude toward Chemistry: A Semantic Differential Instrument for Assessing Curriculum Impacts, *J. Chem. Educ.*, **85**(10), 1440, DOI: 10.1021/ed085p1440.

Bingham G. A., Southee D. J. and Page T., (2015), Meeting the expectation of industry: an integrated approach for the teaching of mechanics and electronics to design students, Eur. J. Eng. Educ., 40(4), 410–431, DOI: 10.1080/03043797.2014.1001813.

Bruck L. B. and Towns M. H., (2009), Preparing Students To Benefit from Inquiry-Based Activities in the Chemistry Laboratory: Guidelines and Suggestions, *J. Chem. Educ.*, **86**(7), 820, DOI: 10.1021/ed086p820.

Chuck J.-A., (2011), Hypothetical biotechnology companies: a role-playing student centered activity for undergraduate science students, *Biochem. Mol. Biol. Educ.*, **39**(2), 173–179.

Cooper M. M. and Sandi-Urena S., (2009), Design and Validation of an Instrument To Assess Metacognitive Skillfulness in Chemistry Problem Solving, *J. Chem. Educ.*, **86**(2), 240, DOI: 10.1021/ed086p240.

Cooper L., Orrell J. and Bowden M., (2010), *Work integrated learning: a guide to effective practice*, Routledge.

Creswell J., (2009), *Research design. Qualitative, Quantitative, and Mixed Methods Approaches*, Thousand Oaks, CA: SAGE Publications.

Cummins R. H., Green W. J. and Elliott C., (2004), ''Prompted'' Inquiry-Based Learning in the Introductory Chemistry Laboratory, *J. Chem. Educ.*, **81**(2), 239, DOI: 10.1021/ed081p239.

DeKorver B. K. and Towns M. H., (2015), General Chemistry Students' Goals for Chemistry Laboratory Coursework, *J. Chem. Educ.*, **92**(12), 2031–2037, DOI: 10.1021/acs.jchemed.5b00463.

Domin D. S., (1999), A Review of Laboratory Instruction Styles, *J. Chem. Educ.*, **76**(4), 543, DOI: 10.1021/ed076p543.

Dopico E., Linde A. R. and Garcia-Vazquez E., (2014), Learning gains in lab practices: teach science doing science, *J. Biol. Educ.*, **48**(1), 46–52, DOI: 10.1080/00219266.2013.801874.

Duch B. J., Groh S. E. and Allen D. E., (2001), Why problem-based learning, *The Power of Problem-Based Learning*, 3–11.

Erhart S. E., McCarrick R. M., Lorigan G. A. and Yezierski E. J., (2016), Citrus Quality Control: An NMR/MRI Problem-Based Experiment, *J. Chem. Educ.*, **93**(2), 335–339, DOI: 10.1021/acs.jchemed.5b00251.

Flaherty A., O'Dwyer A., Mannix-McNamara P. and Leahy J. J., (2017), The influence of psychological empowerment on the enhancement of chemistry laboratory demonstrators' perceived teaching self-image and behaviours as graduate teaching assistants, *Chem. Educ. Res. Pract.*, **18**(4), 710–736, DOI: 10.1039/C7RP00051K.

Fritz C. O., Morris P. E. and Richler J. J., (2012), Effect size estimates: Current use, calculations, and interpretation, *J. Exp. Psychol.: Gen.*, **141**(1), 2–18, DOI: 10.1037/a0024338.

Galloway K. R. and Bretz S. L., (2015), Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory, *J. Chem. Educ.*, **92**(7), 1149–1158, DOI: 10.1021/ed500881y.

Gormally C., Brickman P., Hallar B. and Armstrong N., (2009), Effects of inquiry-based learning on students' science literacy skills and confidence, *Int. J. Scholar. Teach. Learn.*, **3**(2), 16, DOI: 10.20429/ijsotl.2009.030216.

Grove N. and Bretz S. L., (2007), CHEMX: An Instrument To Assess Students' Cognitive Expectations for Learning Chemistry, *J. Chem. Educ.*, **84**(9), 1524, DOI: 10.1021/ed084p1524.

Hanson S. and Overton T., (2010), *Skills required by new chemistry graduates and their development in degree programmes*,Hull, UK: Higher Education Academy UK Physical Sciences Centre.

Hattie J., (2008), *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*, Routledge.

Henry G. T., (1990), *Practical sampling*, Sage, **vol. 21**.

Jansson S., Söderström H., Andersson P. L. and Nording M. L., (2015), Implementation of Problem-Based Learning in Environmental Chemistry, *J. Chem. Educ.*, **92**(12), 2080–2086, DOI: 10.1021/ed500970y.

Johnstone A. and Al-Shuaili A., (2001), Learning in the laboratory; some thoughts from the literature, *Univ. Chem. Educ.*, **5**(2), 42–51.

Kelly O. C. and Finlayson O. E., (2007), Providing solutions through problem-based learning for the undergraduate 1st year chemistry laboratory, *Chem. Educ. Res. Pract.*, **8**(3), 347–361, DOI: 10.1039/B7RP90009K.

Lagowski J. J., (1990), *Entry-level science courses: the weak link*, ACS Publications.

Leal Filho W. and Pace P., (2016), *Teaching Education for Sustainable Development at University Level*, Springer.

Lenhard W. and Lenhard A., (2016), *Calculation of Effect Sizes*, from https://www.psychometrica.de/effect_size.html.

Mc Ilrath S. P., Robertson N. J. and Kuchta R. J., (2012), Bustin' Bunnies: An Adaptable Inquiry-Based Approach Introducing Molecular Weight and Polymer Properties, *J. Chem. Educ.*, **89**(7), 928–932, DOI: 10.1021/ed2004615.

McDonnell C., O'Connor C. and Seery M. K., (2007), Developing practical chemistry skills by means of student-driven problem based learning mini-projects, *Chem. Educ. Res. Pract.*, **8**(2), 130–139, DOI: 10.1039/B6RP90026G.

Norton A. and Cakitaki B., (2016), *Mapping Australian higher education 2016*, Grattan Institute, p. 7.

Pilcher L. A., Riley D. L., Mathabathe K. C. and Potgieter M., (2015), An inquiry-based practical curriculum for organic chemistry as preparation for industry and postgraduate research, *S. Afr. J. Chem.*, **68**, 236–244.

Pringle D. L. and Henderleiter J., (1999), Effects of Context-Based Laboratory Experiments on Attitudes of Analytical Chemistry Students. *J. Chem. Educ.*, **76**(1), 100, DOI: 10.1021/ed076p100.

Ram P., (1999), Problem-Based Learning in Undergraduate Instruction. A Sophomore Chemistry Laboratory, *J. Chem. Educ.*, **76**(8), 1122, DOI: 10.1021/ed076p1122.

Russel C. B., (2008), *Development and Evaluation of a Research-Based Undergraduate Laboratory Curriculum*, PhD, West Lafayette, Indiana: Purdue University.

Sandi-Urena S., Cooper M. and Stevens R., (2012), Effect of Cooperative Problem-Based Lab Instruction on Metacognition and Problem-Solving Skills, *J. Chem. Educ.*, **89**(6), 700–706, DOI: 10.1021/ed1011844.

Santos J. R. A., (1999), Cronbach's alpha: a tool for assessing the reliability of scales, *J. Extension*, **37**(2), 1–5.

Sarkar M., Overton T., Thompson C. and Rayner G., (2016), Graduate Employability: Views of Recent Science Graduates and Employers, *Int. J. Innov. Sci. Math. Educ.*, **24**(3), 31–48.

Velasco J. B., Knedeisen A., Xue D., Vickrey T. L., Abebe M. and Stains M., (2016), Characterizing Instructional Practices in the Laboratory: The Laboratory Observation Protocol for Undergraduate STEM, *J. Chem. Educ.*, **93**(7), 1191–1203, DOI: 10.1021/acs.jchemed.6b00062.

Watson D., (1992), Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale: An Application to Class Consciousness, *Sociol. Method. Res.*, **21**(1), 52–88, DOI: 10.1177/0049124192021001003.

Weaver G. C., Russell C. B. and Wink D. J., (2008), Inquiry-based and research-based laboratory pedagogies in undergraduate science, *Nat. Chem. Biol.*, **4**, 577, DOI: 10.1038/nchembio1008-577.

Wiggins G., (1990), *The Case for Authentic Assessment*, ERIC Digest.

Woolfolk A., (2005), *Educational Psychology*, Boston, MA: Allyn and Bacon.

5.1 Summary of findings

The results indicated that students found the new inquiry/problem/context/industry-based laboratory activities, in comparison to unaltered experiments:

- More challenging, less easy, more contextualised to the work-force, enabled more student decision making, and were less dependent on the laboratory manual.
- Resulted in a greater range of perceived skill development, which now included scientific methodology and a greater focus on transferable skills (such as teamwork or time management).
- Were just as enjoyable but for a greater variety of reasons (e.g. context or inquiry).
- Challenged the students' dependence on guidance from the laboratory manual.

Overall, this work appeared to indicate that the individual laboratories were meeting the aims of the large-scale inclusion of inquiry/problem/context/industry-based experiments i.e. students were articulating a wider range of employability skills, particularly those around inquiry and experimental design. Furthermore, these results imply that the benefits of contextualisation and inquiry/problem-based activities are neither unique to either a given year level nor a specific sub-discipline (e.g. organic chemistry). However, it is difficult to ascertain from this data whether students will continue to recognise these benefits (e.g. be able to recall or articulate their developed skills) as they proceed through their undergraduate careers. Hence, longitudinal studies on the students' perceptions of their undergraduate laboratory experiences were undertaken and will be discussed in Chapter 6.

It is also worth noting that this study generated a survey that could be utilised by other researchers or institutions wishing to monitor the success of their own interventions. This is particularly useful as this survey is highly flexible and is easily used in other science and science-related disciplines (e.g. it is easy to simply change 'chemistry' to 'biology' throughout the survey).

Chapter 6 - Overall impacts of the large-scale inclusion of inquiry/problem/context/industry-based experiments.

It was of interest to monitor the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on a larger or longer scale (i.e. unit/course or year level as opposed to each individual activity) with regards to the following research questions:

- What is the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on:
  - The students' perception of laboratory aims and their expectations of their thoughts, actions and feelings during laboratories? (Page 30, Research Question 1)
  - The students' level of enjoyment of the laboratory exercises? (Page 30, Research Question 2)
  - The development of students' employability skills and their recognition of these skills? (Page 30, Research Question 3)

To consider whether the findings of higher levels of enjoyment and a broader range of articulated skills developed because of the inquiry/problem/context/industry-based experiments activities (as noted in Chapter 5) persists over the larger unit/course level scale of 2-3 months, focus groups were conducted at the end of units/courses before and after the TLL programme had been undertaken. The units/courses chosen were those where more than half of the laboratories were significantly altered or replaced by the TLL programme.

To consider the findings on a year level time scale, prior results in Chapter 3 were considered. The results of those surveys provided a baseline to monitor the change in student perceptions of the aims of laboratory activities and their expectations because of the large-scale inclusion of inquiry/problem/context/industry-based experiments. The original survey was distributed annually to students to track any changes in student perceptions of the aims of laboratory activities and their expectations of how they will think, feel and act during any given laboratory activity. These surveys were disseminated to 2nd and 3rd year students in 2017 and 2018 to track the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' perceptions of the 2nd and 3rd year laboratory activities, respectively.

The data from these surveys and focus groups have been analysed and submitted to the *International Journal of Science Education* and reproduced on the following pages.

Inquiry-, problem-, context- and industry-based laboratories: An investigation into the impact of large-scale, longitudinal redevelopment on student perceptions of teaching laboratories.

Previous work in the School of Chemistry at Monash University has shown that students recognise that inquiry/problem/context/industry- based experiments were better contextualised, more open to decision making and aided in the development of scientific and transferable skills. However, the results were collected immediately after the experiments were carried out. This study investigated whether these gains persisted over a longer time scale alongside the impact of the persistence of a large number of unaltered traditional experiments. Student focus groups were undertaken at the completion of units/courses in which more than half of the laboratory experiments were redesigned to investigate their impact over a semester. Annual surveys were distributed to monitor students' perceptions of the aims of teaching laboratories, and their expectations of their own behaviour. The findings indicated that the positive outcomes of the new experiments were still evident at the end of a semester. The annual survey showed that whilst 2nd year students were more able to appreciate the connection between the experiments and the real world, 3rd year students had apparently lost this appreciation. It is believed that this is likely due to the larger number of expository experiments that remained in the 3rd year. Overall, the large-scale changes away from expository experiments had a positive impact on student enjoyment and perceived skill development but only when meaningful proportions of the experiments were redesigned. It would appear that the continuing existence of many expository experiments undermined the students' perceptions of the benefits gained by the new laboratory experiments.

Keywords: undergraduate, chemistry, quantiative, qualitative, practicals

Introduction

The Transforming Laboratory Learning programme (TLL) ran from early 2016 to late 2018 and aimed to modernise the teaching laboratory experiments at Monash University, a large Australian institution with over 2,000 students studying chemistry. The alteration of the laboratory experiments was designed to meet the current needs of the students and their future employers, whether they be in research, industry or sectors unrelated to chemistry. It has been noted that students are often found to lack the transferable skills that modern employers desire (Sarkar, Overton, Thompson, & Rayner, 2016) such as time management, team-working, independence, self-directed learning, problem solving or critical thinking skills and many others.

In many higher education institutions, laboratory experiments consist of highly prescriptive experiments, i.e. expository or recipe-based experiments, in which students simply follow a procedure rather than devising or testing their own methods or hypotheses. These experiences are often criticised for requiring little to no critical thought (Hodson, 1990), with students simply '*following a prescribed procedure to experience a pre-determined outcome'* (Domin, 2007). It has been noted that students often struggle to recognise the aims of these experiments, even immediately after their completion (Kirschner & Meester, 1988).

The consistent large-scale use of expository experiments was seen to be the case at Monash University where many laboratory experiences were not only expository in style but often lacked a context beyond simply confirming or testing theories covered in a lecture course. Hence, the TLL programme previously described (George-Williams, Soo, Ziebell, Thompson, & Overton, 2018) was devised in order to combat these perceived issues. The TLL programme intended to implement modern teaching and learning practices, alongside on-going communication with industry partners, to enhance the laboratory experiments in order to:

(1) Increase the workforce or real-world context of the laboratory experiments. This was in an attempt to increase student engagement through context-based learning (Pilot & Bulte, 2006) which has been shown to lead to more meaningful learning when the context of the task is perceived as relevant

and/or related to the student's daily lives (Avargil, Herscovitz, & Dori, 2013). The contexts were often found through conversations with industrial partners which led to the inclusion of modern techniques and challenges that would better prepare students for employment.

(2) Increase the number of opportunities for students to develop and test their own methods and hypotheses. This aim led to increased use of inquiry-based learning (Bruner, 1961). When students are given opportunities to ask their own questions, and then supported to seek the answers, they tend to be more engaged and develop higher order skills such as critical thinking or problem solving (Lazonder & Harmsen, 2016).

The use of context- and inquiry-based learning are well known in the literature. There are examples of undergraduate chemistry laboratory experiments that utilise either modern industry/real world contexts (Erhart, McCarrick, Lorigan, & Yezierski, 2015) or student-centred inquiry (Bernard, Britz-McKibbin, & Gernigon, 2007; Kulevich, Herrick, & Mills, 2014; Mills & Guilmette, 2007). However, these examples are often isolated experiences and are rarely used on a large scale (i.e. in more than one unit/course or across year levels), although there are examples of these approaches being used across an entire unit (Gormally, Brickman, Hallar, & Armstrong, 2009; Pilcher, Riley, Mathabathe, & Potgieter, 2015).

The impact of laboratory interventions over a longer time-scale has been investigated by Szteinberg and Weaver (2013). In this case, the effects of a single redesigned semester long unit/course was investigated two and three years after students had completed it. The new unit/course incorporated research/inquiry experiences and required students to come up with and test their own ideas. It was found that students were more likely to recall these laboratories compared to students who completed the traditional course and were more likely to feel prepared in later research experiences. This work indicates the long lasting potential effects of the inclusion of research/inquiry-based experiments on students' preparedness for more realistic scientific environments, but is only focused on a single year level and semester. Hence, further work on more extensive changes (i.e. multiple year levels and units/courses) may be of importance to the education research community.

The TLL programme aimed to transform laboratory experiments across years 1, 2 and 3. The challenge of this research was to evaluate the effect of this large-scale programme as it redesigned or replaced laboratory experiments over different units/courses and year levels (Figure 1) The overarching research question was whether the inclusion of a large number of inquiry/problem/context/industry-based experiments would alter the students' perceptions of laboratory experiments? The research questions that guided this work were - What was the impact of the large scale inclusion of inquiry/problem/context/industry- based experiments on:

(1) students' perceptions of why they enjoyed (or didn't enjoy) specific teaching laboratory experiments?
(2) students' perceptions of their own skill development?
(3) students' perceptions of the aims of laboratory experiments?
(4) students' perceptions of how they will act, feel and think during a given teaching laboratory?

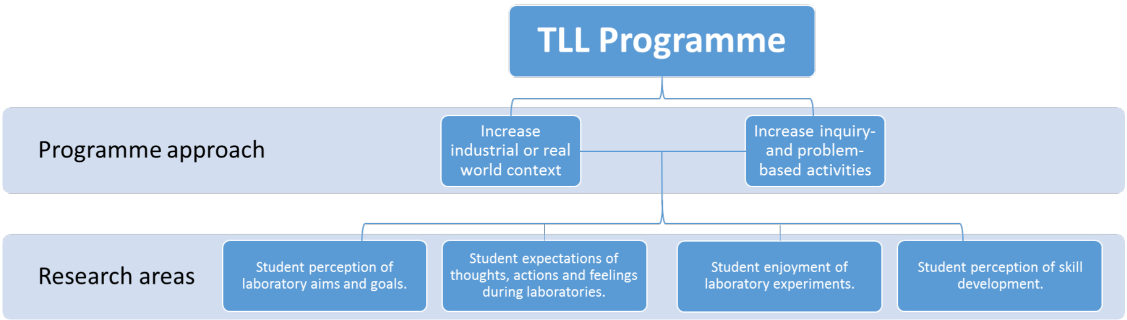


Figure 1 - An overview of the TLL programme aims and research areas.

The effect of the TLL programme on student enjoyment and perceptions of their developed skills directly after a new experiment has been previously published (George-Williams, Soo, et al., 2018). This was carried out using a modified version of a published survey (Russel, 2008). It was noted that students found the new experiments to be more difficult, more open to their own decision making, better contextualised and just as enjoyable as previous laboratory experiences, albeit for a greater variety of reasons (such as the context or the opportunity for inquiry). Students were also able to articulate a much wider range of developed skills, particularly after inquiry-based experiments. Whilst these results were very encouraging, it was not possible to determine whether the effects would persist over a longer time-scale, especially as many expository experiments still existed in the overall three year laboratory programme.

To investigate the longevity of the gains noted, a series of focus groups were undertaken at the completion of units/courses where at least half of the laboratory experiments were significantly redesigned or replaced, i.e. at the end of a semester. Significant work has already been undertaken on students' perceptions of the aims of laboratory experiments. For example, previous work by DeKorver and Towns (2015) and Russell and Weaver (2008) has shown that students in chemistry laboratories tended to focus on affective goals, such as simply completing the task in a short period of time in order to obtain assessment marks. Additionally, it was found (Boud, Dunn, Kennedy, & Thorley, 1980) that when provided with a list of potential aims of laboratory experiments, students tended to rate the development of practical skills and connection to theory above the development of problem-solving skills.

Student perceptions of the aims of laboratory experiments was investigated at Monash University through analysis of student responses to the single question 'What do you think the aims of doing a practical chemistry course are?' (George-Williams, Ziebell, et al., 2018), which indicated that the students were focused on the development of either theoretical understanding or practical skills above either preparation for the workforce or the development of transferable skills. Therefore, there existed a strong baseline dataset through which to measure the effect of the TLL programme on students' the perceptions of the aims of laboratory experiments.

Student expectations of their own thoughts, actions and feelings has been considered in the literature. The Meaningful Learning in the Laboratory Instrument (Galloway & Bretz, 2015a) (MLLI) was designed to measure students' actions during a teaching laboratory through the lens of Novak's theory of meaningful learning (Novak, 1998), which states that meaningful learning only occurs when a student correctly aligns the psychomotor, affective and cognitive domains (i.e. actions, feelings and thoughts) during a learning experience. The original use of this survey, and its subsequent use on a national scale (Galloway & Bretz, 2015b), both showed that a significant mismatch existed between student expectations of their laboratory experiences and what actually occurred. Use of the survey on a large cohort of first, second and third year students at multiple institutions (George-Williams et al., 2019) revealed that these expectations appeared to be unaffected by the laboratory experiments currently provided, which tended to be relatively expository in nature. Hence, this data set provided an opportunity to measure the effect of changes implemented through the TLL programme on the students' expectations of their actions, thoughts and feeling during teaching laboratories.

Method

The aim of this study was to investigate the long-term effects of a large-scale move to context- and inquiry-based experiments on a variety of student perceptions. The theoretical framework that guided this work was constructivism which '*underlies the assumption that learning is an active process where knowledge is constructed based on personal experiences and the continual testing of hypotheses*' (Leal Filho & Pace, 2016)'. It was postulated that the inclusion of a greater amount of context would allow the students to better connect to their daily lives and that the use of inquiry-based activities would provide an opportunity for students to test and investigate their own hypotheses.

The research questions were designed to investigate whether students were able to reconcile the new experiences with their preconceptions of teaching laboratories in terms of the aims of such experiences, how they believe they will think, act and feel, and what skills they might develop. The overall changes to the laboratory program resulted in two new 1st year laboratory experiments (one in each of the two units/courses valuable), fourteen for the 2nd year (spread over three of the five units/courses available) and ten in the 3rd year (spread over four of the eight units/courses available). As there are typically around 8 weeks of laboratory experiments per unit/course, this equated to approximately 15%, 41% and 15% of the 1st, 2nd and 3rd year laboratory experiments being altered or replaced as per year level. Consequently, the proportion of context- and inquiry-based experiments that 3rd year students engaged with was much smaller than for 2nd year students. It is worth noting that the eight 3rd year units/courses compared to just five for the 2nd year contributed to the lower percentage of changes in the 3rd year units/courses. Furthermore, the amount of changes undertaken in the 3rd year units/courses were limited by available time, buy-in of academic staff and limited equipment. Lastly, the 1st year of the degree programme had been overhauled just prior to the TLL programme and was therefore relatively untouched during this time.

Associated pre-laboratory activities were also generated and many laboratory manuals were edited to be consistent in format and style. The data required to monitor the effect of these changes were collected in 2016, 2017 and 2018 using focus groups and paper-based surveys through a cross-sectional, mixed methods research design.

Data collection – Focus groups

Focus groups were used in this study to allow students an opportunity to provide in-depth collective views of the new experiments (Kitzinger, 1994, 1995). Furthermore, the use of focus groups helped to validate the findings of the paper-based survey results - a process known as triangulation (Morgan, 1996). Administration of additional surveys was not desirable at this stage in order to avoid student fatigue with respect to completing paper-based surveys.

