International Journal of Innovation in Science and Mathematics Education, *23*(5), 91-103, 2015.

Using Inquiry-based Practicals to Promote Students' Critical Evaluation of the Scientific Literature and Maturation of their Understanding of the Nature of Scientific Knowledge

Kirsten Zimbardi[a], Anne Loyle-Langholz[b], Judit Kibedi[a], and Kay Colthorpe[a]

Corresponding author: k.zimbardi@uq.edu.au
[a]School of Biomedical Sciences and the Centre for Educational Innovation and Technology, The University of Queensland, Brisbane QLD 4072, Australia
[b]College of Education and Human Development, University of Minnesota-Twin Cities, Minneapolis, MN, USA

Keywords: threshold learning outcomes, scientific reasoning, epistemology, class observations

International Journal of Innovation in Science and Mathematics Education, *23*(5), 91-103, 2015.

Abstract

Science graduates need to be able to tackle the complex, novel problems they will face in the 21st Century workplace. In Australia, these skills have been formalised as a set of national academic standards for science graduates that highlight essential skills in inquiry and problem solving. Such scientific thinking skills have been shown to be enhanced by inquiry-based curricula. However, within this curriculum, the extent of student learning gains depends on implementation and how student engage with class activities. We video recorded students undertaking inquiry classes to investigate how students engage with the curricula, and asked students to annotate their videos to highlight instances of scientific thinking, with specific reference to the Australian national academic standards for scientific thinking. Interviews through the semester elucidated the development of students' critical thinking skills during the classes. This methodology allowed us to unpack which specific experiences within inquiry classes provide the most potent learning experiences and drive the maturation of students' scientific thinking skills. In particular, when students need to integrate their novel or unexpected findings with the scientific literature, most are prompted to develop a more mature understanding of the contestable nature of scientific knowledge and the role of inquiry and experimentation in knowledge creation. This article reports on the places where student thinking appears to go awry, the indicators that signal that students have reached these crucial crossroads, and potential approaches to inquiry curriculum implementation to propel students toward a more mature, nuanced and critical way of interacting with their data and the primary scientific literature.

Introduction