Students were invited to participate in the focus groups either through a broadcasted message on the learning management system (Moodle) or through a direct request during class time. No selection criteria were applied and all who volunteered were accepted. Students were offered food as the sessions were generally run during lunch or dinner time. Students were informed that their responses would remain anonymous, would never be linked to them, and would not affect their academic standing in any way. Any names mentioned (of other students or teaching staff) throughout the session were redacted. The facilitators of the focus groups were either senior researchers with extensive experience of conducting focus groups or a PhD candidate trained by the same senior researchers. Whilst some students participated in multiple pre- or post- focus groups, no students participated in both pre- and post- focus groups for any given unit/course. The focus groups were audio recorded and transcribed. Eight focus groups were undertaken for four units/courses both before and after the large scale alteration of the laboratory activities (2nd year Food Chemistry, 2nd year Inorganic and Organic Chemistry, 3rd year Materials Chemistry and 3rd year Advanced Organic Chemistry). The number of students who participated in each focus group can be found in Table 1 and generally represent less than 10% of their respective cohorts.

Table 1 - The number of students who participated in the focus groups.

	Pre (*N*)	**Post (*N*)**
2nd year Inorganic and Organic Chemistry	6	8
2nd year Food Chemistry	2	5
3rd year Materials Chemistry	2	3
3rd year Advanced Organic Chemistry	7	7

The three main questions that the students were asked were:

(1) Which laboratory experiences did you enjoy? Why?

(2) Which laboratory experiences didn't you enjoy? Why?

(3) What skills do you think you developed throughout these laboratory experiences?

Students were also encouraged to discuss the teaching associates (demonstrators), the pre-laboratory quizzes and the post-laboratory assessment. However, this was undertaken more as an opportunity to gain feedback on the development of teaching materials, rather than to gather evidence for this study. Furthermore, students were invited to raise any other topic relevant to the course at the completion of the questions about the laboratory experiments.

Data collection – Surveys

The amended MLLI survey consisted of 31 closed questions and a 5-point Likert scale (Strongly Disagree, Disagree, Neutral, Agree and Strongly Agree) (see the appendix). The survey was distributed in paper format alongside the open question 'What do you think the aims of doing a practical chemistry course are?' and several demographic questions (gender identity, age and enrolment). Students were surveyed during a teaching laboratory induction session for one 2nd year unit/course and two 3rd year units/courses. These units/courses were chosen for convenience and to avoid clashing with other surveys being distributed during the same semester. Students were invited to complete the survey, but it was not compulsory. The number of students who responded can be found in Table 2.

Table 2 - The number of responses to the MLLI survey and open question.

	***N* (2017)**	***% (2017)***	***N* (2018)**	***% (2018)***
2nd Year	143	~41-47%	128	~37-43%
3rd Year	132	~52-66%	80	~32-40%

Data analysis – open answers and audio transcripts

After transcription, all open questions were analysed through inductive coding techniques i.e. the responses were read multiple times by a single author until themes began to emerge from the data. These themes were assigned a code and then used by 3-5 chemistry education researchers to assign codes to each individual transcript. If the agreement between the researchers was below 90%, the themes/codes were revised, and coding again undertaken. This process is often referred to as inter-rater reliability (Gwet, 2014) and ensures that the themes extracted are not unique to a single researcher. This process was undertaken on both the open answers to the surveys and the recorded transcripts from the focus groups.

For the responses to the open survey questions only, the total number of times a given theme was raised by all respondents was converted to a percentage of the total number of responses to allow for comparison. The frequency of a given theme being raised by different cohorts was also determined in order to investigate potential significant differences between cohorts. This was achieved through the use of the Pearson's chi squared test with a *p* value cut-off of 0.05 (or 95% confidence interval) using SPSS Version 23 (Statistical Package for the Social Sciences). Additionally, Cramer's *V* (ϕ_C) was calculated in order to assign an 'effect size' to any significant differences (i.e. how large or small any difference is). The ranges for ϕ_C were determined first through consideration of the work of Hattie (2008) (on Cohen's *d*), who lowered the original range values (Cohen, 1988) when concluding that the ranges utilised excluded many successful interventions. The new ranges assigned by Hattie were later extended to *r* (Fritz, Morris, & Richler, 2012; Lenhard & Lenhard, 2016). As the original *r* ranges (i.e. before

Hattie and the later extension work) matched those for the Cramer's V (ϕ_C), the altered r ranges were finally applied to Cramer's V (ϕ_C) here (Table 3).

Table 3 - The ranges for ϕ_C according to the work of Hattie (2008) and later extension (Fritz et al., 2012; Lenhard & Lenhard, 2016).

ϕ_C range	**Label**	**Explanation**
0-0.1	'Student' effect size	This refers to the natural variation in any group of students. For example, a more motivated student may respond more positively than a less motivated students.
0.101-0.2	'Teacher' effect size	This refers to the effect of a particularly motivated teacher over the course of a single year (i.e. this effect size could be achieved given time/motivation).
>0.201	'Zone of desired effect'	This refers to interventions that have an immediate impact and are where educators should typically focus their efforts.

Data analysis – closed answers

The responses to the closed MLLI questions were transcribed into Excel as numbers 1-5, where 1=strongly disagree, 2= disagree, 3=neutral, 4=agree and 5=strongly agree. The frequency of responses was used to compare cohorts. The average of this data was never used as this ordinal data requires non-parametric statistical analysis. To ensure different cohorts were not responding in a systematically different way to *all* questions, an omnibus test (here, an F-test) was performed. In this case, if the test showed no significant difference between how a given cohort responded to *all* questions compared to the all responses of all cohorts combined, then further testing was undertaken to investigate how that cohort responded to individual questions.

Significant differences in responses to individual questions between independent cohorts was determined through the use the Pearson's chi squared test.

The p value was decreased through a Bonferroni correction, giving a p cut-off of 0.005 (99.5% confidence interval). This was carried out to lower the chance of a Type I (false positive) error which can result when multiple tests are performed on a single cohort (i.e. a Pearson's chi squared test between two cohorts on a 30 question survey). The effect size of any changes measured were calculated using ϕ_C. However, a complication arises through the comparison of the 5-item Likert scale as this increases the degrees of freedom in the analysis (df=4 in this case). As such, the ranges were again modified from the original ranges to take this into account (Cohen, 1988), with the ranges used shown in Table 4.

Table 4 - The modified ϕ_C values to account for the larger degrees of freedom from the use of a Likert scale.

ϕ_C range	**Effect size**
0.050-0.150	Small effect
0.151-0.250	Medium effect
>0.251	Large effect

Limitations

One limitation of this work is the sample sizes. At Monash University, there are approximately 300-350 students enrolled in 2nd year units/courses and 200-250 students enrolled in 3rd year units/courses. However, cross-sections of these cohorts were surveyed (e.g. two of the four 3rd year units/courses available), which lowers the total maximum possible response rates. Thus, the cross-sections may not be truly representative of the entire cohort. Furthermore, the response rates were between 30% and 80% of the respective cohorts. However, as many students are simultaneously enrolled in both 2nd and 3rd year units/courses, the number of individual students is likely to be lower than the number of enrolments, implying the percentage response rates may in fact by higher than stated. A power analysis was undertaken, using the work of Barlett, Kotrlik, and Higgins (2001), which shows that at the 95% confidence level, these sample sizes are considered representative of the cross-section of the cohorts measured (as per the required cut-offs for categorical data).

Another limitation of this study is the change in scales in the MLLI survey from an electronic slider (0-100%) to a 5-point Likert style which was undertaken due to the paper based nature of survey delivery. As a result, it is difficult to directly compare the results of this study with previous findings. This resulted in the data being treated through a question by question analysis (as compared to grouping questions together) followed by a Pearson's chi-squared test. It was found that a factor analysis did not yield similar results to those raised by Galloway and Bretz (2015a). This is potentially the result of a disconnect between student perception of the questions and their ability to connect to their affective, psychomotor or cognitive domains. Overall, comparison between this data and that collected in the work of Galloway and Bretz will not be achievable at this time due to the different scales utilised.

The numbers of students who participated in the focus groups were small and it is, therefore, difficult to transfer their responses to other students and contexts (i.e. other institutions). In addition, these students were volunteers and may not be typical of students in their cohorts. The use of multiple measurement tools (open and closed survey questions and focus groups) and subsequent triangulation of the results through comparison to one another should negate these issues.

Another limitation is that the use of a Pearson's chi-squared test, alongside a Holm-Bonferroni correction, is very conservative, potentially resulting in an increase of Type II (false negative) errors. Also, with regards to the analysis of the ordinal Likert data, the use of an *F*-test and a Pearson's chi-squared are typically used for continuous and categorical data respectively. However, with no true test for ordinal data widely accepted in the literature, the issue of using tests designed for continuous/categorical data on ordinal data cannot be circumvented.

Another limitation to this study is that other research programs were on-going concurrently at Monash University. One such project focused on the use of skills badges or icons in the laboratory manuals in order to enhance students' awareness of the possible development of transferable skills (similar to the work of Barlett et al. (2001) and Hennah and Seery (2017)). Therefore, it is possible that the responses of students were affected by this dual messaging (i.e. from the TLL programme and the badging programme).

Finally, students were aware of the TLL program between 2016 and 2018 through regular surveys, invitations to participate in focus groups, trials of new experiments and the role of some of the authors as teaching associates or academic staff. It is possible that some of the responses of the students were influenced by this visibility. It is important to note however that when students were responding to surveys and focus groups, they were not explicitly told which laboratories had been altered by the TLL programme.

Results and Discussion

Focus groups

The student focus groups at the end of a semester were used to determine what the impact of the implementation of inquiry/problem/context/industry- based experiments were over a time scale of months as compared to immediately following a laboratory experiment as measured previously (George-Williams, Soo, et al., 2018). The first questions considered were ‘Which laboratory experiences did you enjoy? Why?’. The themes extracted from all focus groups are presented in Table 5.

Table 5. The themes extracted during the focus groups when students were asked ‘Which laboratory experiences did you enjoy? Why?’.

Code	Description	Example
CB	Context-based (specifically real life)	*‘Personally, my favourite lab was the panacetin one because it had that little bit of application to it.’’*
IB	Inquiry-based	*‘I really, really enjoyed the nylon prac ... we had to predict and we actually got to test’*
GI	General interest statement	*‘it was still kind of cool … It was just fun …’*
GW	Group work	*‘But I think working as a group … was a really good thing’*
MW	Multi-week	*‘[It] was good going over a few weeks and was really getting involved with it.’*
WF	Work force context	*‘it made me appreciate ... how a company is doing similar things to what we're doing’*
LL	Link to lectures	*‘you have an understanding of it from the lectures ... by the time you finished a report and submitted it, I felt like I'd learned something’*
OC	Overcoming challenges or obstacles	*‘Even though it was difficult, I think it was satisfying’*

The responses to these questions were compared between the focus groups undertaken at the completion of units/courses before any changes had been made, and after the experiments had been revised in the following year. The findings are presented in Figure 2.

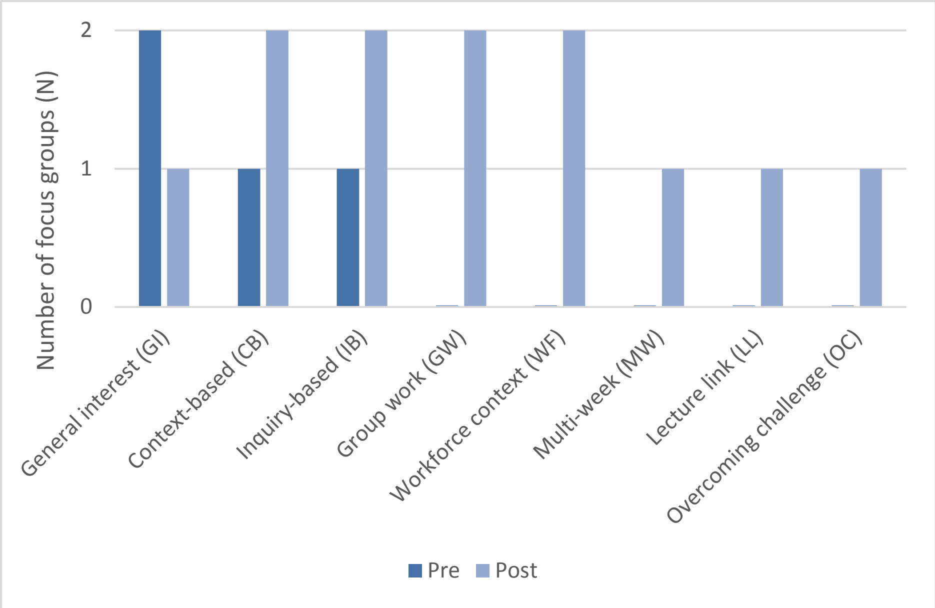


Figure 2. The number of focus groups (*N*) where a given theme was raised pre- and post- the inclusion of the new experiments when considering enjoyment.

This data indicates that the responses of students to the laboratory experiments before the intervention was mostly focused on general interest (GI) and some context/inquiry-based (CB/IB) experiments that already existed in those units/courses. After the inclusion of a large number of inquiry/problem/context/industry- based experiments, the responses of the students changed to include more context/inquiry-based (CB/IB) activities alongside an appreciation for group work (GW), the workforce context (WF) and, in one-case, the use of a multi-week (MW), challenging experiment (OC) that was well connected to the lecture material (LL).

Overall, it would appear that the students enjoyed both the inquiry/problem/context/industry- based experiments. These quotes (Table 5) also show the deep appreciation for the industry-focused experiments. These results match the data obtained for the individual laboratory experiments (George-Williams, Soo, et al., 2018) where students were more likely to state that the experiment enabled them to make their own decisions and was highly contextualised to the real world or the workforce. The longevity of these findings appear to indicate that the effects of the new approach are still prevalent even after the completion of a semester-long course. This is clearly the result of more than half of the experiments being redesigned in these units/courses but also indicates that the remaining expository experiences did not negate the benefits of the new experiments on this timescale.

The next questions considered were ‘Which laboratory experiences didn’t you enjoy? Why?’. The themes extracted from all focus groups are presented in Table 6.

Table 6. The themes extracted during the focus groups when students were asked ‘Which laboratory experiences didn’t you enjoy? Why?’.

Code	**Description**	**Example**
TE	Too easy.	*‘it felt a bit like it could have been done at first year level’*
TG	Too guided (either by TAs or the laboratory manual).	*‘Maybe instruct the demonstrators not to give the answers, just like not to spoon feed us.’*
TB	Too boring or repetitive.	*‘I didn’t like it because it was like, you weighed one thing out and then it was just like squirt, squirt in a test tube ...’*
NLL	No clear link to lecture materials.	*‘Because it seems like they were basically just giving us the lab to teach us something they couldn't fit into the lectures’*
TD	Too difficult	*‘it got a bit too overwhelming because, even though the labs were four hours, it felt like you were trying to squeeze six hours' worth of learning into four hours’*
OQ	Overly qualitative results	*‘one of my main problems with like a lot of the practice we've done so far had been haven't been particularly qualitative’*
LG	Lacking guidance (either by TAs or the laboratory manual)	*‘we actually got confused. Like not just us, the TAs as well ... the TAs didn't seem to know, or deliver how we should do it as effectively as they could have’*

Again, the changes in themes raised were considered before and after the implementation of inquiry/problem/context/industry- based experiments and the findings presented in Figure 3.

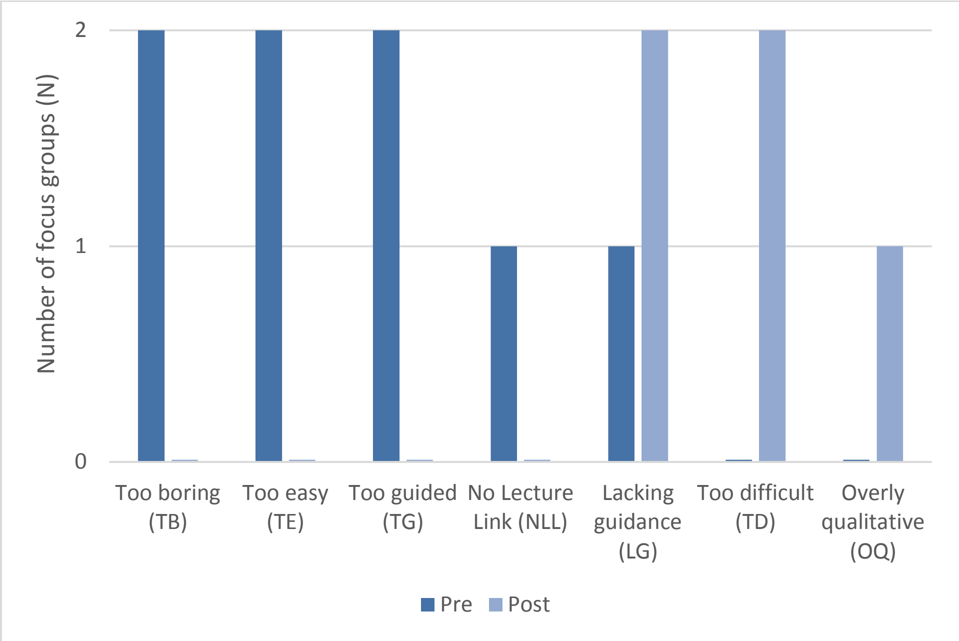


Figure 3. The number of focus groups (*N*) where a given theme was raised pre- and post-the inclusion of the new experiments when considering non-enjoyable experiments.

Figure 3 shows a clear shift from a student perception of laboratory experiments that were considered boring, easy, overly guided, or poorly linked to lectures to too difficult, lacking guidance or overly qualitative results. Again, this matches the previous study (George-Williams, Soo, et al., 2018) in which students were more likely to state that the new laboratory experiments were challenging and sometimes raised a lack of guidance as a significant issue with the experiences. Issues with teaching staff is a common finding in the literature for inquiry-based experiences (Gormally et al., 2009). Whilst it is important to ensure good quality support and guidance for students, these findings are considered as a positive outcome as the experiments now stretch the students and encourage them to overcome challenges that they are likely to encounter in the workforce.

Finally, students were asked to reflect on their skill development – 'What skills do you think you developed throughout these laboratory experiences?'. The skills raised are shown in Figure 4 including their breakdown as per pre and post the intervention.

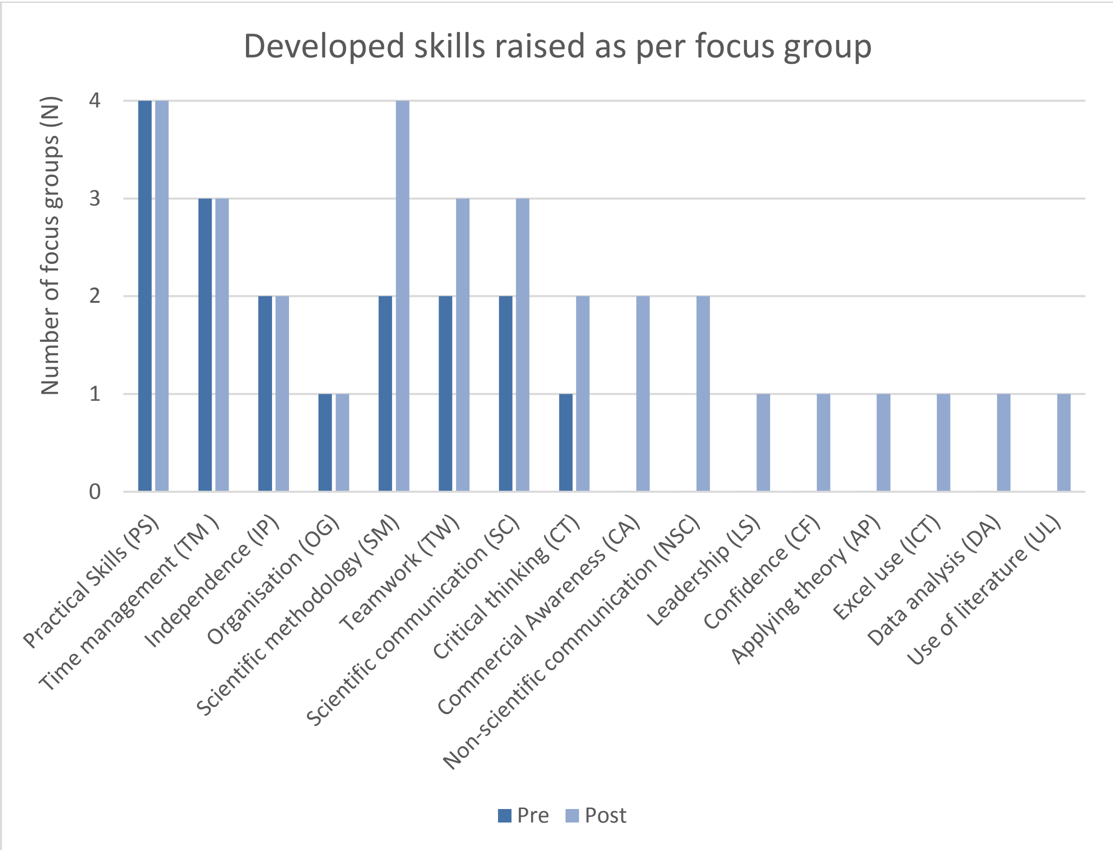


Figure 4. The number of focus groups (N) where a given skill was raised pre- and post- the inclusion of the new experiments.

The skills in Figure 4 show that the development of practical skills, time management, independence or organisational skills was unaffected by the inclusion of the inquiry/problem/context/industry- based experiments. Four skills - scientific methodology, teamwork, scientific communication and critical thinking - were noted to increase after the inclusion of the new experiments whilst eight new skills emerged, with two of them (commercial awareness non-scientific communication) being raised in more than one focus group. It is important to note that the increase in non-scientific communication and commercial awareness raised by students was only in cases where the assessment was more aligned with real-work practices (i.e. authentic assessment (Gulikers, Bastiaens, & Kirschner, 2006)), such as the generation of executive summaries to corporate entities.

Overall, these results are again in agreement with prior survey data (George-Williams, Soo, et al., 2018) which showed a broader appreciation of skill development in the redesigned experiments, indicating that the effects of the new approach persisted after 2-3 months even in the presence of unaltered expository experiments. With regards to the small number of participants in several focus groups, and the issue of potential non-representation, the strong overlap between these results and those collected via the previously published survey indicate that these issues appeared to have a limited effect on the results.

Annual surveys - Perceptions of the aims of teaching laboratories

With the focus group data showing the longevity of the benefits of inquiry/problem/context/industry- based experiments on a multi-month scale, it was important to consider the effects on an annual scale. Previous work (George-Williams, Ziebell, et al., 2018) showed that six major themes were raised by at least 10% of the students when they were asked to comment on the aims of laboratory experiments; TU (enhancing theoretical understanding), AP (application of theory), PS (developing practical skills), EX (gaining general laboratory experience), WF (preparation for the workforce) and TS (developing transferable skills). These aims were previously tracked over all three year levels in 2016 (George-Williams, Ziebell, et al., 2018) but only 2nd and 3rd year students were surveyed in 2017 and 2018 as the TLL programme did not result in large changes to the 1st year laboratory experiments. Themes that were raised by more than 10% of the respondents to the survey for 2016, 2017 and 2018 are shown below in Table 7.

Table 7. The aims raised by at least 10% of the students by year level 2016-2018. *Significantly different as per the Pearson's chi squared test when compared to 1st years in 2016. ‡Significantly different as per the Pearson's chi squared test when compared to previous years (i.e. 2nd year students in 2017 or 2018 vs 2nd year students in 2016).

2016						
	Most common themes → Least common themes					
1st year students	TU	AP	PS	EX	WF	TS
2nd year students	EX	PS	TU	WF	AP	TS
3rd year students	PS*	AP	TU*	EX*	WF	TS*
2017						
2nd year students	PS‡	WF	AP	TU=EX‡	TS	CX‡
3rd year students	PS	TU=AP	WF	EX‡	TS	-
2018						
2nd year students	WF‡	PS	AP	TU	CX‡=TS	EX‡
3rd year students	WF‡	PS	AP	TU	TS	EX

It was previously noted that the students in 2016 appeared to become more focused on the development of skills (particularly practical skills or PS) at the expense of a focus on theoretical understanding (TU) as they moved through the laboratory programme. In addition, students in their 2nd or 3rd year did not change their perceptions around the importance of preparing for the workforce (WF), contextualisation to the real world (CX) or developing an appreciation of scientific methodology (SM).

With regards to the 2nd year student responses in 2017 and 2018, the increase in the recognition of the importance of contextualisation to the real world (CX) was a welcome one (p= 0.008, ϕ_C= 0.169). This is consistent with previous results (George-Williams, Soo, et al., 2018) which showed that students generally perceived the redesigned experiments as more contextualised. However, this was generally not found as a major theme in the responses of 3rd year students. This is likely the result of the fact that many of the 3rd year laboratory experiments were relatively untouched by the TLL programme (with only 15% redesigned or replaced as compared to 41% in 2nd year). In these cases, the laboratory experiments remained fairly traditional and lacked the strong connections to the real world that was noted in the redesigned units/courses and it is likely that students reflected on their most recent experiences in the 3rd year units/courses.

Gaining general laboratory experience (EX) was found to become a less common theme (-20%, p= 0.004, ϕ_C= -0.237) but it is possible that students tended to write this when they had little else to say. Far more interestingly, preparation for the workforce (WF) became the most commonly raised theme in 2018. This was consistent between both year levels and was found to have a moderate effect size (2nd years – p= 0.026, ϕ_C= 0.133, 3rd years - p= 0.049, ϕ_C= 0.142). Overall, this is a positive outcome in line with the aims of the TLL programme, even though the changes are not large.

Overall, the new approach using inquiry/problem/context/industry- based learning appeared to cause students to gain a greater appreciation that the teaching laboratory can, and should, prepare them for the workforce. They maintained their belief that they should develop practical and transferable skills but failed to recognise the importance of scientific methodology. Finally, students started to raise the importance of contextualisation, but it would seem that this theme was lost in their 3rd year, potentially due to the persistence of traditional laboratory experiences in the 3rd year of their undergraduate degree.

Annual surveys – Perceptions of expected student behaviour

Previous work (George-Williams et al., 2019) has indicated that students tended to have a fairly positive outlook towards their laboratory experiences, and their perceptions did not change as they moved through the undergraduate programme. Whether this was due to the students' perceptions being difficult to shift, the laboratory program continually meeting their expectations or a failure on behalf of the MLLI survey to detect such changes, was unclear. It was of interest to see whether the redesigned laboratory experiments would have any effect on student perception as measured by the MLLI survey and it was again disseminated to 2nd and 3rd year students at the start of semester 2 in 2017 and 2018.

Only one significant change was noted in the responses from 2nd year students. They were more likely to feel that they would learn chemistry useful to their lives. The use of the Pearson's chi-squared test, alongside the use of the Bonferroni corrected cut-off value (p=0.005), is very conservative and is perhaps partially responsible for the lack of observed changes. However, there were no changes detected in the responses from 3rd year students. These results appear to match the results seen for the perception of laboratory aims i.e. gains noted in more positive student perceptions with regards to the 2nd year laboratory experiments were no longer noted in the 3rd year of the undergraduate program, probably due to the persistence of mainly expository experiences in the 3rd year. These results would appear to contradict the work of Szteinberg and Weaver (2013), in which longer lasting effects of their research/inquiry unit/course was noted. However, in that case, students were directly asked about the specific intervention rather than commenting on all laboratory experiences in a general manner (which is the case in this work).

Conclusions

In order to investigate what impact the large scale inclusion of inquiry/problem/context/industry- based experiments was having on student perceptions over a long time scale, data from eight focus groups and two annual surveys in 2016, 2017 and 2018 (N=80-165) were collected and transcribed. The outcomes of the focus groups showed that the benefits identified through the individual experiment exit survey (i.e. broader reasons for enjoyment, broader skill development, greater appreciation for scientific methodology and more recognition of context) were still notable after 2-3 months at the end of the semester. The longevity of these results indicates that the large scale inclusion of these experiments can have positive outcomes that persist for a longer timescale than just immediately after the experience.

The annual MLLI survey and open question (‘What do you think the aims of doing a practical chemistry course are?’) indicated that minimal changes to student perception were measurable on an annual timescale, except that preparation for the workforce became the major laboratory aim. However, there were other positive effects noted in the responses of 2nd year students, such as the inclusion of contextualisation to the real world as an aim of laboratory experiments, and the perception that students would learn chemistry that was useful to their lives. This is consistent with the large number of units/courses in the 3rd year that still consisted of a large number of expository experiences.

Overall, the annual surveys imply that the large-scale inclusion of inquiry/problem/context/industry- based experiments had more significant effects in the 2nd year of the undergraduate degree programme than in the 3rd year. These effects were consistent with prior findings and showed that the large scale implementation of the inquiry/problem/context/industry- based experiments was successful in creating a modern laboratory experience more in line with the needs for the modern workforce. However, these results strongly imply that the positive impact of such changes will be lost if students are subsequently exposed to traditional, expository experiments. Implementation of inquiry/problem/context/industry- based laboratory experiments should be consistent over all years of a programme and this can present challenges.

Acknowledgements

The authors would like to acknowledge and thank, first and foremost, the students who participated in this study, whose honest feedback will result in a better learning experience for all. The authors would also like to express their gratitude to the technical staff at Monash University for their constant support and advice. The authors acknowledge Monash University for funding, and hosting, the Transforming Laboratory Learning program. Ethics approval was obtained from

the Monash University Human Ethics Research Committee, application number 2016000584.

Conflicts of interest

There are no conflicts to declare.

References

Avargil, S., Herscovitz, O., & Dori, Y. J. (2013). Challenges in the transition to large-scale reform in chemical education. *Thinking Skills and Creativity, 10*, 189-207. doi: http://dx.doi.org/10.1016/j.tsc.2013.07.008

Barlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information technology, learning, and performance journal, 19*(1), 43.

Bernard, E., Britz-McKibbin, P., & Gernigon, N. (2007). Resveratrol Photoisomerization: An Integrative Guided-Inquiry Experiment. *Journal of Chemical Education, 84*(7), 1159. doi: 10.1021/ed084p1159

Boud, D. J., Dunn, J., Kennedy, T., & Thorley, R. (1980). The Aims of Science Laboratory Courses: a Survey of Students, Graduates and Practising Scientists. *European Journal of Science Education, 2*(4), 415-428. doi: 10.1080/0140528800020408

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*, 21-32.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn: Erlbaum Associates, Hillsdale.

DeKorver, B. K., & Towns, M. H. (2015). General Chemistry Students' Goals for Chemistry Laboratory Coursework. *Journal of Chemical Education, 92*(12), 2031-2037. doi: 10.1021/acs.jchemed.5b00463

Domin, D. S. (2007). Students' perceptions of when conceptual development occurs during laboratory instruction. *Chemistry Education Research and Practice, 8*(2), 140-152. doi: 10.1039/B6RP90027E

Erhart, S. E., McCarrick, R. M., Lorigan, G. A., & Yezierski, E. J. (2015). Citrus Quality Control: An NMR/MRI Problem-Based Experiment. *Journal of Chemical Education*. doi: 10.1021/acs.jchemed.5b00251

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18. doi: 10.1037/a0024338

Galloway, K. R., & Bretz, S. L. (2015a). Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory. *Journal of Chemical Education, 92*(7), 1149-1158. doi: 10.1021/ed500881y

Galloway, K. R., & Bretz, S. L. (2015b). Measuring Meaningful Learning in the Undergraduate Chemistry Laboratory: A National, Cross-Sectional Study. *Journal of Chemical Education, 92*(12), 2006-2018. doi: 10.1021/acs.jchemed.5b00538

George-Williams, S. R., Karis, D., Ziebell, A. L., Kitson, R. R. A., Coppo, P., Schmid, S., . . . Overton, T. L. (2019). Investigating student and staff perceptions of students' experiences in teaching laboratories through the lens of meaningful learning. *Chemistry Education Research and Practice*. doi: 10.1039/C8RP00188J

George-Williams, S. R., Soo, J. T., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2018). Inquiry and industry inspired laboratories: the impact on students' perceptions of skill development and engagements. *Chemistry Education Research and Practice, 19*(2), 583-596. doi: 10.1039/C7RP00233E

George-Williams, S. R., Ziebell, A. L., Kitson, R. R. A., Coppo, P., Thompson, C. D., & Overton, T. L. (2018). 'What do you think the aims of doing a practical chemistry course are?' A comparison of the views of students and teaching staff across three universities. *Chemistry Education Research and Practice, 19*(2), 463-473. doi: 10.1039/C7RP00177K

Gormally, C., Brickman, P., Hallar, B., & Armstrong, N. (2009). Effects of inquiry-based learning on students' science literacy skills and confidence.

Gulikers, J., Bastiaens, T., & Kirschner, P. (2006). Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional framework. *Journal of Vocational Education and Training, 58*(3), 337-357.

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*: Advanced Analytics, LLC.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*: Routledge.

Hennah, N., & Seery, M. K. (2017). Using Digital Badges for Developing High School Chemistry Laboratory Skills. *Journal of Chemical Education, 94*(7), 844-848. doi: 10.1021/acs.jchemed.7b00175

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review, 70*(256), 33-40.

Kirschner, P. A., & Meester, M. A. M. (1988). The Laboratory in Higher Science Education: Problems, Premises and Objectives. *Higher Education, 17*(1), 81-98. doi: 10.1007/BF00130901

Kitzinger, J. (1994). The methodology of focus groups: the importance of interaction between research participants. *Sociology of health & illness, 16*(1), 103-121.

Kitzinger, J. (1995). Introducing focus groups. *British medical journal, 311*(7000), 299-303.
Kulevich, S. E., Herrick, R. S., & Mills, K. V. (2014). A Discovery Chemistry Experiment on Buffers. *Journal of Chemical Education, 91*(8), 1207-1211. doi: 10.1021/ed400377a
Lazonder, A. W., & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning:Effects of Guidance. *Review of Educational Research, 86*(3), 681-718. doi: 10.3102/0034654315627366
Leal Filho, W., & Pace, P. (2016). Teaching Education for Sustainable Development at University Level: Springer.
Lenhard, W., & Lenhard, A. (2016). Calculation of Effect Sizes., from https://www.psychometrica.de/effect_size.html
Mills, K. V., & Guilmette, L. W. (2007). Thermochemical Analysis of Neutralization Reactions: An Introductory Discovery Experiment. *Journal of Chemical Education, 84*(2), 326. doi: 10.1021/ed084p326
Morgan, D. L. (1996). *Focus groups as qualitative research* (Vol. 16): Sage publications.
Novak, J. D. (1998). *Learning, creating, and using knowledge*: Mahwah, NJ: Erlbaum.
Pilcher, L. A., Riley, D. L., Mathabathe, K. C., & Potgieter, M. (2015). An inquiry-based practical curriculum for organic chemistry as preparation for industry and postgraduate research. *South African Journal of Chemistry, 68*, 236-244.
Pilot, A., & Bulte, A. M. W. (2006). Why Do You "Need to Know"? Context-based education. *International Journal of Science Education, 28*(9), 953-956. doi: 10.1080/09500690600702462
Russel, C. B. (2008). *Development and Evaluation of a Research-Based Undergraduate Laboratory Curriculum.* (PhD), Purdue University, West Lafayette, Indiana.
Russell, C. B., & Weaver, G. (2008). Student perceptions of the purpose and function of the laboratory in science: A grounded theory study. *International Journal for the scholarship of teaching and learning, 2*(2), 9. doi: 10.20429/ijsotl.2008.020209
Sarkar, M., Overton, T., Thompson, C., & Rayner, G. (2016). Graduate Employability: Views of Recent Science Graduates and Employers. *International Journal of Innovation in Science and Mathematics Education., 24*(3), 31-48.
Szteinberg, G. A., & Weaver, G. C. (2013). Participants' reflections two and three years after an introductory chemistry course-embedded research experience. *Chemistry Education Research and Practice, 14*(1), 23-35. doi: 10.1039/C2RP20115A

Appendix

The questions asked of the students in the MLLI survey are shown below in Table 8.

Table 8. Questions asked on the MLLI survey.

When performing experiments in a chemistry laboratory course, I expect to …
1 learn chemistry that will be useful in my life.
2 worry about finishing on time.
3 make decisions about what data to collect.
4 feel unsure about the purpose of the procedures.
5 experience moments of insight.
6 be confused about how the instruments work.
7 learn critical thinking skills.
8 be excited to do chemistry.
9 be nervous about making mistakes.
10 consider if my data makes sense.
11 think about what the molecules are doing.
12 feel disorganized.
13 develop confidence in the laboratory.
14 worry about getting good data.
15 find the procedures simple to do.
16 be confused about the underlying concepts.
17 "get stuck" but keep trying.
18 be nervous when handling chemicals.
19 think about chemistry I already know.
20 worry about the quality of my data.
21 be frustrated.
22 interpret my data beyond only doing calculations.
23 Please select both agree and disagree for this question.
24 focus on procedures, not concepts.
25 use my observations to understand the behaviour of atoms and molecules.
26 make mistakes and try again.
27 be intrigued by the instruments.
28 feel intimidated.
29 be confused about what my data mean.
30 be confident when using equipment.
31 learn problem solving skills.

6.1 Summary of findings

This chapter (and publication) investigated four main research questions over an annual time scale. The conclusions drawn from this part of the study are discussed below with regards to each question:

What is the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' perception of laboratory aims and expectations of their thoughts, actions and feelings during laboratories? (Now over a longer time scale)

The use of the laboratory aims and expectations survey painted a complex picture with regards to students' perceptions of laboratory aims and expectations. Students became more likely, as compared to the baseline data in Chapter 3, to raise preparation of the workforce as an aim for teaching laboratories (+14%) but less likely to raise gaining general laboratory experience (-20%) in their 3rd year of their undergraduate degree. It should also be noted that students began to raise contextualisation (+12%) in their 2nd year, but that this was not found in the responses of 3rd year students (i.e. <10% of students raised the theme). It is possible that the lack of the contextualisation theme in the responses from 3rd year students is simply due to the larger amount of unaltered laboratory activities in the 3rd year of the undergraduate programme. Whilst the TLL programme altered or replaced an almost equal number of teaching laboratories in both year levels, the impact is potentially less noticeable in the 3rd year of the program as there are more units/courses in the 3rd year overall as compared to the 2nd year (eight vs. five). As students are likely responding to the survey with their most recent experiences in mind, the presence of many traditional activities in the 3rd year would be expected to give results that match the 2016 dataset.

With regards to the students' expectations of their thoughts, feelings and actions during a given laboratory (as measured by the MLLI survey), a similar trend to the responses of students towards the aims of teaching laboratories was noted. In their 2nd year, students were much more likely to believe that they would learn chemistry that was useful to their lives. However, by their 3rd year, this change was no longer noted with no significant differences measured between the responses of 3rd year students before or after the large-scale inclusion of inquiry/problem/context/industry-based experiments. As before, these results are likely the result of the large number of traditional laboratory activities still present in the 3rd year of the undergraduate laboratory programme.

What is the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the students' level of enjoyment of the laboratory exercises?

During the focus groups, when asked about laboratory experiences that they enjoyed, students raised a greater number of reasons for enjoyment and regularly raised the inclusion of inquiry or student driven activities and the new industrial or real-world contexts. These results are consistent with the individual activity surveys discussed in Chapter 5 and seem to indicate once again that the positive outcomes of these laboratory activities could still be measured when a complete unit/course was considered rather than an individual activity. No similar question was asked to students at the completion of their program so no further analysis can be made with regards to this research question.

What is the impact of the large-scale inclusion of inquiry/problem/context/industry-based experiments on the development of students' employability skills and their recognition of these skills?

Analysis of the focus group transcripts indicated that the findings of the survey discussed in Chapter 3, in which students showed a broader appreciation for skill development (notably including the development of investigation of scientific methodology skills) was again noted. Hence, this outcome for the large-scale inclusion of inquiry/problem/context/industry-based experiments was still measurable on the slightly longer time scale of one unit/course rather than a single laboratory. It should be noted that whenever authentic assessment was incorporated, commercial awareness and non-scientific methodology also became significant themes regularly raised in the focus groups. Hence, how the laboratory activities were assessed had a marked outcome on the students' perceptions of their own skill development, which is consistent with literature (Darling-Hammond & Snyder, 2000).

Overall, these results appear to indicate that the large-scale inclusion of inquiry/problem/context/industry-based experiments was successful with regards to its aims (higher student enjoyment with more student recognition of skill development) on both the unit/course-level time scale and potentially in the 2nd year of the undergraduate laboratory program. The individual laboratory survey and focus groups show positive outcomes just after the completion of individual laboratories and after the completion of significantly redesigned units/courses. However, the annual survey highlights that the persistence of a significant number of traditional laboratory activities in the 3rd year of the undergraduate chemistry course are potentially undermining these positive outcomes and causing students to no longer raise previous positive gains - such as an appreciation for scientific methodology or the contextualisation of teaching laboratories being recognised as important aims of teaching laboratories. Whilst it is of course possible that the survey utilised was simply unable to detect any significant changes in the 3rd year, the consistency of the lack of changes over the two main measurements (teaching laboratory aims and student expectations of themselves during teaching laboratories) would imply that this is unlikely.

Chapter 7 – Combating variation in marks and the quality of feedback.

Whilst this work predominately focused on the perceptions of students with regards to the new laboratory activities, significant strides were also taken towards affecting the practices of the teaching associates, particularly with regards to the marking of assessment pieces (Page 30, Research Question 4). Through student surveys and other informal feedback processes, it was apparent that at Monash University that there exists significant variation between assessors with respect to marks assigned for laboratory reports and the level and quality of feedback provided to the students. Several interventions were attempted between 2016 and 2018 which included the use of more detailed marking criteria and enhanced training of teaching associates, the use of a specifically designed Excel marking rubric and the use of automated marking processes on the online learning management system (Moodle in this case). The results have this have published in *Assessment and Evaluation in Higher Education* (George-Williams, Carroll, Ziebell, Thompson, & Overton, 2019) and this is reproduced on the following pages[4]:

[4] NB: The online publication also encloses a blank version of the spreadsheet with instructions on its use.

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
https://doi.org/10.1080/02602938.2018.1545897

Curtailing marking variation and enhancing feedback in large scale undergraduate chemistry courses through reducing academic judgement: a case study

Stephen George-Williams, Mary-Rose Carroll, Angela Ziebell, Christopher Thompson and Tina Overton

Monash University, Clayton, Australia

ABSTRACT

Variation in marks awarded, alongside quality of feedback, is an issue whenever large-scale assessment is undertaken. In particular, variation between sessional teaching staff has been studied for decades resulting in many recorded efforts to overcome this issue. Attempts to curtail variation range from moderation meetings, extended training programmes, electronic tools, automated feedback or even audio/video feedback. Decreased marking variation was observed whenever automated marking was used, potentially due to less academic judgment being used by the markers. This article will focus on a case study of three interventions undertaken at Monash University that were designed to address concerns around the variability of marking and the feedback between sessional teaching staff employed in the chemistry teaching laboratories. The interventions included the use of detailed marking criteria, Excel marking spreadsheets and automated marked Moodle reports. Results indicated that more detailed marking criteria had no effect whilst automated processes caused a consistent decrease. This was attributed to a decrease in the academic judgment markers were expected to use. Only the Excel spreadsheet ensured the provision of consistent feedback to students. Sessional teaching staff commented that their marking loads were reduced and the new methods were easy to use.

KEYWORDS
Electronic marking; sessional teaching staff; marking criteria; large cohorts

Introduction

It is well known that when the work produced by a large cohort of students is marked by many assessors, marking variation is likely to occur, i.e. a single piece of work may receive different marks from different assessors. This issue was first significantly studied in the 1980s (Edwards 1979; Bell 1980; Byrne 1980; Hall and Daglish 1982; Collier 1986). The situation is further complicated by having constantly changing sessional teaching staff engaged in marking which can also affect the quality of the marking undertaken (Smith and Coombe 2006). It is worth noting that whilst this issue is common with *any* teaching staff, sessional teaching staff provide a significant challenge simply due to their large number and frequent turnover in many teaching institutions.

One of the more common ways to counter this issue is through the use of structured training (or calibration), which is a process in which markers discuss their viewpoints around a particular

CONTACT Stephen George-Williams [redacted]

Supplemental data for this article can be accessed here.

assessment piece in order to reach a consensus before marking takes place. On such example is the developing understanding of assessment for learning (DUAL) programme which guided sessional teaching staff through the marking process (Bird and Yucel 2013) and was found to lower marking variation. However, arguments against such processes (Bloxham 2009; Bloxham et al. 2016) highlight that a significant amount of effort by all teaching staff is required (i.e. large amounts of time or personal commitment) for these gains to occur.

Research shows that this variation in the marks awarded can often be attributed to the variation in the core beliefs (i.e. what markers may feel is more important or worthy of attention) of the teaching staff themselves (Hunter and Docherty 2011), which would in turn lead to differing levels of use of academic judgment (i.e. how much a given marker would use their internal judgment to decide a mark). Additionally, the degree of marking variation was seen to be discipline specific with a wider spread of marks with greater variation noted from more quantitative subjects (e.g. mathematics) compared to more qualitative subjects (e.g. philosophy) (Bridges et al. 1999).

These studies highlight that marking variation is a complicated issue. Dealing with marking variation becomes even more complex if the amount and quality of the feedback given to students is also considered. It has been shown that students are not always unanimous in the type of feedback they desire, ranging from requesting 'precise guidance' to more general feedback (Bell, Mladenovic, and Price 2013). Furthermore, with the rise of digital technologies, the way in which feedback is delivered is significantly changing. Electronic means of feedback (e.g. digitally annotated reports) have been met with similar levels of satisfaction by students compared to hand written notes and have been noted to reduce the marking time required by teaching staff (Sopina and McNeill 2015). Other forms of electronic feedback are also available, including automated feedback (i.e. computer generated/assigned) (Watt et al. 2002), audio or video feedback (Lunt and Curran 2010), or through use of in-house designed applications (Campbell 2005; Heinrich et al. 2009). These studies also highlight that electronic feedback was more readily accessed/read by students but required significant retraining of teaching staff.

There is evidence also of the limited effects that feedback can have on the students' overall learning if it is not correctly utilised by the student (Crisp 2007). Typically, this is due to the conventional 'transmission' of comments that may or may not be utilised by the students. There is a growing argument in the literature for the use of 'dialogue feedback' methods (Ajjawi and Boud 2017, 2018; Carless 2015) which utilise a conversation between teaching staff and students to guide student learning rather than simply having teaching staff provide a list of comments at the completion of an assessment piece. Whilst fundamentally sound, such a shift in the feedback procedures would require a significant amount of training time for sessional teaching staff who may not remain in post training and would likely limit the amount of assessment that could be potentially undertaken on a given content area. As this work sought to address issues of scale and consistency with regards to marking and feedback, this new framework was not utilised here.

Herein lies the challenge of this work. How should the issue of marking variation and provision of limited feedback by sessional teaching staff be addressed? Additionally, how can improved marking consistency and enhanced feedback be achieved whilst not increasing the time commitment of sessional teaching staff and whilst also avoiding the need for significant retraining?

At Monash University, marking variation and variable quality feedback between different sessional teaching staff had been raised by students through both formal (e.g. course evaluations) and informal means (e.g. direct complaints to academic staff). This case study monitored the effects of three interventions designed to lower the variation in marks awarded for student work and increase the amount of consistent and constructive feedback provided. Ideally, this would be achieved without increasing the workload of sessional teaching staff or requiring a significant amount of retraining. The underlying idea throughout the interventions was that reducing academic judgment should decrease marking variation between markers. Additionally, electronic marking rubrics should ensure students receive at least a base level of feedback.

The focus of this case study was primarily in relation to the assessment of laboratory activities of first year chemistry students. Three interventions were designed and their effect on marking and feedback consistency were investigated. The three interventions were:

1. The use of increased detail in marking criteria and specific marking training.
2. The use of an in-house generated Excel marking spreadsheet.
3. The use of automated rubrics in Moodle (the learning management system used at Monash University). In this case, most of the marking would be performed by the learning management system itself.

Methods

The aim of this study was to monitor the effect of several major changes to the marking procedures in first year chemistry in 2016, 2017 and 2018. This was achieved through analysis of marks awarded to students from 2015 to 2018 and surveys with sessional teaching staff in 2017 and 2018.

Research theoretical framework

The underlying framework of this study is that electronic means of marking and providing feedback might be a useful tool for overcoming marking variation borne from using large numbers of sessional teaching staff. In theory, these processes should reduce the academic judgment required by the sessional teaching staff during marking, which should in turn reduce variation between different markers.

Furthermore, the responses from the sessional teaching staff to the surveys and focus groups were guided by the underlying concept of constructivism, as their responses were influenced by their prior experiences with marking student reports (Leal Filho and Pace 2016). Therefore, the responses from participants to being asked to reflect on their experiences with any given marking process will be as a result of any prior conclusions that they have built for themselves. The use of open questions to encourage full explanation of the answers chosen ('Please explain why you chose the previous answer?') allowed the respondents to fully draw from their own personal experiences and understanding rather than being overly prompted by a survey with more specific closed questions.

Lastly, a case study approach was taken to this research. A case study can generate an in-depth view of a complex issue within a real-life context. In this case, the case study approach was used to generate an in-depth view of the factors that mitigate variation in marking and provided feedback within the context of a large first year undergraduate chemistry laboratory course.

Increased detail in marking criteria and new marking training

Before any intervention was undertaken, the marking criteria provided to the sessional teaching staff were often vague. For example, several discussion sections (typically worth more than 10% of the total marks awarded) were merely assigned a mark with no details given on what was expected to be written in that section. As such, it was generally up to the marker to decide what they felt was required for the discussion section resulting in a large requirement for academic judgment. Therefore, as part of the first intervention, all marking criteria were enhanced to ensure that they included detailed expectations for each mark to be allocated. These new

	0 pts		0.5 pts		1 pts	
				Section One		
				Section Two		
Criteria	Sub-criteria	X	Sub-criteria		Sub-criteria	
Criteria	Sub-criteria		Sub-criteria	X	Sub-criteria	
	Sub-criteria		Sub-criteria		Sub-criteria	X

Figure 1. An example Excel marking spreadsheet showing the auto-coloured cells.

criteria were distributed to the markers before the beginning of semester 2, 2016. At this stage, students reported the outcomes of their laboratory activities by completing proformas containing the questions to be answered, tables for results and details of calculations to be completed. The detailed marking criteria were not shared with the students in this case as the marks were often assigned to very specific concepts that the students could simply copy in order to obtain marks. Feedback provided by the sessional teaching staff only existed as electronic annotations on the proformas and was not monitored in this study (as feedback in previous years was paper-based and not captured for comparison).

Alongside this, the training programme provided for all sessional teaching staff was modified to include a marking activity in which they marked reports with and without detailed marking criteria. The results of this activity were shared during the training session and marking variation and feedback were discussed.

Generation of Excel spreadsheet

The automated Excel spreadsheet was first generated in early 2017 as an electronic version of a marking rubric for a higher year chemistry laboratory. Over several iterations, the final version adopted a rubric style with marking criteria in each cell and an additional cell adjacent to each item in the rubric. This cell could only be set to either blank or an X using a dropdown box. If an X did populate this cell, it automatically became coloured (red for a zero mark, yellow for an intermediate mark and green for full marks) and would automatically generate the correct mark on the right side of the rubric table. This is shown in Figure 1.

The mark was assigned using an embedded IF function, which is an Excel logical test that will return a different value depending on if a given condition is met or not (which, in this case, was where the X had been placed in the spreadsheet). Another column on the right-hand side of the rubric allowed sessional teaching staff to provide additional feedback if required. Sections of the spreadsheet were designed to collapse to ease reading ('accordion style') and every box (except the dropdown boxes and general feedback column) was set to 'locked' to prevent changes being made by the markers. A macro was required to allow the collapsible sections to be opened and closed whilst allowing for this global protection. The total mark was automatically calculated by the spreadsheet to avoid arithmetic error. Finally, any unmarked section read 'No mark' and was highlighted light red to ensure the marker realised what sections still needed to be marked. This spreadsheet will be supplied as supplementary information.

This marking process was utilised in four experiments in the second semester of 2017. Its use was not made compulsory but 80% of the sessional teaching staff choose to use it throughout the semester.

Generation of online Moodle reports

The automated Moodle reports were generated in such a manner to include similar questions originally given to the students in previous years within the proformas. Consequently, there

Table 1. Questions asked in the online survey of sessional teaching staff.

Question	Closed answer options			
In terms of difficulty, how did you find using the new marking method?	Easy	Hard		Neither
Compared to marking in previous years, how was your rate of marking affected?	Faster	Slower		Neither
Compared to marking in previous years, how did the amount of feedback you provided change?	Increased	Decreased		Neither
Compared to marking in previous years, how did the quality of feedback that you provided change?	Increased	Decreased		Similar
Using the new Moodle marking, how long did it take you to mark a full set of reports per group for a given experiment (on average)?	Under 1 hour	Between 1 and 2 hours	Between 2 and 3 hours	Greater than 4 hours

were no longer in-class proformas to be completed and students were expected to record all relevant data into a laboratory notebook for which they were assigned a portion of the marks. Some items present in the automated Moodle assessment included, but were not limited to:

- An electronic table for results where students could input numerical values which were automatically marked.
- Dragable items that could either be used to insert the correct words into incomplete method sections or to insert missing items into blank or incomplete chemical equations.
- Multiple choice or dropdown questions to probe conceptual understanding.
- Essay style questions to allow students to input full scientific discussions and conclusions. These sections would be marked directly by sessional teaching staff.
- Submission sections where students could upload scanned documents (e.g. laboratory notes or graphs) to be marked directly by sessional teaching staff.

This style of submission and marking was undertaken for all experiments in semester 1, 2018. Some automated feedback was provided for the closed questions whilst the sessional teaching staff were asked to leave guidance on the open questions that they marked.

Data collection

Anonymised student marks were collected through Moodle alongside information about group allocations, allowing for marks to be assigned to a specific sessional teacher. The enrolments in the monitored first year chemistry subjects were typically 1000–1200. These data were collected from 2015 to 2018.

Surveys were disseminated to sessional teaching staff through the use of an online Google form in 2017 and 2018. There were typically 30–60 sessional teaching staff assigned to the first year laboratories. The questions asked are shown in Table 1. The final question was only asked in 2018. Every closed question was also followed by a blank box where sessional teaching staff were encouraged to expand on why they chose their given answer. A total of 22 sessional teaching staff responded in 2017 (out of 26 who used the Excel spreadsheet) and 35 in 2018 (out of 53 who taught in that year).

Data analysis – marking variation

The student marks data was first analysed in order to address whether or not marking variation was effected by the marking process utilised. In this case, marking variation was defined as whenever a given group of students, who were all marked by a single assessor, received an average mark that was considered 'low' or 'high' as compared to the average mark of the entire year level.

To achieve this, individual student marks were first separated into their assigned groups (of 12–16 students) for each experiment. This data was then analysed using SPSS to generate an average mark for each group of students for each week. Excel was then utilised to determine whether a given average mark was 10% higher or lower than the overall average mark for the entire student cohort for that experiment (e.g. a group average mark of 70% when the cohort average was 85% would be marked as 'low'). These results were then grouped into four categories:

i. **Baseline** – 2015, semester 1 and 2 with 2016, semester 1. This dataset represented marking variation observed before any of the three interventions.
ii. **Increased detail** – 2016, semester 2 (all weeks), 2017, semester 1 (all weeks) and 2017, semester 2, weeks 3, 4 and 7. This dataset represented marking variation observed for experiments marked after the introduction of the enhanced marking criteria.
iii. **Excel** – 2017, semester 2, weeks 1, 2, 5 and 6. This dataset represented marking variation observed for experiments marked using the Excel marking spreadsheets.
iv. **Moodle** – 2018, semester 1. This dataset represented marking variation in experiments marked using the Moodle marking system.

These categorised datasets were then treated with a Pearson's chi squared test to measure whether any significant differences between reports marked by different processes could be noted (i.e. $p < 0.05$). This was followed by a calculation of Cramer's V to determine the effect size of any change in marking variation noted between the different marking methods. The cut-offs for the 'size' of the effects were determined through the work of Hattie (2008), which was later extended (Fritz, Morris, and Richler 2012; Lenhard and Lenhard 2016) to r. As the original cut-offs for r matched those for Cramer's V (Cohen 1988) the same altered ranges were used in this analysis. As such, the ranges were defined as:

1. $0 \leq V \leq 0.1$. 'Student effect size'. This refers to the natural variation in any group of students. For example, a more motivated student may respond more positively than a less motivated student.
2. $0.1 < V \leq 0.2$. 'Teacher effect size'. This refers to the effect of a particularly motivated teacher over the course of a single year (i.e. this effect size could be achieved given time/motivation).
3. $V > 0.2$. 'Zone of desired effect'. This refers to interventions that have an immediate impact and are where educators should typically focus their efforts.

Additionally, the number of times a group average mark was noted to be either high or low was also tallied. This was then converted to a percentage of the total number of groups (e.g. if there were 10 groups out of 72 that were marked either high or low, then 14% of the groups experienced large marking variation). This was performed so that the differences could be readily graphed and visually compared.

Data analysis – surveys of sessional teaching staff

The surveys delivered to the sessional teaching staff were analysed in order to determine how the sessional teaching staff perceived the new marking procedures with regards to:

1. their ease of use,

2. the marking loads resulting from them,
3. the quality or amount of their provided feedback.

Quantitative data resulting from the surveys remained untreated beyond representing the responses as a percentage of respondents selecting a given answer. Qualitative data (i.e. responses to the open questions) were read multiple times by a single researcher until prominent themes were identified. These themes (and the original data) were then given to two other chemistry education researchers to ensure the themes held inter–rater reliability to at least an 80% agreement level (i.e. that the themes extracted were independent of a single researcher). Once the themes had been generated, the original data was coded to them. Themes raised by more than 10% of the respondents were then recorded for comparison.

Limitations

There are three major limitations to this work:

1. As the number of sessional teaching staff was quite small (30–60) and only from one institution, their responses cannot be generalised beyond Monash University. Additionally, the responses may not be fully representative of the sessional teaching staff at Monash University as the response rates were typically 66–85%.
2. The viewpoints of the students were not considered and could easily form the basis of future work.
3. As these changes occurred over one semester each, it is possible that the sessional teaching staff were simply gaining experience over time with the assessment materials themselves (i.e. the teaching laboratories). However, with a high turnaround of sessional teaching staff, this is considered unlikely.

Results and discussion

The overall demographics of the sessional teaching staff at Monash University has been previously reported (George-Williams et al. 2018). In 2017, there was an even split of male and female-identifying sessional teaching staff and the majority were over 22 years of age. The level of teaching experience varied greatly with over 40% having taught for less than one year. Approximately 75% were enrolled as postgraduate students at Monash University. The survey results in 2018 indicated that these values still held.

The timeline of the changes to the laboratory programme from 2015 to 2018 are shown in Table 2. Note that some changes were made simultaneously and, with the exception of the enhanced training and providing greater marking guidance, changes were trialled for a single semester before analysis. It is important to note that the bulk of the questions asked in the post-laboratory assessment did not significantly change between years (unless a new experiment was introduced). Hence, only the mechanics by which the marking was undertaken was significantly altered.

Marking variation

The percentage of groups marked either 10% above or below the cohort average are shown in Figure 2. The Pearson chi-squared test results between the sub-groups are also shown in Table 3.

In Figure 2, it would appear that (in comparison to the baseline dataset), the use of increased detail in the marking criteria potentially caused a slight increase in marking variation. However,

Table 2. The changes to the laboratory programme delivered in either semester 1 or semester 2 from 2015 to 2018.

Year	Semester 1	Semester 2
2015	Baseline. Seven experiments.	Baseline. Six experiments.
2016	No significant changes made.	New experiment 4. *Increased detail in marking criteria and enhanced training.*
2017	*Increased detail in marking criteria and enhanced training.*	*Use of automated Excel spreadsheets.*
2018	*Use of automated Moodle marking.* New experiment 1.	N/A

Note. The items in italics are the three major changes being monitored in this study.

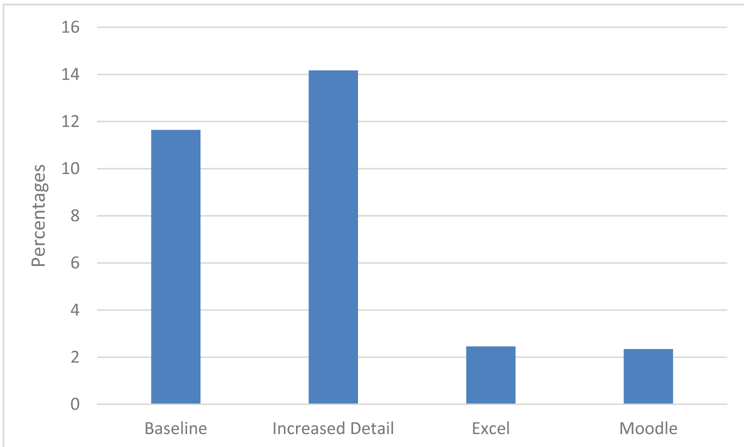


Figure 2. The percentage of groups marked 10% above or below the cohort average for laboratories assessed from 2015 to 2018.

Table 3. Pearson chi-squared results comparing the baseline, increased detail, Excel and Moodle datasets.

	p	φ_c
Baseline vs. Increased detail	0.063	–
Increased detail vs. Excel	<0.0005	0.224
Increased detail vs. Moodle	<0.0005	0.228
Moodle vs. Excel	0.929	–

this change is not significant according to the Pearson's chi-squared test ($p = 0.063$) which implies that increasing the detail in the marking criteria, and providing more specific moderation exercises during teaching associate training, had little to no effect on the overall marking variation measured. It would appear that this intervention had no impact on the academic judgment being used by the sessional teaching staff.

The next comparisons undertaken were between the increased detail subset and either the Excel or Moodle datasets. No comparison was made with the Baseline dataset as the Excel and Moodle marking interventions were undertaken after the marking criteria were altered and the new training program was put in place. Hence, any effect caused by the increased detail intervention should also effect the Excel and Moodle marking datasets. Additionally, before direct comparisons could be made, the large difference in the sample size between the increased detail dataset (1157) and either the Excel data set (203) or the Moodle data set (553) needed to be taken into account. Hence, only every fifth result from the increased detail dataset (231 in total) was used to compare to the Excel dataset and every second result (578 in total) used to compare to the Moodle dataset. This method of analysis has been shown to be valid and accounts for the sensitivity of the Pearson's chi-squared to large variations in sample sizes (Barlett, Kotrlik, and Higgins 2001).

The next result that can clearly be seen in Figure 2 is that the use of either the Excel marking spreadsheet or the Moodle automated marking appeared to lower the marking variation as compared to the increased detail or baseline dataset. These changes were found to not only be significant by the Pearson's chi squared test ($p < 0.0005$ for either comparison) but were also shown to have similar effect sizes within the 'zone of desired effect' ($\varphi_c = 0.224$ and $\varphi_c = 0.228$ for the Excel and Moodle comparisons, respectively). Interestingly, the Excel and Moodle datasets were also found not to be significantly different from one another ($p = 0.929$), implying that they had similar effects on the overall marking variation noted. This decrease in marking variation for either the Excel spreadsheets or the Moodle marking system is consistent with the literature on the use of electronic marking. This decrease also holds to the hypothesis that decreasing the academic judgment required by the marker would indeed consistently lower marking variation between sessional teaching staff.

Whilst these results are encouraging, it is important to consider what other effects the revised marking processes may have had. During the marking process, the Excel spreadsheets were automatically coloured (green for full marks, yellow for partial and red for zero marks) so that students could easily identify where they had gained or lost marks throughout their work. As the Excel spreadsheet was then electronically delivered to the student, students could see both where they performed well (positive feedback) and where they could improve (constructive feedback). Hence, all students received a consistent level of feedback regardless of the sessional teaching staff assigned to their group. It is of course hard to gauge if students read this as feedback but previous literature (Sopina and McNeill 2015) would suggest that electronic feedback is likely read no more or less than physical (i.e. non-electronic) feedback.

The use of the online Moodle report was designed to indicate where students lost or gained marks (e.g. if they had received 0, 1 or 2 marks on a given question). However, it was not populated with automated feedback stating exactly *why* the student had obtained the mark they did (i.e. it did not then provide the correct answer). Clearly, this is a potential area for improvement in future iterations of this work. The only directed feedback that a student could receive through marking on Moodle was in the sections marked by a sessional teacher (the discussion and conclusion sections and any uploaded scanned documentation). Indeed, an analysis of the marking variation in just the discussion and conclusion sections for experiment 2, 2018 (semester one) showed that up to 35% of the group marks for the discussion and conclusion respectively were found to deviate from the cohort average. This would seem to imply that the issue of sessional teaching staff marking variation (i.e. their overuse of their own personal academic judgement) was not actually decreased through the use of the Moodle marking system as academic judgement was still being highly utilised by the markers. However, overall marking variation was lowered as the large amount of automated marking overcame the smaller amount of marks awarded by the sessional teaching staff.

Lastly, it is important to consider whether the altered marking processes potentially affected student learning. One might consider that the use of automated Moodle marking could potentially reward students for more lower-level thinking (e.g. simple recall) rather than encouraging the more higher-level thinking required for writing scientific discussions (or similar tasks). Whilst this is possible, the breakdown of marks was not altered throughout each intervention, such that the more open written sections were still marked to the same total marks in every semester. Whether students focused less on these sections in the more automated processes is difficult to determine without direct student responses – which (as mentioned previously) could form the basis of future work.

Survey responses from sessional teaching staff

In late 2017 and mid-2018, sessional teaching staff were asked a number of questions through a Google survey. The percentage responses from 22 to 35 sessional teaching staff in 2017 and 2018, respectively, are shown in Table 4.

Table 4. Sessional teaching staff responses to the survey questions.

In terms of difficulty, how did you find using the:			
	Easy	Hard	Neither
Excel spreadsheet	77.3%	9.1%	13.6%
Moodle marking report	62.9%	2.9%	34.3%
Compared to marking in previous years, how was your rate of marking affected?			
	Faster	Slower	Neither
Excel spreadsheet	63.6%	10.5%	25.8%
Moodle marking report	85.0%	5.0%	10.0%
Compared to marking in previous years, how did the amount of feedback you provided change?			
	Increased	Decreased	Neither
Excel spreadsheet	13.6%	54.5%	31.8%
Moodle marking report	40.0%	30.0%	30.0%
Compared to marking in previous years, how did the quality of feedback that you provided change?			
	Higher	Lower	Similar
Excel spreadsheet	27.3%	27.3%	45.5%
Moodle marking report	60.0%	10.0%	30.0%

Note. The shaded values are those where more than 60% of respondents choose a given response.

The majority of sessional teaching staff appeared to find both the Excel marking spreadsheets and the Moodle marking process easy to use (77% and 63%, respectively). Those that found the Excel spreadsheet difficult referred to a belief that it either slowed down their marking rate ('*It was more time consuming*') or was tedious and removed their autonomy ('*so long and a little too automated*'). Positive responses tended to focus on either the appearance ('*Well laid out and designed*'), the lower amount of interpretation required ('*Felt like less thinking/interpretation*') or how the spreadsheet was '*was quick and easy to use*'. The Excel spreadsheets appeared to increase the marking rate of many sessional teaching staff (64%), but some (26%) commented on the difficulty of marking student responses that were not covered by the spreadsheet ('*they might not have that point covered in excel spreadsheet and it becomes little tough to award the marks*'). It should be noted that whilst the issue of the prescriptive nature of the Excel spreadsheet can be potentially mitigated through carefully crafted rubrics, the goal of this work was to remove the academic judgement of the sessional teaching staff. Hence, sessional teaching staff who are uncomfortable with less freedom in their marking will always likely have this issue with the Excel spreadsheets.

The Moodle marking was noted to result in '*less to mark overall*'. This was further indicated through an increased rate of marking, with 85% of sessional teaching staff stating that they felt that their marking rate had increased. Neutral or negative responses tended to focus on the apparent compromise between speed and the lack of functionality ('*with the previous methods … I could highlight and address points in their discussion more clearly*').

With regards to the feedback provided to the students, sessional teaching staff clearly felt that they provided less when using the Excel spreadsheets (55%) than in previous years. This was attributed to the belief that the Excel spreadsheet '*was detailed enough that feedback was within it*'. As such, sessional teaching staff appeared to feel that they *personally* provided less feedback (as compared to previous years) rather than the student receiving less feedback (as it was potentially already contained in the Excel spreadsheet).

Interestingly, there was a split in opinion with regards to the amount of feedback provided using the Moodle marking process with around a third of respondents believing they gave more (40%), less (30%) or just as much (30%) as they used to. Sessional teaching staff who believed they gave less feedback stated '*obviously those sections no longer being marked by me I am not providing feedback on. So overall I would be providing less feedback*'. Sessional teaching staff who stated that they gave more feedback raised that '*I have more time to address the issues with each individual student*', whilst those who felt that no change had occurred did not readily provide a reason for this belief.

Lastly, the sessional teaching staff's perception of the quality of the feedback they provided was investigated. With regards to the Excel spreadsheet, sessional teaching staff were somewhat divided with as many stating that the feedback quality was higher (27%) as those stating that it was lower (27%) than in previous years. Those that felt their quality was higher, generally stated that they felt they '*can give extra feedback addressed to certain personalized problems for the student not included in the spreadsheet content*'. Those that felt their feedback was of lower quality commented on the spreadsheet being too prescriptive ('*I used to be much more holistic in being able to take into account student understanding, rather than whether they hit specific boxes*'). With regards to the Moodle marking reports, sessional teaching staff tended to feel as though the feedback they gave was of higher quality (60%) as they had '*more time to*' give feedback and that '*I think it [the feedback] was focussed more on the bigger picture, not just telling them why their calcs were wrong*'.

Overall, it would appear that sessional teaching staff believed that they gave more, higher quality feedback when using the Moodle marking reports. Further research would be required to measure these potential questions through the perspective of the students.

Conclusions

A case study at Monash University wherein three teaching interventions designed to reduce the effects of academic judgment on marking variation, alongside increasing feedback to students, has been described. This case study took place between 2016 and 2018 and included the use of enhanced marking criteria and teaching associate training, Excel spreadsheets as automated electronic rubrics and automated Moodle marking reports. Statistical analysis of student marks and two online surveys delivered to sessional teaching staff were used to monitor the effects of these interventions.

Detailed analysis of the marking variation between sessional teaching staff revealed that the use of increased detail in the marking criteria and enhanced teaching training had little consistent effect on whether sessional teaching staff were marking high or low (i.e. they were marking with an average mark 10% below or above the cohort average). This would seem to imply that typical moderation procedures (i.e. only performed once during a yearly training programme) have a relatively low impact on this issue, which is consistent with several studies in the literature.

The use of the Excel marking spreadsheets or the Moodle marking reports was shown to consistently reduce the marking variation. Overall, it would appear that electronic means of marking (Excel and Moodle) were more effective at reducing the impact of the varied academic judgment of the sessional teaching staff. This appeared to be caused by the nature of the Excel spreadsheet forcing a marker to use the marking guideline as provided, or by the nature of the Moodle marking system simply isolating the variation caused by sessional teaching staff to smaller sections of the marked assessments.

Both Excel spreadsheets and the Moodle marking reports appeared to (on average) reduce the marking load of the sessional teaching staff, with the Moodle marking reports having a larger effect. Both processes were considered easy to use, but it was believed that the Moodle marking reports resulted in a higher quality of feedback as per the responses of the sessional teaching staff. This is particularly interesting as only the use of the Excel marking spreadsheet actually ensured that all students received at least a base level of feedback.

It is worth noting that sessional teaching staff may begin to feel that they are no longer impactful in the marking process if their ability to utilise their academic judgment is too reduced (as was noted through the use of the Excel marking spreadsheet). Whether this concern is outweighed by the gains seen (i.e. less marking variation and greater control over provided feedback) is a matter for further consideration. However, involving the sessional teaching staff, where possible, in the generation of these materials may provide a sense of ownership and, in turn, a solution to this issue.

Overall, automated marking can simplify the marking process, reduce marking variation and ensure a certain level of feedback is provided to all students. This is generally achieved through the removal of personal, academic judgement which results in more reliable, consistent marking practices. These results are consistent with the current literature on automated marking processes, providing another example of their use. It is worth finally noting that the Excel marking spreadsheets are being used in higher year levels where students are expected to write longer open reports (and fewer marks are assigned to background theory questions). The Moodle marking process is continuing to be used in the first year laboratories due to the more controlled marking loads of sessional teaching staff combined with the significantly lowered marking variation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Stephen George-Williams is a PhD candidate (Chemistry Education) at Monash University. His current work primarily focuses on studying the effects of the use of large scale inquiry- and context-based learning in undergraduate teaching laboratories.

Mary-Rose Carroll is a senior laboratory manager in the School of Chemistry at Monash University. Her primary role is to ensure the smooth operation of the undergraduate laboratories but regularly contributes to their content and training of sessional teaching staff.

Angela Ziebell is an Education Developer at Monash University. She is currently focused on a programme called Transforming Laboratory Learning which seeks to overall the undergraduate teaching laboratories to increase the employability of the students.

Christopher Thompson is both an Associate Professor of chemistry education and the Associate Dean of Education at Monash University. His work primarily focuses on the use of performance in science to enhance theoretical understanding whilst increasing student engagement and imparting a range of other transferable skills.

Tina Overton is a Professor of chemistry education at Monash University. Her illustrious career has focused mostly on the use of problem-based learning to enhance the problem solving ability of undergraduate students.

ORCID

Stephen George-Williams http://orcid.org/0000-0002-2578-1187
Mary-Rose Carroll http://orcid.org/0000-0003-0362-9560
Christopher Thompson http://orcid.org/0000-0001-6913-9922
Tina Overton http://orcid.org/0000-0001-5141-6194

References

Ajjawi, R., and D. Boud. 2017. "Researching Feedback Dialogue: An Interactional Analysis Approach." *Assessment & Evaluation in Higher Education* 42 (2):252–265. doi:10.1080/02602938.2015.1102863

Ajjawi, R., and D. Boud. 2018. "Examining the Nature and Effects of Feedback Dialogue." *Assessment & Evaluation in Higher Education* 43 (7):1106–1119. doi:10.1080/02602938.2018.1434128

Barlett, J. E., J. W. Kotrlik, and C. C. Higgins. 2001. "Organizational Research: Determining Appropriate Sample Size in Survey Research." *Information Technology, Learning, and Performance Journal* 19 (1):43–50.

Bell, A., R. Mladenovic, and M. Price. 2013. "Students' Perceptions of the Usefulness of Marking Guides, Grade Descriptors and Annotated Exemplars." *Assessment & Evaluation in Higher Education* 38 (7):769–788. doi:10.1080/02602938.2012.714738.

Bell, R. C. 1980. “Problems in Improving the Reliability of Essay Marks.” *Assessment in Higher Education* 5 (3): 254–263. doi:10.1080/0260293800050303.

Bird, F. L., and R. Yucel. 2013. “Improving Marking Reliability of Scientific Writing with the Developing Understanding of Assessment for Learning Programme.” *Assessment & Evaluation in Higher Education* 38 (5): 536–553. doi:10.1080/02602938.2012.658155.

Bloxham, S. 2009. “Marking and Moderation in the UK: false Assumptions and Wasted Resources.” *Assessment & Evaluation in Higher Education* 34 (2):209–220. doi:10.1080/02602930801955978.

Bloxham, S., B. den-Outer, J. Hudson, and M. Price. 2016. “Let’s Stop the Pretence of Consistent Marking: Exploring the Multiple Limitations of Assessment Criteria.” *Assessment & Evaluation in Higher Education* 41 (3):466–481. doi:10.1080/02602938.2015.1024607.

Bridges, P., B. Bourdillon, D. Collymore, A. Cooper, W. Fox, C. Haines, D. Turner, H. Woolf, and M. Yorke. 1999. “Discipline-Related Marking Behaviour Using Percentages: A Potential Cause of Inequity in Assessment.” *Assessment & Evaluation in Higher Education* 24 (3):285–300. doi:10.1080/0260293990240303.

Byrne, C. 1980. “Tutor Marked Assignments at the Open University: A Question of Reliability.” *Assessment in Higher Education* 5 (2):150–167. doi:10.1080/0260293800050203.

Campbell, A. 2005. “Application of ICT and Rubrics to the Assessment Process Where Professional Judgement Is Involved: The Features of an e-Marking Tool.” *Assessment & Evaluation in Higher Education* 30 (5):529–537. doi:10.1080/02602930500187055.

Carless, D. 2015. *Excellence in University Assessment: Learning from Award-Winning Practice*. London: Routledge.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum Associates.

Collier, M. 1986. “A Specific Investigation of Relative Performance of Examination Markers.” *Assessment & Evaluation in Higher Education* 11 (2):130–137. doi:10.1080/0260293860110204.

Crisp, B. R. 2007. “Is It Worth the Effort? How Feedback Influences Students’ Subsequent Submission of Assessable Work.” *Assessment & Evaluation in Higher Education* 32 (5):571–581. doi:10.1080/02602930601116912.

Edwards, D. 1979. “A Study of the Reliability of Tutor Marked Assignments at the Open University.” *Assessment in Higher Education* 5 (1):16–44. doi:10.1080/0260293790050102.

Fritz, C. O., P. E. Morris, and J. J. Richler. 2012. “Effect Size Estimates: Current Use, Calculations, and Interpretation.” *Journal of Experimental Psychology: General* 141 (1):2–18. doi:10.1037/a0024338.

George-Williams, S. R., A. L. Ziebell, R. R. A. Kitson, P. Coppo, C. D. Thompson, and T. L. Overton. 2018. “What Do You Think the Aims of Doing a Practical Chemistry Course Are? A Comparison of the Views of Students and Teaching Staff across Three Universities.” *Chemistry Education Research and Practice* 19 (2):463–473. doi:10.1039/C7RP00177K.

Hall, C. G. W., and N. D. Daglish. 1982. “Length and Quality: An Exploratory Study of Inter-Marker Reliability.” *Assessment & Evaluation in Higher Education* 7 (2):186–191. doi:10.1080/0260293820070209.

Hattie, J. 2008. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge.

Heinrich, E., J. Milne, A. Ramsay, and D. Morrison. 2009. “Recommendations for the Use of e-Tools for Improvements around Assignment Marking Quality.” *Assessment & Evaluation in Higher Education* 34 (4):469–479. doi:10.1080/02602930802071122

Hunter, K., and P. Docherty. 2011. “Reducing Variation in the Assessment of Student Writing.” *Assessment & Evaluation in Higher Education* 36 (1):109–124. doi:10.1080/02602930903215842.

Leal Filho, W., and P. Pace. 2016. *Teaching education for sustainable development at university level*. Berlin: Springer.

Lenhard, W., and A. Lenhard. 2016. Calculation of effect sizes. https://www.psychometrica.de/effect_size.html

Lunt, T., and J. Curran. 2010. “Are You Listening Please? The Advantages of Electronic Audio Feedback Compared to Written Feedback.” *Assessment & Evaluation in Higher Education* 35 (7):759–769. doi:10.1080/02602930902977772.

Smith, E., and K. Coombe. 2006. “Quality and Qualms in the Marking of University Assignments by Sessional Staff: An Exploratory Study.” *Higher Education* 51 (1):45–69.

Sopina, E., and R. McNeill. 2015. “Investigating the Relationship between Quality, Format and Delivery of Feedback for Written Assignments in Higher Education.” *Assessment & Evaluation in Higher Education* 40 (5):666–680. doi:10.1080/02602938.2014.945072.

Watt, S., C. Simpson, C. McKillop, and V. Nunn. 2002. “Electronic Course Surveys: Does Automating Feedback and Reporting Give Better Results?” *Assessment & Evaluation in Higher Education* 27 (4):325–337. doi:10.1080/0260293022000001346.

7.1 Summary of findings

This case study provided some interesting insights. The results indicated that:

- Using more detailed marking criteria and enhancing the annual training programme had no significant effect on the marking variation between teaching associates.
- The more automated processes (Excel and Moodle), which were designed to require less academic judgment to be used by the teaching associates, significantly reduced marking variation between them. These processes were also found to reduce the marking load on the teaching associates, with the Moodle-based method showing the biggest reported reduction.
- Feedback by the teaching associates was of reportedly higher quality (according to the teaching associates) when the Moodle marking system was utilised. However, a baseline level of feedback to ALL students could only be ensured when the Excel spreadsheet was utilised.
- Some teaching associates struggled with the use of automated marking as they felt it removed their opportunity to make an impact on student learning.

Overall, these interventions showed that automated marking processes, when properly utilised, can provide a powerful means of reducing marking variation through reduced use of academic judgment. Furthermore, with the use of Excel marking rubrics, the level of feedback can be applied in a more consistent manner to all students. However, how much this feedback is utilised by students is a topic for further study.

Chapter 8 – Conclusions, implications for practice/research and future work.

Between early 2016 and late 2018, the Transforming Laboratory Learning programme resulted in 27 new laboratory activities at Monash University that were generated utilising context-, inquiry-, problem- and industry-based pedagogies. As a result, over 40% of 2nd year and approximately 15% of 3rd year teaching laboratory activities were either significantly redesigned or replaced. Additionally, new assessment techniques were also generated utilising either Excel or Moodle and disseminated to teaching associates for marking of student reports in 1st year.

Previous literature has shown that the use of context- and inquiry-based activities has significant benefits such as higher student enjoyment, an increased appreciation of scientific methodology and a chance to develop students' higher-level cognitive skills (such as critical thinking and problem-solving skills). This study indicates that this can be achieved on scale over multiple year levels and sub-disciplines (e.g. organic, physical, analytical or inorganic chemistry) with retention of these benefits still found. Furthermore, the use of industry or other work environments as a source of context was noted to further increase student enjoyment and was particularly beneficial to broadening students' sense of skill development when assessment was altered to match the new work-based context.

Students generally found the new inquiry/problem/context/industry-based laboratory activities more difficult, more open, more contextualised and were less dependent on the laboratory manual for success. Furthermore, their recognition of skill development in the new activities was broader and often included a deepened appreciation of scientific methodology and experimental planning. Whilst students regularly raised the desire to have more guidance, they were also more likely to raise the context, challenge and open nature of the new laboratory activities as reasons for enjoyment.

The findings also indicated that students enjoyed the new activities for a greater variety of reasons, particularly the use of context- and inquiry-based activities. When the assessment was significantly changed to use more authentic tasks, students were more likely to recognise the development of commercial awareness and non-scientific communication skills which highlights the importance of authentic assessment in meaningful learning and the development of employment skills. The positive outcomes of the large-scale use of the inquiry/problem/context/industry based laboratory activities were not limited to immediately after the completion of the new activities but were also measurable at a unit/course level.

Student perceptions of laboratory aims were first noted to be relatively narrow and focused predominately on aims more aligned with traditional, expository laboratories. Little mention was made of scientific methodology or experimental planning. As students progressed through their undergraduate degrees, their focus on theoretical understanding decreased and they developed a greater appreciation of the development of skills, both practical and transferable. In contrast, teaching associates and academic staff held a slightly narrower view of aims of laboratory activities and did not typically see them as a chance to develop general experience or to prepare the students for the workforce. It is important to note that most teaching staff, like the students, failed to raise either scientific methodology or experimental planning. This is likely the result of the large number of expository experiments that both students and teaching staff have either encountered or taught.

Before any intervention had taken place, students were found to hold very positive expectations towards their undergraduate laboratory experiences, undertaking activities that would contribute towards their meaningful learning. There were few changes over year levels or between institutions implying that the laboratory programmes being offered either had no effect on or were meeting student expectations. The viewpoints of teaching staff were very different, with staff far more likely to believe that students would undertake activities that would detract from their meaningful learning.

Whilst teaching associates matched the students' positivity with regards to activities that would contribute towards their meaningful learning, academic staff were less inclined to believe that students would have a positive experience. These results indicated a chasm between student and teaching staff expectations (as per the discussion around laboratory aims).

After the large-scale inclusion of inquiry/problem/context/industry-based experiments, it was noted that students enrolled in 2^{nd} year units/courses recognised the contextualisation of the theory into the real-world as a major laboratory aim and were more likely to expect to learn chemistry that would be useful in their lives. However, no such changes were noted in the responses of the 3^{rd} year students, who undertook a still largely traditional laboratory programme. This would seem to imply that whilst students were beginning to recognise the benefits of the industry- and context-based activities, this was lost when students began undertaking mostly expository experiences later on. Hence, whilst attitudes and perceptions may shift because of new experiences, these same attitudes and perceptions may revert to their previous state if the learning environment provided also reverts to the original, more traditional state.

With regards to the new assessment techniques, when the level of academic judgement was reduced either using the Excel spreadsheet or the Moodle marking system, there was a significant decrease in the variation between markers. Additionally, teaching associates commented that the electronic marking processes were fast and easy to use although many teaching associates raised concerns with their perceived lack of ability to impact on student learning through a decreased control of the marks assigned.

Overall, this work highlights the importance of consistent change over all year levels and all units/courses if truly significant outcomes are going to be measured from teaching interventions on a large-scale. Whilst altering individual activities or units/courses is a highly important process, true impact on student perceptions and their ability to articulate their skill development requires a large-scale and consistent change of teaching practices.

Lastly, assessment processes need to be considered and better linked to real-life or 'authentic' environments to better direct student perceptions of their development of transferable skills. Additional to this, large variation between different markers due to an overuse of personal academic judgement can be mitigated through the use of electronic marking processes such as Excel or online Learning Management Systems.

8.1 Implications for practice

Large-scale use of inquiry/problem/context/industry-based laboratory activities should be implemented in order to enhance student enjoyment and their perceptions of skill development, particularly with regards to transferable skills and an ability to undertake scientific investigations. Authentic assessment should be used wherever possible in order to significantly affect student perception of development of meaningful and relevant transferable skills. However, as these gains can be undermined if change is not consistent over all year levels, any pedagogic changes should be implemented over an entire programme of study wherever possible. However, it is important to note that these are only the perceptions of students and that the true benefits of inquiry/problem/context/industry-based laboratory activities may persist regardless of the findings stated here.

Lastly, whenever marking variability is an issue, academic judgement should be minimised in order to lead to more consistent marking. Automated marking techniques should be further explored to fully capitalise on this finding.

8.2 Implications for research

The generation and dissemination of the survey designed to evaluate the student's perceptions of individual laboratory activities provides an easily accessible tool for other researchers working in the same area. Furthermore, the data set provided during this study provides a strong comparison piece for any future work on similar interventions.

The framework generated to study the responses of staff and students when considering the aims of practical courses will allow other researches to use a similar methodology and analysis procedure to ensure reliable and valid comparisons between current and future studies. The use of the MLLI survey has also been expanded to contexts beyond the United States, increasing its validity beyond a single culture and location.

8.3 Future work

There are several avenues of future research still to be considered as a result of the findings in this thesis:

1) Would changing *all* laboratory activities through the lens of the TLL programme have a greater effect on the perceptions of students and teaching staff?
2) Does prior industry or research experience influence the perceptions of students towards the TLL generated activities?
3) Would revising the pre-laboratory activities to better match the TLL programme influence student enjoyment and their perceptions of the activities and their skill development?
4) Would more TLL focused training for the Teaching Associates result in greater changes to the perceptions of students and their levels of enjoyment?
5) Would a greater amount of authentic assessment further influence the perceptions of students and teaching staff?

References

Agustian, H. Y., & Seery, M. K. (2017). Reasserting the role of pre-laboratory activities in chemistry education: a proposed framework for their design. *Chemistry Education Research and Practice, 18*(4), 518-532. doi: 10.1039/C7RP00140A

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: an interactional analysis approach. *Assessment & Evaluation in Higher Education, 42*(2), 252-265.

Ajjawi, R., & Boud, D. (2018). Examining the nature and effects of feedback dialogue. *Assessment & Evaluation in Higher Education*, 1-14.

Avargil, S., Herscovitz, O., & Dori, Y. J. (2013). Challenges in the transition to large-scale reform in chemical education. *Thinking Skills and Creativity, 10*, 189-207. doi: http://dx.doi.org/10/1016/j.tsc.2013.07.008

Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and children, 46*(2), 26.

Barlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information technology, learning, and performance journal, 19*(1), 43.

Basey, J. M., Maines, A. P., Francis, C. D., & Melbourne, B. (2014). An Evaluation of Two Hands-On Lab Styles for Plant Biodiversity in Undergraduate Biology. *CBE-life sciences education, 13*(3), 493-503. doi: 10.1187/cbe.14-03-0062

Bell, A., Mladenovic, R., & Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education, 38*(7), 769-788. doi: 10.1080/02602938.2012.714738

Bell, R. C. (1980). Problems in improving the reliability of essay marks. *Assessment in Higher Education, 5*(3), 254-263. doi: 10.1080/0260293800050303

Bennett, J., & Lubben, F. (2006). Context-based Chemistry: The Salters approach. *International Journal of Science Education, 28*(9), 999-1015. doi: 10.1080/09500690600702496

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher education*, *32*(3), 347-364. doi: 10.1007/BF00138871

Bird, F. L., & Yucel, R. (2013). Improving marking reliability of scientific writing with the Developing Understanding of Assessment for Learning programme. *Assessment & Evaluation in Higher Education, 38*(5), 536-553. doi: 10.1080/02602938.2012.658155

Bloxham, S. (2009). Marking and moderation in the UK: false assumptions and wasted resources. *Assessment & Evaluation in Higher Education, 34*(2), 209-220. doi: 10.1080/02602930801955978

Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education, 41*(3), 466-481. doi: 10.1080/02602938.2015.1024607

Boud, D. J., Dunn, J., Kennedy, T., & Thorley, R. (1980). The Aims of Science Laboratory Courses: a Survey of Students, Graduates and Practising Scientists. *European Journal of Science Education, 2*(4), 415-428. doi: 10.1080/0140528800020408

Bowers, W. G. (1924). The importance of laboratory work as compared with text book work, etc., in the study of chemistry. *School Science and Mathematics, 24*(6), 606-613. doi: 10.1111/j.1949-8594.1924.tb04867.x

Bretz, S. L., Fay, M., Bruck, L. B., & Towns, M. H. (2013). What Faculty Interviews Reveal about Meaningful Learning in the Undergraduate Chemistry Laboratory. *Journal of Chemical Education, 90*(3), 281-288. doi: 10.1021/ed300384r

Bretz, S. L., Galloway, K. R., Orzel, J., & Gross, E. (2016). Faculty Goals, Inquiry, and Meaningful Learning in the Undergraduate Chemistry Laboratory *Technology and Assessment Strategies for Improving Student Learning in Chemistry* (Vol. 1235, pp. 101-115): American Chemical Society.

Bridges, P., Bourdillon, B., Collymore, D., Cooper, A., Fox, W., Haines, C., . . . Yorke, M. (1999). Discipline-related Marking Behaviour Using Percentages: a potential cause of inequity in assessment. *Assessment & Evaluation in Higher Education, 24*(3), 285-300. doi: 10.1080/0260293990240303

Broman, K., & Parchmann, I. (2014). Students' application of chemical concepts when solving chemistry problems in different contexts. *Chemistry Education Research and Practice, 15*(4), 516-529. doi: 10.1039/C4RP00051J

Bruck, A. D., & Towns, M. (2013). Development, Implementation, and Analysis of a National Survey of Faculty Goals for Undergraduate Chemistry Laboratory. *Journal of Chemical Education, 90*(6), 685-693. doi: 10.1021/ed300371n

Bruck, L. B., Towns, M., & Bretz, S. L. (2010). Faculty Perspectives of Undergraduate Chemistry Laboratory: Goals and Obstacles to Success. *Journal of Chemical Education, 87*(12), 1416-1424. doi: 10.1021/ed900002d

Buck, L. B., Bretz, S. L., & Towns, M. H. (2008, 2008 September-October). Characterizing the level of inquiry in the undergraduate laboratory: discrepancies abound in use of the word "inquiry." We propose a quantitative rubric to characterize inquiry in undergraduate laboratories. *Journal of College Science Teaching, 38,* 52.

Bulte, A. M. W., Westbroek, H. B., de Jong, O., & Pilot, A. (2006). A Research Approach to Designing Chemistry Education using Authentic Practices as Contexts. *International Journal of Science Education, 28*(9), 1063-1086. doi: 10.1080/09500690600702520

Byrne, C. (1980). Tutor marked assignments at the open university: a question of reliability. *Assessment in Higher Education, 5*(2), 150-167. doi: 10.1080/0260293800050203

Campbell, A. (2005). Application of ICT and rubrics to the assessment process where professional judgement is involved: the features of an e-marking tool. *Assessment & Evaluation in Higher Education, 30*(5), 529-537. doi: 10.1080/02602930500187055

Campbell, B., Lazonby, J., Millar, R., Nicolson, P., Ramsden, J., & Waddington, D. (1994). Science: The Salters' approach-a case study of the process of large-scale curriculum development. *Science Education, 78*(5), 415-447. doi: 10.1002/sce.3730780503

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*: Routledge.

Carnduff, J., & Reid, N. (2003). *Enhancing Undergraduate Chemistry Laboratories: Pre-laboratory and Post-laboratory Exercises*: Royal Society of Chemistry.

Chatterjee, S., Williamson, V. M., McCann, K., & Peck, M. L. (2009). Surveying Students' Attitudes and Perceptions toward Guided-Inquiry and Open-Inquiry Laboratories. *Journal of Chemical Education, 86*(12), 1427. doi: 10.1021/ed086p1427

Collier, M. (1986). A specific investigation of relative performance of examination markers. *Assessment & Evaluation in Higher Education, 11*(2), 130-137. doi: 10.1080/0260293860110204

Creswell, J. (2009). *Research design. Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: SAGE Publications.

Crisp, B. R. (2007). Is it worth the effort? How feedback influences students' subsequent submission of assessable work. *Assessment & Evaluation in Higher Education, 32*(5), 571-581. doi: 10.1080/02602930601116912

Cummins, R. H., Green, W. J., & Elliott, C. (2004). "Prompted" Inquiry-Based Learning in the Introductory Chemistry Laboratory. *Journal of Chemical Education, 81*(2), 239. doi: 10.1021/ed081p239

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and teacher education, 16*(5-6), 523-545.

DeKorver, B. K., & Towns, M. H. (2015). General Chemistry Students' Goals for Chemistry Laboratory Coursework. *Journal of Chemical Education, 92*(12), 2031-2037. doi: 10.1021/acs.jchemed.5b00463

Diller, K. R., & Phelps, S. F. (2008). Learning outcomes, portfolios, and rubrics, oh my! Authentic assessment of an information literacy program. *portal: Libraries and the Academy, 8*(1), 75-89.

Domin, D. S. (2007). Students' perceptions of when conceptual development occurs during laboratory instruction. *Chemistry Education Research and Practice, 8*(2), 140-152. doi: 10.1039/B6RP90027E

Donner, R. S., & Bickley, H. (1993). Problem-based learning in American medical education: an overview. *Bulletin of the Medical Library Association, 81*(3), 294-298.

Dunnivant, F. M. (2002). Analytical Problems Associated with the Analysis of Metals in a Simulated Hazardous Waste. *Journal of Chemical Education, 79*(6), 718. doi: 10.1021/ed079p718

Edwards, D. (1979). A study of the reliability of tutor marked assignments at the open university. *Assessment in Higher Education, 5*(1), 16-44. doi: 10.1080/0260293790050102

Elliott, M. J., Stewart, K. K., & Lagowski, J. J. (2008). The Role of the Laboratory in Chemistry Instruction. *Journal of Chemical Education, 85*(1), 145. doi: 10.1021/ed085p145

Erhart, S. E., McCarrick, R. M., Lorigan, G. A., & Yezierski, E. J. (2015). Citrus Quality Control: An NMR/MRI Problem-Based Experiment. *Journal of Chemical Education*. doi: 10.1021/acs.jchemed.5b00251

Esson, J. M., Scott, R., & Hayes, C. J. (2018). Chemistry and Art: Removal of Graffiti Ink from Paints Grounded in a Real-Life Scenario. *Journal of Chemical Education*. doi: 10.1021/acs.jchemed.7b00536

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*(3), 1-13.

Galloway, K. R., & Bretz, S. L. (2015a). Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory. *Journal of Chemical Education, 92*(7), 1149-1158. doi: 10.1021/ed500881y

Galloway, K. R., & Bretz, S. L. (2015b). Measuring Meaningful Learning in the Undergraduate Chemistry Laboratory: A National, Cross-Sectional Study. *Journal of Chemical Education, 92*(12), 2006-2018. doi: 10.1021/acs.jchemed.5b00538

Galloway, K. R., Malakpa, Z., & Bretz, S. L. (2016). Investigating Affective Experiences in the Undergraduate Chemistry Laboratory: Students' Perceptions of Control and Responsibility. *Journal of Chemical Education, 93*(2), 227-238. doi: 10.1021/acs.jchemed.5b00737

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research: Competencies for analysis and applications (8th ed.)*. Upper Saddle River: Pearson Merril Prentice Hall.

George-Williams, S. R., Carroll, M.-R., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2019). Curtailing marking variation and enhancing feedback in large-scale first year undergraduate chemistry courses through reducing academic judgement: A case study. *Assessment & Evaluation in Higher Education, Accepted, awaiting publication.* doi: 10.1080/02602938.2018.1545897

George-Williams, S. R., Karis, D., Ziebell, A. L., Kitson, R. R. A., Coppo, P., Schmid, S., . . . Overton, T. L. (2019). Investigating student and staff perceptions of students' experiences in teaching laboratories through the lens of meaningful learning. *Chemistry Education Research and Practice*. doi: 10.1039/C8RP00188J

George-Williams, S. R., Soo, J. T., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2018). Inquiry and industry inspired laboratories: The impact on students' perceptions of skill development and engagements. *Chemistry Education Research and Practice*. doi: 10.1039/C7RP00233E

George-Williams, S. R., Ziebell, A. L., Kitson, R. R. A., Coppo, P., Thompson, C. D., & Overton, T. L. (2018). 'What do you think the aims of doing a practical chemistry course are?' A comparison of the views of students and teaching staff across three universities. *Chemistry Education Research and Practice, 19*(2), 463-473. doi: 10.1039/C7RP00177K

George-Williams, S. R., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2018a). Electronic Waste – A case study in attempting to reduce cognitive overload whilst increasing context and student control in an undergraduate laboratory. *Monash Education Academy Digest, Submitted, awaiting review*.

George-Williams, S. R., Ziebell, A. L., Thompson, C. D., & Overton, T. L. (2018b). Enhancing inquiry in the teaching laboratory; a consecutive expository/inquiry-based laboratory exercise. *Monash Education Academy Digest, 1*(1).

Gilbert, J. K. (2006). On the Nature of “Context” in Chemical Education. *International Journal of Science Education, 28*(9), 957-976. doi: 10.1080/09500690600702470

Given, L. M. (2008). *The Sage encyclopedia of qualitative research methods*: Sage Publications.

Good, H. G. (1936). On the early history of Liebig's laboratory. *Journal of Chemical Education, 13*(12), 557. doi: 10.1021/ed013p557

Graham, J. G. U. (1992). Fisher's Exact Test. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 155*(3), 395-402. doi: 10.2307/2982890

Gregory, S.-J., & Trapani, G. D. (2012). A Blended Learning Approach to Laboratory Preparation. *International Journal of Innovation in Science and Mathematics Education, 20*(1), 56-70.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Educational technology research and development, 52*(3), 67-86.

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*: Advanced Analytics, LLC.

Hall, C. G. W., & Daglish, N. D. (1982). Length and quality: an exploratory study of inter-marker reliability. *Assessment & Evaluation in Higher Education, 7*(2), 186-191. doi: 10.1080/0260293820070209

Harris, H. H. (1997). The Chemistry Classroom: Formulas for Successful Teaching (Herron, J. Dudley). *Journal of Chemical Education, 74*(10), 1167. doi: 10.1021/ed074p1167.3

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*: Routledge.

Hawkes, S. J. (2004). Chemistry Is Not a Laboratory Science. *Journal of Chemical Education, 81*(9), 1257. doi: 10.1021/ed081p1257

Healey, M., & Jenkins, A. (2009). *Developing undergraduate research and inquiry* (p. 152). York: Higher Education Academy.

Heinrich, E., Milne, J., Ramsay, A., & Morrison, D. (2009). Recommendations for the use of e-tools for improvements around assignment marking quality. *Assessment & Evaluation in Higher Education, 34*(4), 469-479. doi: 10.1080/02602930802071122

Herron, M. D. (1971). The Nature of Scientific Enquiry. *The School Review, 79*(2), 171-212. doi: 10.1086/442968

Hicks, R. W., & Bevsek, H. M. (2012). Utilizing Problem-Based Learning in Qualitative Analysis Lab Experiments. *Journal of Chemical Education, 89*(2), 254-257. doi: 10.1021/ed1001202

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational psychology review, 16*(3), 235-266. doi: 10.1023/B:EDPR.0000034022.16470.f3

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review, 70*(256), 33-40.

Hodson, D., & Hodson, J. (1998). From constructivism to social constructivism: A Vygotskian perspective on teaching and learning science. *School Science Review*, 79(289), 33-41.

Holmes, N. G., & Wieman, C. E. (2018). Introductory physics labs: We can do better. *Physics Today, 71*(1), 38-45. doi: 10.1063/pt.3.3816

Huck, L. A., & Leigh, W. J. (2010). A Better Sunscreen: Structural Effects on Spectral Properties. *Journal of Chemical Education, 87*(12), 1384-1387. doi: 10.1021/ed1004867

Hulien, M. L., Lekse, J. W., Rosmus, K. A., Devlin, K. P., Glenn, J. R., Wisneski, S. D., . . . Aitken, J. A. (2015). An Inquiry-Based Project Focused on the X-ray Powder Diffraction Analysis of Common Household Solids. *Journal of Chemical Education, 92*(12), 2152-2156. doi: 10.1021/acs.jchemed.5b00008

Hunter, K., & Docherty, P. (2011). Reducing variation in the assessment of student writing. *Assessment & Evaluation in Higher Education, 36*(1), 109-124. doi: 10.1080/02602930903215842

ICSU-CTS. (1979). *Learning strategies in university science*. Cardiff: University College Cardiff Press.

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of computer assisted learning, 7*(2), 75-83.

Johnstone, A. H. (1997). Chemistry Teaching - Science or Alchemy? 1996 Brasted Lecture. *Journal of Chemical Education, 74*(3), 262. doi: 10.1021/ed074p262

Johnstone, A. H., & Al-Shuaili, A. (2001). Learning in the laboratory; some thoughts from the literature. *Univ. Chem. Educ., 5*(1), 41-50.

Kirschner, P. A., & Meester, M. A. M. (1988). The Laboratory in Higher Science Education: Problems, Premises and Objectives. *Higher Education, 17*(1), 81-98. doi: 10.1007/BF00130901

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi: 10.1207/s15326985ep4102_1

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *International Journal of Computer-Supported Collaborative Learning, 13*(2), 213-233. doi: 10.1007/s11412-018-9277-y

Koehler, B. P., & Orvis, J. N. (2003). Internet-Based Prelaboratory Tutorials and Computer-Based Probes in General Chemistry. *Journal of Chemical Education, 80*(6), 606. doi: 10.1021/ed080p606

Kulevich, S. E., Herrick, R. S., & Mills, K. V. (2014). A Discovery Chemistry Experiment on Buffers. *Journal of Chemical Education, 91*(8), 1207-1211. doi: 10.1021/ed400377a

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681-718.

Leal Filho, W., & Pace, P. (2016). Teaching Education for Sustainable Development at University Level: Springer.

Lennon, D., Freer, A. A., Winfield, J. M., Landon, P., & Reid, N. (2002). An undergraduate teaching initiative to demonstrate the complexity and range of issues typically encountered in modern industrial chemistry. *Green Chemistry, 4*(3), 181-187. doi: 10.1039/B202541H

Letton, K. M. (1987). *A study of the factors influencing the efficiency of learning in a undergraduate chemistry laboratory,.* (M Phil), Jordanhill College of Education,, Glasgow, Scotland.

Linn, M. C. (1997). The Role of the Laboratory in Science Learning. *The Elementary School Journal, 97*(4), 401-417. doi: 10.1086/461873

Lowry, R. (2014). Concepts and applications of inferential statistics.

Lunt, T., & Curran, J. (2010). 'Are you listening please?' The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education, 35*(7), 759-769. doi: 10.1080/02602930902977772

Lynch, P. P., & Ndyetabura, V. L. (1983). Practical work in schools: An examination of teachers' stated aims and the influence of practical work according to students. *Journal of Research in Science Teaching, 20*(7), 663-671. doi: 10.1002/tea.3660200707

Mahaffy, P. (2004). THE FUTURE SHAPE OF CHEMISTRY EDUCATION. *Chemistry Education Research and Practice, 5*(3), 229-245. doi: 10.1039/B4RP90026J

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 43-52. doi: 10.1207/S15326985EP3801_6

Mc Ilrath, S. P., Robertson, N. J., & Kuchta, R. J. (2012). Bustin' Bunnies: An Adaptable Inquiry-Based Approach Introducing Molecular Weight and Polymer Properties. *Journal of Chemical Education, 89*(7), 928-932. doi: 10.1021/ed2004615

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica, 22*(3), 276-282.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica, 23*(2), 143-149.

McKelvy, G. M. (2000). Preparing for the Chemistry Laboratory: An Internet Presentation and Assessment Tool. *University Chemistry Education*(4), 46-49.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. doi: 10.1037/h0043158

Miller, G. L. (1992). *The history of science: an annotated bibliography*. Pasadena, Calif.: Salem Press.

Morse, J. M. (2003). Principles of mixed methods and multimethod research design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research.* (pp. 189-208). Thousand Oaks, CA: Sage publications, Inc.

Nielsen, S. E., Scaffidi, J. P., & Yezierski, E. J. (2014). Detecting Art Forgeries: A Problem-Based Raman Spectroscopy Lab. *Journal of Chemical Education, 91*(3), 446-450. doi: 10.1021/ed400319k

Norton, A., & Cakitaki, B. (2016). Mapping Australian higher education 2016. *Grattan Institute, 7*.

OECD. (2002). *Responding to Student Expectations*: OECD Publishing.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist, 38*(1), 1-4. doi: 10.1207/S15326985EP3801_1

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational psychology review, 6*(4), 351-371.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors, 35*(4), 737-743. doi: 10.1177/001872089303500412

Pea, R. D. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *Journal of the Learning Sciences, 13*(3), 423-451. doi: 10.1207/s15327809jls1303_6

Phillips, M. (1981). Early history of physics laboratories for students at the college level. *American Journal of Physics, 49*(6), 522-527. doi: 10.1119/1.12664

Pilcher, L. A., Riley, D. L., Mathabathe, K. C., & Potgieter, M. (2015). An inquiry-based practical curriculum for organic chemistry as preparation for industry and postgraduate research. *South African Journal of Chemistry, 68*, 236-244.

Pilot, A., & Bulte, A. M. W. (2006). Why Do You "Need to Know"? Context-based education. *International Journal of Science Education, 28*(9), 953-956. doi: 10.1080/09500690600702462

Rajapaksha, S. M., Samarasekara, D., Brown, J. C., Howard, L., Gerken, K., Archer, T., . . . Mlsna, D. (2018). Determination of Xylitol in Sugar-Free Gum by GC–MS with Direct Aqueous Injection: A Laboratory Experiment for Chemistry Students. *Journal of Chemical Education, 95*(11), 2017-2022. doi: 10.1021/acs.jchemed.8b00424

Reid, N., & Shah, I. (2007). The role of laboratory work in university chemistry. *Chemistry Education Research and Practice, 8*(2), 172-185. doi: 10.1039/B5RP90026C

Russell, C. B. (2008). *Development and Evaluation of a Research-Based Undergraduate Laboratory Curriculum.* (PhD), Purdue University, West Lafayette, Indiana.

Russell, C. B., & Weaver, G. (2008). Student perceptions of the purpose and function of the laboratory in science: A grounded theory study. *International Journal for the scholarship of teaching and learning, 2*(2), 9. doi: 10.20429/ijsotl.2008.020209

Saldaña, J. (2015). *The coding manual for qualitative researchers*: Sage.

Samarasekara, D., Hill, C., & Mlsna, D. (2018). Analysis and Identification of Major Organic Acids in Wine and Fruit Juices by Paper Chromatography. *Journal of Chemical Education, 95*(9), 1621-1625. doi: 10.1021/acs.jchemed.8b00129

Sandi-Urena, S., Cooper, M., & Stevens, R. (2012). Effect of Cooperative Problem-Based Lab Instruction on Metacognition and Problem-Solving Skills. *Journal of Chemical Education, 89*(6), 700-706. doi: 10.1021/ed1011844

Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension, 37*(2), 1-5.

Sarkar, M., Overton, T., Thompson, C., & Rayner, G. (2016). Graduate Employability: Views of Recent Science Graduates and Employers. *International Journal of Innovation in Science and Mathematics Education., 24*(3), 31-48.

Schepmann, H. G., & Mynderse, M. (2010). Ring-Closing Metathesis: An Advanced Guided-Inquiry Experiment for the Organic Laboratory. *Journal of Chemical Education, 87*(7), 721-723. doi: 10.1021/ed100248a

Schmid, S., & Yeung, A. (2005). The influence of a pre-laboratory work module on student performance in the first year chemistry laboratory. *Research and Development in Higher Education, 28*, 471-479.

Schmidt, H. G., Rotgans, J. I., & Yew, E. H. J. (2011). The process of problem-based learning: what works and why. *Medical Education, 45*(8), 792-806. doi: 10.1111/j.1365-2923.2011.04035.x

Seery, M. K., Agustian, H. Y., & Zhang, X. (2018). A Framework for Learning in the Chemistry Laboratory. *Israel Journal of Chemistry, 0*(0). doi: doi:10.1002/ijch.201800093

Shah, I. (2007). *Making University Laboratory Work in Chemistry More Effective.* (M. Phil.), University of Glasgow, Scotland.

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*: crc Press.

Shultz, G. V., & Li, Y. (2016). Student Development of Information Literacy Skills during Problem-Based Organic Chemistry Laboratory Experiments. *Journal of Chemical Education, 93*(3), 413-422. doi: 10.1021/acs.jchemed.5b00523

Shymansky, J. A., & Penick, J. E. (1979). Use of systematic observations to improve college science laboratory instruction. *Science Education, 63*(2), 195-203. doi: 10.1002/sce.3730630207

Smith, E., & Coombe, K. (2006). Quality and qualms in the marking of university assignments by sessional staff: An exploratory study. *Higher Education, 51*(1), 45-69.

Sopina, E., & McNeill, R. (2015). Investigating the relationship between quality, format and delivery of feedback for written assignments in higher education. *Assessment & Evaluation in Higher Education, 40*(5), 666-680. doi: 10.1080/02602938.2014.945072

Stout, R. P. (2016). CO2 Investigations: An Open Inquiry Experiment for General Chemistry. *Journal of Chemical Education, 93*(4), 713-717. doi: 10.1021/ed5006932

Sundberg, M. D., & Armstrong, J. E. (1993). The Status of Laboratory Instruction for Introductory Biology in U.S. Universities. *The American Biology Teacher, 55*(3), 144-146. doi: 10.2307/4449610

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation, 27*(2), 237-246.

Torp, L., & Sage, S. (1998). *Problems as possibilities: Problem-based learning for K-12 education*: ASCD.

Tosun, C., & Taskesenligil, Y. (2013). The effect of problem-based learning on undergraduate students' learning about solutions and their physical properties and scientific processing skills. *Chemistry Education Research and Practice, 14*(1), 36-50.

Vgotsky, L. (1978). Mind in society: The development of higher mental processes: Cambridge, MA: Harvard University Press.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med, 37*(5), 360-363.

Watt, S., Simpson, C., McKillop, C., & Nunn, V. (2002). Electronic Course Surveys: Does automating feedback and reporting give better results? *Assessment & Evaluation in Higher Education, 27*(4), 325-337. doi: 10.1080/0260293022000001346

Wilcox, B. R., & Lewandowski, H. (2017). Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students' beliefs about experimental physics. *Physical Review Physics Education Research, 13*(1), 010108.

Wilkinson, J., & Ward, M. (1997). A comparative study of students' and their teacher's perceptions of laboratory work in secondary schools. *Research in Science Education, 27*(4), 599-610. doi: 10.1007/bf02461483

Woolnough, B. E., & Allsop, T. (1985). *Practical work in science*: Cambridge University Press.

Zahra, D., Robinson, I., Roberts, M., Coombes, L., Cockerill, J., & Burr, S. (2017). Rigour in moderation processes is more important than the choice of method. *Assessment & Evaluation in Higher Education, 42*(7), 1159-1167. doi: 10.1080/02602938.2016.1236183

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of modern applied statistical methods, 6*(1), 4.

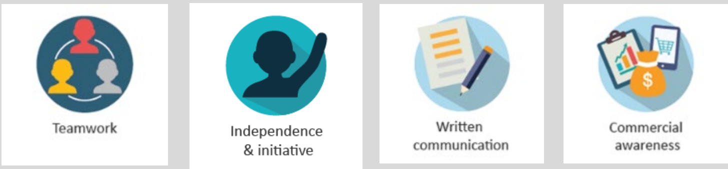
Appendix 1 – Making effective sunscreen ingredients using Claisen-Schmidt condensation

Exercise 8: Making Effective Sunscreen Ingredients using Claisen-Schmidt Condensation

Learning outcomes:

1. Perform an aldol condensation, including use of a dropping funnel and amendment of a related method to suit your purpose.
2. Use structure-function relationship between conjugation and light absorbance to assess a sunscreen reagents effectiveness.
3. Perform a melting point analysis, UV spectroscopy, 1H NMR spectral analysis and calculate % yield.

Employability skills:

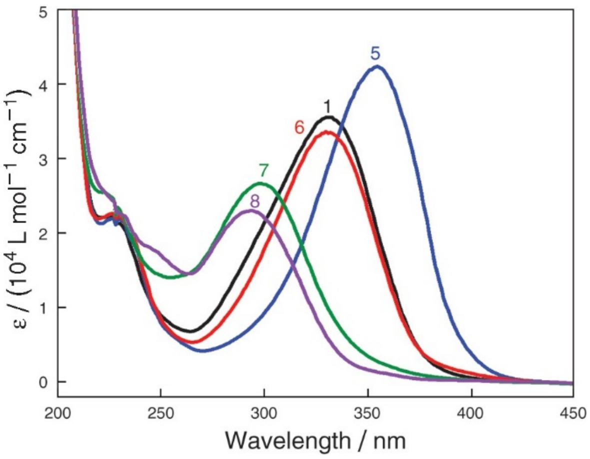


8.1 Safety

A Hazard Identification and Risk Assessment sheet must be completed ***prior*** to the laboratory session. The related Safety Data Sheets (SDSs) are on Moodle. Please ensure your risk assessment is signed by yourself and your teaching associate before you start any laboratory work.

8.2 Introduction

Sunscreens are a multimillion dollar industry in Australia. Sunscreen protects the skin by absorbing or reflecting UV irradiation so that it doesn't reach the skin and cause damage. The organic compounds generally absorb UV irradiation while mineral ingredients (TiO_2 and ZnO) reflect the UV irradiation. High-intensity UV rays are absorbed by excitation of the organic molecule to a higher energy state. The extensive conjugation in these molecules effectively distributes the absorbed energy across the molecule. The UV profile of dibenzalacetone (**DBA-1**) and some related compounds can be seen in figure 8.1. In this case dibenzylidene cyclopentanone (**DBA-5**) is the best sunscreen reagent based on UV profile because it protects highest into the damaging UVA range (315-400 nm) and it also has high absorbtivity in that range. This is due to the high level of conjugation in these molecules which distributes the energy across the molecule allowing the energy to be absorbed.



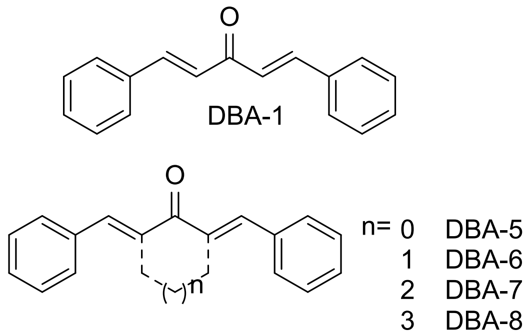
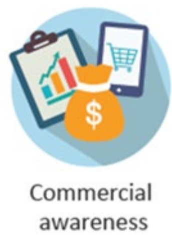


Figure 8.4: UV absorbtivity of **DBA-1**, **5**, **6**, **7**, and 8 recorded in 95% ethanol (left), Dibenzylidene structures important in this investigation (above).

8.3 Scenario



Monash Consulting have been engaged by Rationale (a personal care company) to assess an idea for a potential start-up company. The main focus of the company would be the manufacture and marketing of a new-and-improved sunscreen lotion. Rationale Chemist Dr Sarah Jane has suggested modifying the structure of a molecule already used in some commercial sunscreens, dibenzylideneacetone (**DBA-1**), and instead looking at cyclic versions listed in figure 8.1. The cyclic ketones are of interest because they are not always covered in the patents that cover the dibenzalacetone (e.g. US 7014842 B2). Rationale or their start up may therefore be able to patent the use of the ingredient in sunscreens.

Dr Jane reasoned that the introduction of rings into the central core of the molecule would strengthen the rigidity of the system and force planarity over the entire structure. This planarity would in turn increase the conjugation over the molecule and potentially increase its ability to absorb ultraviolet radiation in the particularly deleterious UV-A region (315-400 nm). She was right! Rationale or their start up may therefore be able to patent the use of the ingredient in sunscreens**.**

Figure 8.1 shows the absorbance of UV light of the 5 **DBA** derivatives. **DBA-5**, the structure containing a 5-membered carbon ring in the centre, showed both the greatest amount of absorptivity at its maxima (~4.2 x 10^4 L mol^{-1} cm^{-1}) and covered the entire UV absorption range. Unfortunately, it came at a cost. The cyclic ketones are slow to react and need heating which would be very expensive on a large scale.

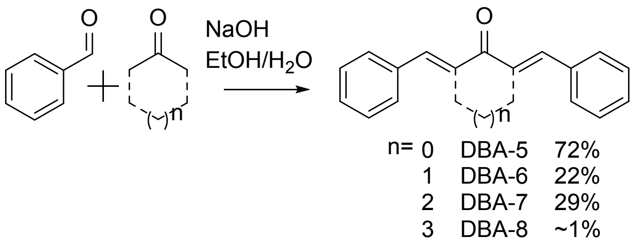
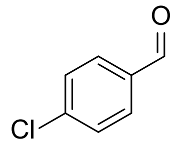


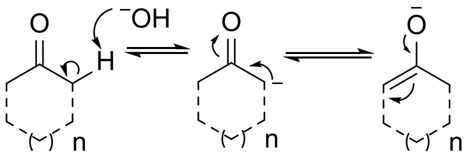
Figure 8.5: The synthesis of the **DBA** derivatives through an aldol condensation of benzaldehyde with a variety of cyclic ketones.

At this time, Rationale are currently unable to continue work on this particular project and have requested that Monash Consulting look into potential solutions to this issue. As seen in the pre-laboratory quiz, logic would dictate using the chlorinated derivative of the benzaldehyde to potentially allow for formation of the desired species at, or below, room temperature. However, we need data:



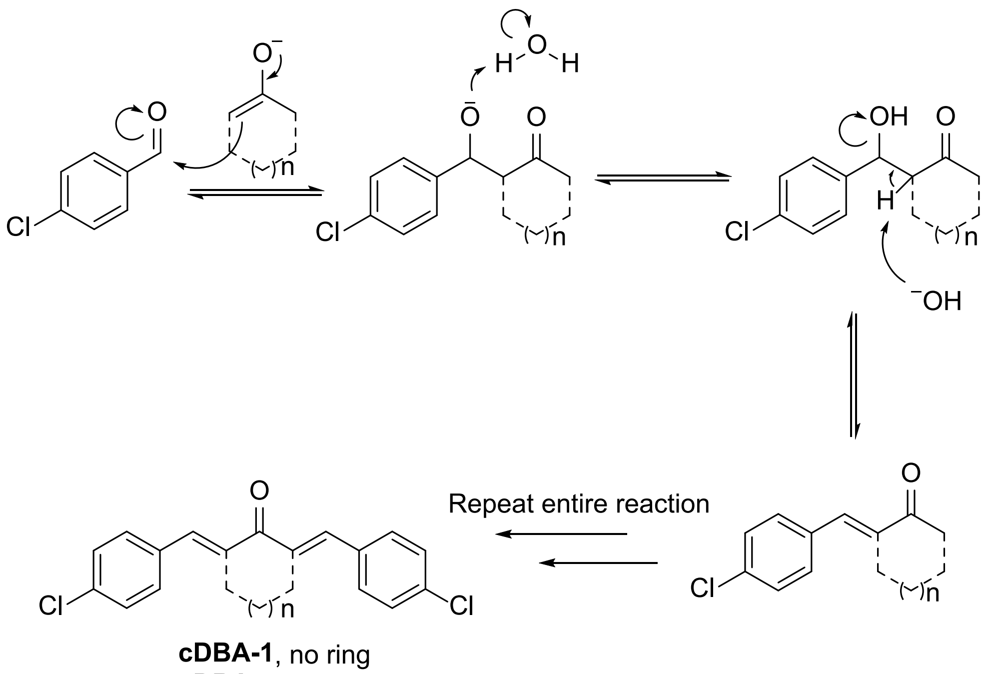
1) The solubility of the starting material (4-chlorobenzaldehyde) in the reaction solvents ethanol and water.
2) The yields obtained after 30 minutes at room temperature.
3) The purity of the product(s) obtained.
4) The effect of the chlorine substituent on the maximum absorbance in the UV-A region (315-400 nm) region.
5) The effect of the chlorine substituent on the range of absorbance in the UV spectrum.

Your job, as a group of three, is to generate the three chlorinated derivatives of **DBA-1**, **5** and **6** and to measure the UV –absorbance of the molecules. Rationale suggest that **DBA-8** be excluded from future tests due to difficulty of synthesis and purification. Furthermore, chlorinated **DBA-7** does not form at room temperature and, as such, is not a part of this study.



Acetone, no ring
Cyclopentanone, n=0
Cycohexanone, n=1

Scheme 1 Mechanism for enolate formation for ketones used in experiment 8.



cDBA-1, no ring
cDBA-5, n=0
cDBA-6, n=1

Scheme 2 Mechanism for chlorinated dibenzylidene formation of **cDBA-1**,**5** and **6**.

8.4 Experimental

a) Synthesis of bischloro dibenzylidene ketones.

Teamwork

Below is the original synthesis of the non-chlorinated derivatives Dr Jane has published in the research literature. You can use this published procedure as the basis of your method but you must change the amounts of reagents as per the pre-laboratory exercise because you are using different reagents. A photo of the glassware set up is provided in figure 8.3.

Synthesis of dibenzalacetone: A solution of sodium hydroxide (0.625 g), water (7.5 mL) and ethanol (7.5 mL) was generated in a reaction vessel suspended above a water bath. A solution of benzaldehyde (0.625 mL, 6.25 mmol), acetone (0.225 mL, 3.125 mmol) and ethanol (2.5 mL) was prepared and added dropwise whilst the temperature (20-25°C) was maintained with a water bath. After thirty minutes, the yellow reaction mixture was filtered and washed with water (2 x 10 mL) followed by ice-cold methanol (2 x 10 mL). The product was left to dry *via* suction for 10 minutes.

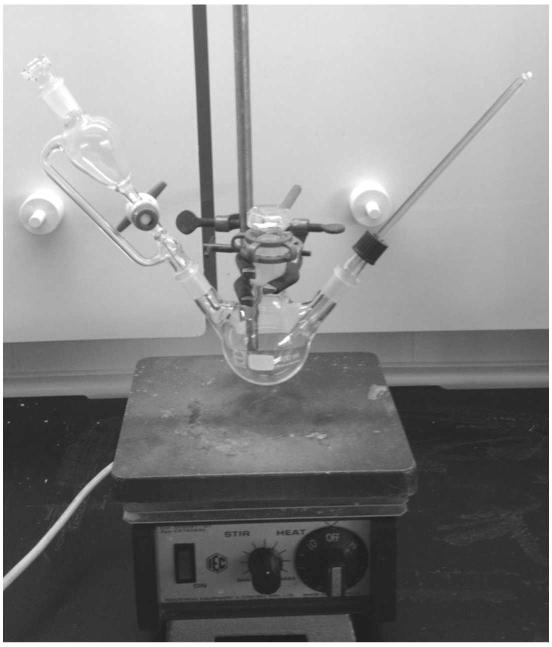


Figure 8.3: Experimental set up for air sensitive nickel alkene isomerisation

Starting your reaction:

1. Use the number of moles of the ketone and benzaldehyde to calculate the mass of each of the reagents you will use (remember n = mass/Mr). This is from the prelab, fill in the spaces below.
2. Be sure to make all transfers in the fume hood, including filtration. Vials can be filled in the fumehood and then take out to weigh with the lid on. Have the reaction running outside the fumehood.
3. Maintain a stable temperature. The reaction is exothermic (produces heat) and it is **important** to maintain the temperature at <25 °C for a **good yield.**
4. Think about what colour your product will be and why.

Approximately ___________ g of chlorobenzaldehyde needed.
Approximately ___________ g of chosen ketone needed.

b) Isolation and analysis of your product:

5. Filter (Hirsch) and wash it thoroughly with water to remove alkali. Keep a small portion of this for your analysis.

6. Dry the product *via* suction.
7. Use the appropriate Standard Operating Procedure (SOP) to establish purity through obtaining a melting point, and ^{1}H NMR spectrum.

c) UV analysis:

Make a 10 mL 2 x 10^{-5} M solution of your compound through the following steps:

1. Accurately weigh approximately 5-10 mg of your compound and add to a 25 mL volumetric flask and make to the mark with chloroform. Calculate the concentration of this solution.
2. Transfer the amount of solution required to make a final concentration (C_2) of 2 x 10^{-5} M to a 10 mL volumetric flask (using $C_1V_1 = C_2V_2$) and make to the mark with chloroform.
3. Run a UV-scan from 315-400 nm and compare to chloroform and the other compounds made by the rest of your group.

8.5 Report: Executive summary

Produce a full report for this experiment. Your report should include the following sections: Introductions & Aims, Method (refer to manual and indicate any changes you made), Results, Discussion, Conclusions, References. See guidance on Moodle.

In the workplace, experimental results are often delivered to more senior members of staff in the form of an executive summary. An executive summary gives these often time-poor individuals the key points of information about the outcome of the experiment without overloading them with unnecessary detail, such as every individual observed signal in your ^{1}H NMR spectrum. An *executive summary* provides a quick overview or synopsis of a longer report, summarising the essential parts. It outlines the purpose of the report, the methods used to conduct the research, key findings and recommendations. They are often presented in a structured way, such as dot point lists.

As part of the assessment for this experiment you are to produce a **one page** executive summary addressing the five points Rationale requested Monash Consulting to investigate. Extra relevant data (such as assigned NMR or UV spectra) should be added as an appendix of no more than 6 pages. Include the results of your whole group (3 chlorinated derivatives). Recall that Dr Jane is a trained chemist and published the original synthesis which means that you don't have to explain your methodology in great detail. Ensure you discuss the yields, ease of synthesis, the ^{1}H NMR spectrum, TLC, melting point and, most importantly, the UV results. Discuss the UV results and whether they follow the pattern you would expect and how you see that borne out. If your results do not fit the pattern seen in figure 8.1, then discuss why that might be.

You will be marked on format and presentation as well as the quality of your science so it must be typed up using standard word processing software. Include a brief introduction of your task (NOT the overall science), summary of key points, and a conclusion with recommendations. Remember, this summary is being submitted to a client and should therefore appear professional. Submit with the pro forma cover sheet.

Total length = 2.5 – 3.5 pages

- Must utilise the writing guide in the student handbook.
- Must be word processed (11 or 12 pt font) and suitably formatted.
- Marks deducted if beyond 4 pages in length or if word limits exceeded by 10% (data, spectra and references are not included in the length limits).

Section	**To get full marks …**
Introduction and Aim (2 marks) 1-2 paragraphs 200-250 words max.	3-4 good introductory/background points, including theory and real-world context. Introductory points backed up with references. Clear aims/goals and means provided.
Method Marks deducted if missing.	You must reference the method you used (e.g. as per CHM2911 lab manual, page XX-XX) and list any changes made throughout the experiment. Use past tense.
Results (5 marks) Half to one page	Results presented in a professional (e.g. no screenshots), easy to interpret manner with visually appealing formatting. Only relevant results shown with additional data placed in an appendix. All raw data (e.g. all absorbance values, all titres etc.), including relevant spectra, calculations and scanned lab notes present in the appendix. All results matches with range of expected values and all calculations are correctly performed.
Discussion (6 marks) Half to one page 300-350 words max.	Results fully relate back to original experimental aims. Most (or all) scientific statements use experimental data to justify claims. Most (or all) important results are reported and explained. Multiple errors/factors are raised and are scientific/experimental in nature. If human errors were made (e.g. a spill occurred), they are raised in a specific manner (i.e. the exact issue is noted). The discussion is written in the past tense, uses full sentences and is in the passive voice.
Conclusion (2 marks) 1-2 paragraphs 150-200 words max.	Relates back to aim. Most (if not all) scientific statements use experimental data to justify claims. Most (if not all) important results are covered.
Executive summary (5 marks) ~ Half page 250-300 words max.	Concise, not overly detailed. Professional presentation. Clear intro and conclusion, with irrelevant information not given. Only relevant and easy to interpret results present.

Appendix 2– Investigation into the Efficacy of a Digestive Enzyme (Beano™)

Experiment 4: Investigation into the Efficacy of a Digestive Enzyme

Learning Outcomes:

1. Calibrate a commercially available glucometer to monitor glucose production.
2. Investigate the action of a commercially available enzyme on the extract of split peas under specified conditions.
3. Identify potential improvements to the methodology.
4. Determine the effect of temperature on the enzyme activity.
5. Investigate the impact of changing conditions on the α-galactosidase performance.

Introduction

Not all carbohydrates that humans consume can be broken down by the action of acids and enzymes in the gut. This is the case with legumes and many cruciferous vegetables like cauliflower and cabbage, and can also be true for tofu. Beans for example contain raffinose family oligomers (verbascose, raffinose and stachyose) which remain in the digestive tract as they are too large to cross the bowel wall. These sugars become an energy source for intestinal tract microbes causing symptoms such as flatulence, diarrhoea and bloating. Although popularity varies regionally, digestive aids that combat these issues can be bought. These are generally over the counter tablets that contain enzymes which break down short chain sugars into their basic monomers (e.g. galactose, glucose and fructose.) and claim to relieve digestive problems (Figure 1). One of those products is Beano™ containing an α-galactosidase. An α-galactosidase works on raffinose family sugars as shown in figure 1.

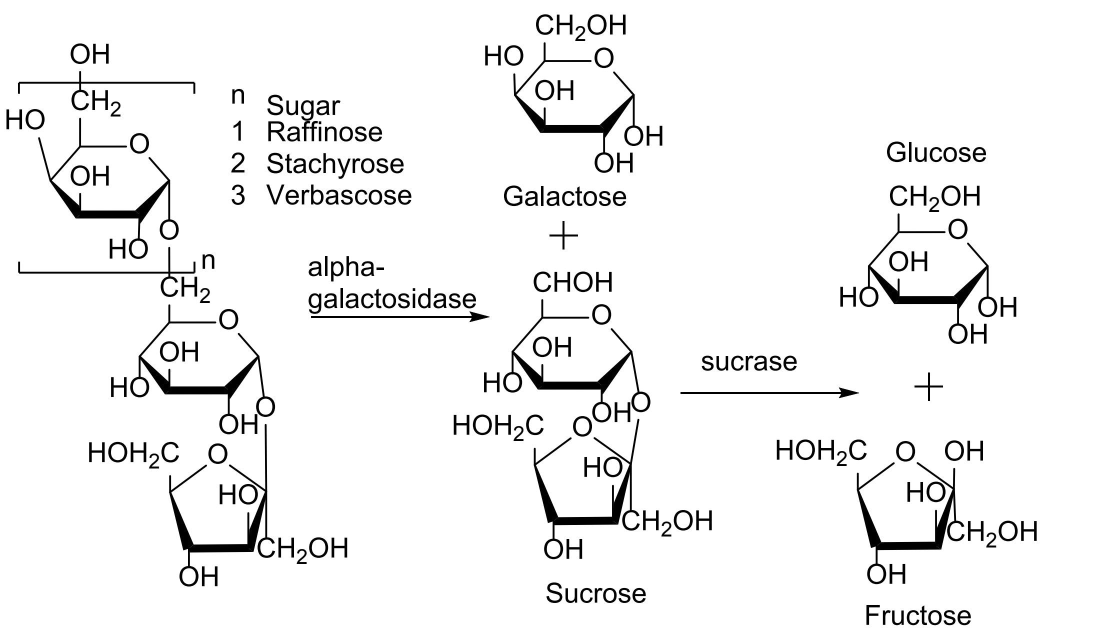


Figure 1: The enzymatic production of oligomers from their polymeric starting materials.

Over the next two weeks, you will be monitoring and studying this process through the use of a readily available glucometer (Figure 2), a small electronic device designed to detect the amount of glucose in human blood. This glucometer works due to the presence of glucose oxidase (GOD) in the tip of the strip which oxidises glucose exclusively (NB: galactose and fructose will not react due to the different stereochemistry). The GOD is reduced and regenerated by phenazine methosulfate which in turn will get reduced. The phenazine methosulfate in turn transfers two electrons to the electrode. The electron transfer produces an electrical signal that is detected and used to determine the concentration of glucose in the solution. As the reaction in Figure 1 proceeds, glucose is formed and its production can be used to monitor the rate at which the oligomers are being broken down into their monomers. There are many different variables that could affect this rate (such as temperature, pH, sugar source or dissolved metals) and this two-week practical will give you a chance to design an investigation of your own choosing in a group of three students.

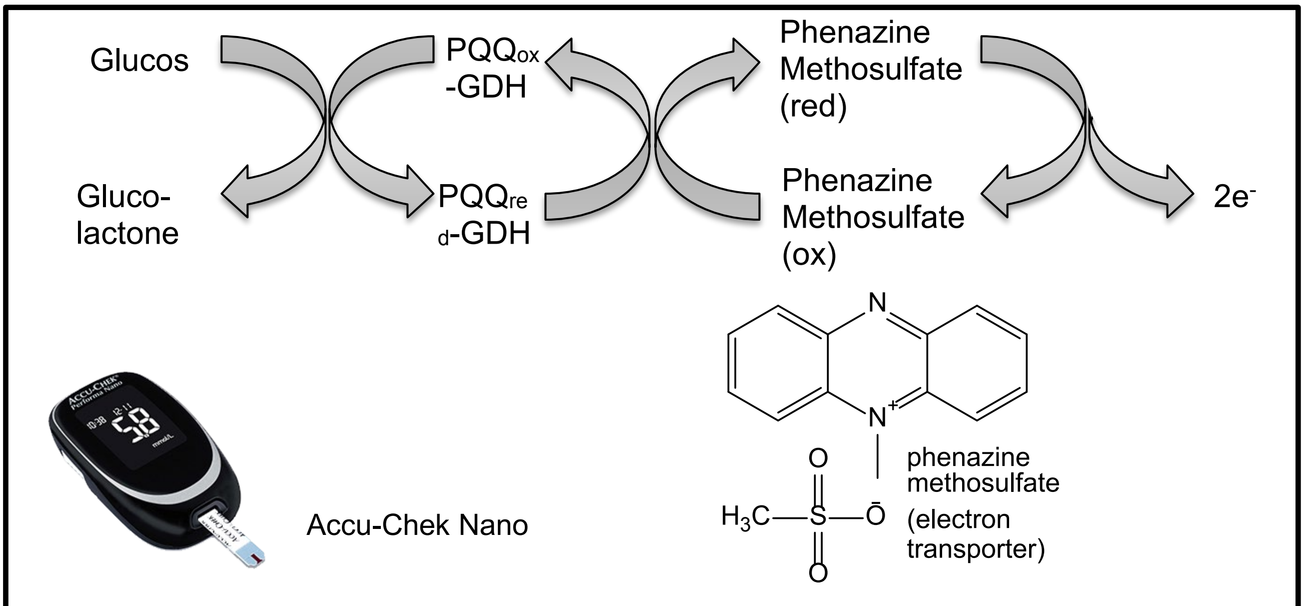


Figure 2: Enzyme linked glucose assay used in the Accu-Chek Nano

Method

Calibration curve and glucometer use.

1. Generate a series of standard solutions of 1, 5, 10, 15 and 25 mmol/L (+) glucose solutions.

a. Whilst preparing the standard solutions, weigh out approximately 25 g of split peas into a 100 mL quick fit conical flask (with lid) using a top loading balance. Add 50 mL of water, seal the bottle and allow it to soak for 10-15 minutes, shaking occasionally (record the actual time!). Once the soaking time has passed, pour the solution off of the split peas into a new, clean bottle.

2. Generate a calibration curve for the glucometer using the standard solutions.

a. Obtain a glucometer and a small vial of test strips.

b. Whilst ensuring that you are wearing gloves, take a test strip and place it into the glucometer so the ‘Accu-Chek’ faces up and the metal end is inside the glucometer. Three 2 s should appear on the screen followed by an image of a strip with a flashing droplet on the tip. Do not get out extra strips, these might get contaminated before you get to use them.

c. Using a glass pipette, place a small amount of the 25 mmol/L solution directly on the end of the tip. Ensure you are adding the liquid to the actual end of strip instead of on top of it (see below). After a moment, record the number on the screen.

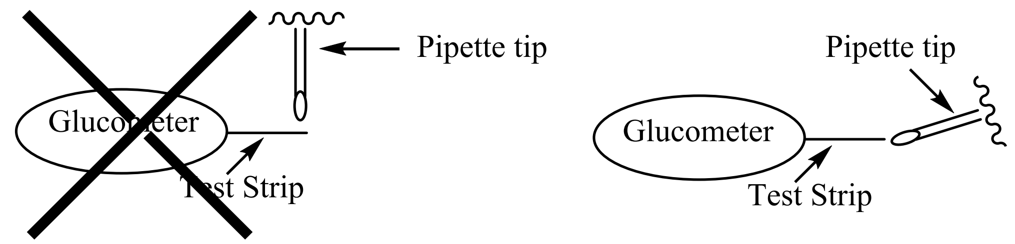


Figure 3: Enzyme linked glucose assay used in the Accu-Chek Nano

d. Repeat the procedure with a new test strip and the 1, 5, 10 and 15 mmol/L solutions.

e. Using excel, graph the recorded values against your known values (those written on the flasks). If the result is not linear, consult your demonstrator.

Sugar Assay

3. Test the ability of the glucometer to monitor the glucose produced by Beano™ and split pea extract.

a. Place ~ 300 mg of Beano™ powder into a 10 mL volumetric flask. Add distilled water to the mark and shake vigorously to dissolve the Beano. Because the Beano™ is in tablet form it has a lot of other ingredients present which are used to hold the tablet together. These will not all dissolve.

b. Heat the aliquots (as per table below) of the supernatant Beano™ and the split pea extract in separate tubes for at least 2 minutes in the pre-heated water bath (45°C ± 5 °C). Avoid the solid when you draw out the Beano™ supernatant.

	A (mL)	B (mL)	C (mL)
Beano™	0.5	1	1.5
Split pea extract	1.5	1	0.5

c. Prepare a stopwatch and hit start once the heated split pea extract is completely added into the heated Beano™ tube.

d. Using the glucometer, test the solutions for the presence of glucose every two minutes. Record the exact times and glucose readings into Excel as you go (NB: DO NOT pretend your time point was exactly 2 minutes if your time point was 20 seconds later. In that case, write 2 minutes and 20 seconds). It is more important to record the exact time than hit the pre-planned time points. **Explain why this is important**.

e. After 10 minutes have passed, take a reading at 15 and 20. If your last two results are not within 1 mmol/L of each other record a 30 and possibly 45 minute point too. Just keep checking if the glucose reading is stabilising (stable is two results within 1 mmo/L). When time permits, begin graphing the data and discuss your results with your demonstrator.

f. Once each member of your team has collected data for up to at least 20 minutes start discussing improvements you can implement for next time you run a curve e.g. change time intervals, run longer, more split pea extract to get a stronger curve, more Beano™ to reach your maximum faster, change time points, lengthen or shorten your run etc. Select the best condition to take through to the temperature investigation and if you are making changes other than the temperature explain what changes are made.

4. Investigate the effect of temperature on the rate of glucose production.

a. As before, add a labelled vial to either one of the other pre-heated water baths (~25 °C and ~80 °C) or the ice bath and add the supernatant Beano™ solution (the volume of which you deemed ideal after 3.). Ensure the solution has at least 2 minutes to reach the temperature of the water bath. What might happen if you leave the Beano™ at extreme temperatures for longer than the suggested two minutes?

b. Prepare a stopwatch and add the split pea extract into the pre-heated vials (again, the volume is determined by your previous results).

c. Using the glucometer, test the solutions for the presence of glucose after the addition of the split pea extract (determine your own time periods). Record the exact times and glucose readings into Excel as you go.

d. Readings up to 15-20 minutes will give you enough data to compare initial glucose production rates so do not exceed these times. (NB: If no glucose is produced after 5-6 minutes there is no need to collect any more data).

In Depth Investigation

5. Choose one of the following series of questions and design a collection of experiments (week 1) to be implemented in order to discover the answers (week 2). Not all the answers are known for these investigations so see what you can discover! Fill in the template found on Moodle individually and discuss as a group to decide your plan for the investigation. Discuss with demonstrator at the start of the second week before you start.

Does beano™ survive the gut?

- What pH is the gut? Is there a range?
- What conditions change the pH in the gut?
- What acid is in the gut?
- If there is deactivation how do we know it is deactivation (destruction) of the enzyme and not just failure to function at that pH?

Can you take beano™ with supplements like Ca?

- What assumptions might you make about pH, volume of meal, type of acid, other components present?
- What forms of Ca do people take, when and how much?
- What other supplements might people take and at what strength?

What bean is best to substitute to decrease intestinal discomfort without using beano?

- Do different beans vary by much? How?
- How do we know that the sugars that come out at the start are indicative of the whole composition of the beans?
- How might we investigate that?
- Who has a bean to bring in?

How long and how hot should I soak my beans to decrease my intestinal discomfort?

- How do I work out how much RFO's are being extracted over time?
- How do I know when all of the RFO's are extracted?
- What are the best methods for extracting? Can these be done in the kitchen?

What metal solutions could interfere with the action of beano?

- How much of those metals could be present before inhibition was noted?
- How much until complete inhibition is noted?
- How likely is this to be relevant? For example, would I ever ingest enough of a particular metal to interfere with beano?

What happens when I drink something alcoholic when I take beano?

- Is the enzyme inhibited by the ethanol?
- Does the ethanol interfere with the measurement of the glucose level?
- Does the ethanol do anything to the bean extract?

Appendix 3 – Electronic Waste

Exercise 4: Electronic Waste

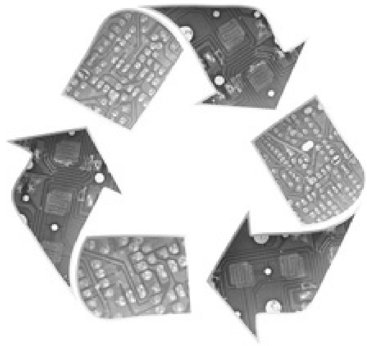
Learning Outcomes:

1. Safely conduct small-scale reactions in micro test-tubes.
2. Identify certain inorganic reaction types, including the formation and dissolution of precipitates and coloured complexes.
3. Make observations and relate the observations back to inorganic chemical theory.
4. Write balanced equations for inorganic reactions.

Hazard Identification and Risk Assessment			
Identify the Hazard (the potential to do harm)	**Risk** (the probability that harm may result)	**Control the Risk** (preventing an incident)	**Disposal of waste**
Zinc nitrate (0.1 M) Irritating to eyes and skin	Solution is dilute. Low	Avoid contact with skin and wash immediately with water if spill occurs.	Heavy metal corrosives carboy in fumehood.
Copper (II) sulfate (0.1 M) Irritating to eyes and skin	Solution is dilute. Low	Avoid contact with skin and wash immediately with water if spill occurs.	Heavy metal corrosives carboy in fumehood.
Cobalt (III) chloride (0.1M) Irritating to eyes and skin	Solution is dilute. Low		Heavy metal corrosives carboy in fumehood.
Iron (III) nitrate (0.1 M) Corrosive. Irritating to eyes and skin	Solution is dilute. Low		Heavy metal corrosives carboy in fumehood.
Ammonia (0.1 M, 0.5 M) Corrosive. Causes burns. Very toxic to aquatic organisms.	Low due to concentration	Avoid contact with skin and wash immediately with water if spill occurs.	Heavy metal corrosives carboy in fumehood.
Ammonia (4 M, Conc.) Corrosive. Causes burns. Very toxic to aquatic organisms.	Medium	Use in fumehood to avoid fumes. Avoid contact with skin and wash immediately with water if spill occurs.	Heavy metal corrosives carboy in fumehood.
Sodium hydroxide (25%) Corrosive. Causes burns, damaging to eyes.	Medium	Handle with care. Wash skin immediately under water if spill occurs.	Heavy metal corrosives carboy in fumehood.
Sodium hydroxide (0.5 M) Corrosive.	Low due to concentration		Heavy metal corrosives carboy in fumehood.
Hydrochloric acid (conc) Corrosive. Causes burns, damaging to eyes.	Medium	Handle with care. Avoid contact and wash immediately with water if spill occurs.	Heavy metal corrosives carboy in fumehood.
General glassware Cuts	Medium	Handle with care, dispose of broken glass using dustpan and brush. If cut occurs, see your demonstrator and seek 1st aid.	Labelled broken glass bin.

Introduction

The electronic devices we use (and most of us love) are full of high value metals (41 at last count). As these consumer goods get replaced we need to ensure that the metals they contain don't go to landfill. In landfill they are a wasted resource, but they also degrade and leach a toxic mix of dissolved metals into the ground. By recycling our old devices (e-waste recycling) we keep valuable metals in the supply chain, and prevent landfill contamination. It is also often cheaper to recycle these metals rather than dig them out of a dwindling supply in the ground in a remote area. In this way, electronic waste recycling is the mining of the 21st century, only this time we are mining waste.



This week you will be investigating precipitation, solubilisation and complexation, and using what you learn to determine the identity of the metals contained in a solution of e-waste leachate. Solubilisation and precipitation are very important in industry as they can be cheap and easy to perform at very large scales. Metals of all types from electronic and electric waste are often recycled by dissolution and selective precipitation.

You will start by looking at cobalt in detail (Part A), exploring how cobalt can be changed from one form to another, in order to understand how and why these changes are happening. You will also get a lot of practice writing out chemical equations, a vital skill in chemistry. Some of the equations are included in this experiment; they help us follow what is happening chemically and assist us to accurately record observations. Work out which ones describe each reaction and practice writing them out correctly.

In Part B you will observe and record how iron, copper and zinc solutions behave when dilute ammonia is added. This will allow you to observe how varied metals behave in this system, and to record your observations for use in identifying your e-waste metals in Part C.

When writing your balanced equations, leave out any spectator ions. Spectator ions are ions which are not involved in actual chemical change; they appear as both a reactant and a product in an ionic equation. To find out which ions are spectator ions:

1. **Write your molecular equation.** $AgNO_{3(aq)} + NaCl_{(aq)} \rightleftharpoons AgCl_{(s)} + NaNO_{3(aq)}$
2. **Ionise your equation.** $Ag^{+}_{(aq)} + NO_3^{-}{}_{(aq)} + Na^{+}_{(aq)} + Cl^{-}_{(aq)} \rightleftharpoons AgCl_{(s)} + Na^{+}_{(aq)} + NO_3^{-}{}_{(aq)}$
3. **Cancel spectator ions to obtain.** $Ag^{+}_{(aq)} + Cl^{-}_{(aq)} \rightleftharpoons AgCl_{(s)}$

Skills tests: There are steps in this lab class where your skills will be checked by the demonstrator so you can be advised where your technique can improve. These are called practical skills tests. If your demonstrator is busy when you get up to a certain step, do not stop and wait. The demonstrator can observe you during Part A and Part B of the experiment; thus, it is always best to get started rather than wait.

Table 4.1: Skills test table (copy this into your laboratory notebook, and have your TA (demonstrator) observe you satisfactorily complete the following tasks)

Practical Skill	Satisfactory?	Comments from TA
Safe addition of the solutions (use of hood, gloves, not holding test tube)		
Correct observations		
Good note taking practice		

Experimental procedure

Part A – Cobalt Equilibria

Starting with the solution of cobalt chloride hexahydrate, written as $[Co(H_2O)_6]Cl_2$ or $[Co(H_2O)_6]^{2+}$, follow the cobalt equilibria scheme on the next page, and record the results in your laboratory notebook. You will be adding a range of reagents to $[Co(H_2O)_6]Cl_2$ (0.1 M) and observing the results. The reagents you will be working with include:

- Ammonia, NH_3 (forms ammonium hydroxide, NH_4OH, when in dilute solution)
- Sodium hydroxide, NaOH
- Hydrochloric acid, HCl

For **each addition**, record both your **observations** and the **equations** which describe the chemistry which is occurring, including specifying whether each compound is in the aqueous form (aq) or the solid (s). Use the equations over the page to assist you with this.

Generally, start with ~0.5 cm^3 or 0.5 mL in a **small** test tube (there are a couple of exceptions). You will need multiple tubes but don't take more than 8 tubes – it is better to clean and reuse the test tubes as you go.

If you think you have missed an observation, talk to your demonstrator about what to do. You will be practicing taking concise but fully descriptive notes on your observations. For example, you might see a reaction that has a blue precipitate forming, which settles over time leaving the solution clear and colourless. You could write several sentences on this but today you are asked to write concise dot point notes. We ask you to practise this because it is also important to learn how to write concise notes which organize your thoughts and observations without loss of detail.

As you complete Part A, find the question number in the cobalt scheme and consider the following questions:

Q1: Why does this reaction form a hydroxide upon the addition of ammonia? Support your answer with a chemical equation.

Q2: How would you reverse this reaction?

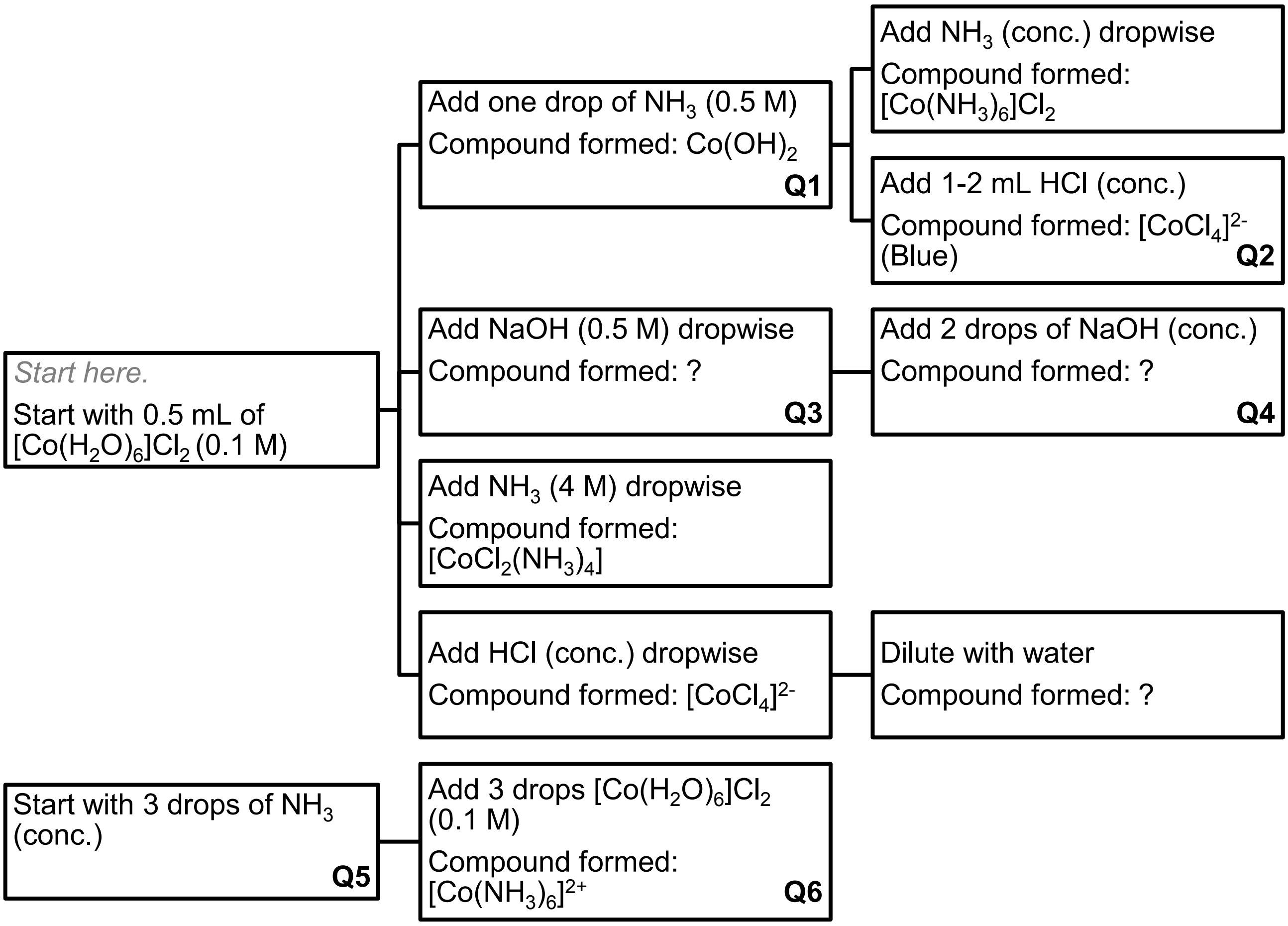
Q3: Are these reactions reversible? Which arrows should you use?

Q4: Are these reactions reversible? Which arrows should you use?

Q5: How would you reverse this reaction?

Q6: What would the counter ions be?

Cobalt Equilibria Scheme – Part A



Part B – Ammonia Addition

For each metal solution (copper, iron, zinc and cobalt), record your observations about the changes in the solution when you add 0.5 M NH_3 solution, and write out the equations.

Start with 0.5 cm^3 of your metal solution in a small test tube, and add 0.5 M NH_3 solution in two steps: first dropwise (recording your observations as you go), and then adding a further 1-2 mL. For some of the solutions you will only need to add 2-3 drops of ammonia solution to see changes. For others, you may need to add up to 2 mL. For each step, record the amount of ammonia solution you needed to add in order to observe a change, and write out the corresponding equation. Make sure you add the ammonia solution slowly and record changes. If you go too quickly you will miss your observation and have to repeat that step. If you think you have overshot, try again with some more of the same metal solution in a new test tube.

Repeat this experiment four times, once with each metal solution (copper, iron, zinc and cobalt). Make sure you note your observations and equations in your laboratory notebook, as this is the information you will use to determine the main metal components of your e-waste unknown.

Part C – Electronic Waste Unknown Identification

Find out your assigned e-waste leachate and repeat the ammonia addition for your unknown solution as you did in Part B for the known metal solutions. There are two metals in each unknown. From your observations propose which two metals are present in your e-waste solution. Be sure to record all results and observations.

Possible Equations

$$Ag^{+}_{(aq)} + Br^{-}_{(aq)} \rightleftharpoons AgBr_{(s)}$$
$$Ag^{+}_{(aq)} + Cl^{-}_{(aq)} \rightleftharpoons AgCl_{(s)}$$
$$2\ Ag^{+}_{(aq)} + 2\ OH^{-}_{(aq)} \rightleftharpoons Ag_2O_{(s)} + H_2O_{(l)}$$
$$Ag(X)_{(s)} + 2\ NH_{3(aq)} \rightleftharpoons [Ag(NH_3)_2]^{+}_{(aq)} + X^{-}_{(aq)}$$
$$[Co(H_2O)_6]^{2+}_{(aq)} + 4\ Cl^{-}_{(aq)} \rightleftharpoons [CoCl_4]^{2-}_{(aq)} + 6\ H_2O_{(l)}$$
$$[Co(H_2O)_6]^{2+}_{(aq)} + 4\ NH_{3(aq)} \rightleftharpoons [Co(NH_3)_4]^{2+}_{(aq)} + 6\ H_2O_{(l)}$$
$$[Co(H_2O)_6]^{2+}_{(aq)} + 6\ NH_{3(aq)} \rightleftharpoons [Co(NH_3)_6]^{2+}_{(aq)} + 6\ H_2O_{(l)}$$
$$[Co(H_2O)_6]^{2+}_{(aq)} + 2\ OH^{-}_{(aq)} \rightleftharpoons Co(OH)_{2(s)} + 6\ H_2O_{(l)}$$
$$Co(OH)_{2(s)} + H_2O_{(l)} \rightleftharpoons [Co(H_2O)_6]^{2+}_{(aq)}$$
$$Co(OH)_{2(s)} + 6\ NH_{3(aq)} \rightleftharpoons [Co(NH_3)_6]^{2+}_{(aq)} + 2\ OH^{-}_{(aq)}$$
$$Cu^{2+}_{(aq)} + 2\ OH^{-}_{(aq)} \rightleftharpoons Cu(OH)_{2(s)}$$
$$Cu(OH)_{2(s)} + 4\ NH_{3(aq)} \rightleftharpoons [Cu(NH_3)_4]^{2+}_{(aq)} + 2\ OH^{-}_{(aq)}$$
$$Fe^{3+}_{(aq)} + 3\ OH^{-}_{(aq)} \rightleftharpoons Fe(OH)_{3(s)}$$
$$\mathbf{NH_{3(aq)} + H_2O_{(l)} \rightleftharpoons NH_4^{+}{}_{(aq)} + OH^{-}{}_{(aq)}}$$
$$NH_4^{+}{}_{(aq)} + H_2O_{(l)} \rightleftharpoons NH_{3(aq)} + H_3O^{+}_{(aq)}$$
$$Zn^{2+}_{(aq)} + 2\ OH^{-}_{(aq)} \rightleftharpoons Zn(OH)_{2(s)}$$
$$Zn(OH)_{2(s)} + 2\ OH^{-}_{(aq)} \rightleftharpoons [Zn(OH)_4]^{2-}_{(aq)}$$
$$Zn(OH)_{2(s)} + 4\ NH_{3(aq)} \rightleftharpoons [Zn(NH_3)_4]^{2+}_{(aq)} + 2\ OH^{-}_{(aq)}$$

Clean up: Make sure you've put all residues into the appropriate waste container in the fume hood. Carefully clean your glassware (this might require you leaving some of the dirtier tubes in the hot soapy water while you complete all of the equations). Ensure your demonstrator checks your 8 tubes are clean before you put them back in the cupboard. Ensure you clean your hands in the hand basins before you leave.

Report

Complete your laboratory report online, *via* Moodle. Ensure you have completed all sections, and uploaded your laboratory notes to Moodle.

Your online report must be finalised within 7 days of the completion of your class.

Appendix 4 – The individual laboratory survey

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
This lab experience …					
presented real science to students, similar to what scientists do in real research labs.	O	O	O	O	O
was *not* very similar to real research.	O	O	O	O	O
made me realize I could do science research in a real science laboratory (for instance at a university, or with a pharmaceutical company).	O	O	O	O	O
did *not* make me learn.	O	O	O	O	O
gave me a better understanding of the process of scientific research as a result of this experiment.	O	O	O	O	O
helped me understand how the topics that are covered in chemistry lecture are connected to real research.	O	O	O	O	O
helped me better understand chemistry, in general, as a result of completing the chemistry lab.	O	O	O	O	O
was *not* well organized.	O	O	O	O	O
has made me more interested in a science career.	O	O	O	O	O
has made me *less* interested in science.	O	O	O	O	O
has made me more interested in chemistry.	O	O	O	O	O
was interesting.	O	O	O	O	O
was worthwhile.	O	O	O	O	O
was easy.	O	O	O	O	O
was well contextualised to real-life/the workforce.	O	O	O	O	O
was open enough to allow me to make decisions.	O	O	O	O	O
was challenging.	O	O	O	O	O
In this lab …					
the instructional materials provided me with sufficient guidance for me to carry out the experiments.	O	O	O	O	O
I can be successful by simply following the procedures in the lab manual.	O	O	O	O	O
the instructional materials did *not* provide me with explicit instructions about my experiment.	O	O	O	O	O
I did *not* repeat experiments to check results.	O	O	O	O	O
General questions.					
Finding answers to real research questions motivated me to do well in the chemistry lab.	O	O	O	O	O
Having the opportunity to use chemistry instruments made this course *less* interesting for me.	O	O	O	O	O
Having the opportunity to use chemistry instruments helped me learn course topics.	O	O	O	O	O
Finding answers to real world questions motivated me to do well in the chemistry lab.	O	O	O	O	O
Please select both agree and disagree for this question.	O	O	O	O	O
In my life, I will *not* use skills I've learned in this chemistry lab.	O	O	O	O	O
Even if I don't end up working in a science related job, the laboratory experience will still benefit me.	O	O	O	O	O

What skills did you develop throughout today's experience?
Was there anything that could be improved about today's lab?
Did you enjoy today's lab? Why/why not?

Appendix 5 – The laboratory aims and expectations survey

What gender do you identify as?

Male O Female O Rather not say O

What course are you currently enrolled in (Overall, not individual courses)?

____

What age group do you belong to?

16-18 O 19-21 O 22+ O

You are currently enrolled as:

An international student. O A domestic student. O

What do you think the aims of doing a practical chemistry course are?

SD = Strongly Disagree	D = Disagree	N = Neutral	A = Agree	SA = Strongly agree

When performing experiments in a chemistry laboratory course, I expect…

	SD	D	N	A	SA
1 *to learn chemistry that will be useful in my life.*	O	O	O	O	O
2 *to worry about finishing on time.*	O	O	O	O	O
3 *to make decisions about what data to collect.*	O	O	O	O	O
4 *to feel unsure about the purpose of the procedures.*	O	O	O	O	O
5 *to experience moments of insight.*	O	O	O	O	O
6 *to be confused about how the instruments work.*	O	O	O	O	O
7 *to learn critical thinking skills.*	O	O	O	O	O
8 *to be excited to do chemistry.*	O	O	O	O	O
9 *to be nervous about making mistakes.*	O	O	O	O	O
10 *to consider if my data makes sense.*	O	O	O	O	O
11 *to think about what the molecules are doing.*	O	O	O	O	O
12 *to feel disorganized.*	O	O	O	O	O
13 *to develop confidence in the laboratory.*	O	O	O	O	O
14 *to worry about getting good data.*	O	O	O	O	O
15 *to find the procedures simple to do.*	O	O	O	O	O
16 *to be confused about the underlying concepts.*	O	O	O	O	O
17 *to "get stuck" but keep trying.*	O	O	O	O	O
18 *to be nervous when handling chemicals.*	O	O	O	O	O
19 *to think about chemistry I already know.*	O	O	O	O	O
20 *to worry about the quality of my data.*	O	O	O	O	O
21 *to be frustrated.*	O	O	O	O	O
22 *to interpret my data beyond only doing calculations.*	O	O	O	O	O
23 *Please select both agree and disagree for this question.*	O	O	O	O	O
24 *to focus on procedures, not concepts.*	O	O	O	O	O
25 *to use my observations to understand the behaviour of atoms and molecules*	O	O	O	O	O
26 *to make mistakes and try again.*	O	O	O	O	O
27 *to be intrigued by the instruments.*	O	O	O	O	O
28 *to feel intimidated.*	O	O	O	O	O
29 *to be confused about what my data mean.*	O	O	O	O	O
30 *to be confident when using equipment.*	O	O	O	O	O
31 *to learn problem solving skills.*	O	O	O	O	O

Appendix 6 –Teaching associate survey

Appendix Table 3.1 – Questions asked in the online survey of sessional teaching staff.

Question	Closed answer options			
In terms of difficulty, how did you find using the new marking method?	Easy	Hard	Neither	
Compared to marking in previous years, how was your rate of marking affected?	Faster	Slower	Neither	
Compared to marking in previous years, how did the amount of feedback you provided change?	Increased	Decreased	Neither	
Compared to marking in previous years, how did the quality of feedback that you provided change?	Increased	Decreased	Similar	
Using the new Moodle marking, how long did it take you to mark a full set of reports per group for a given experiment (on average)?	Under 1 hour	Between 1 and 2 hours	Between 2 and 3 hours	Greater than 4 hours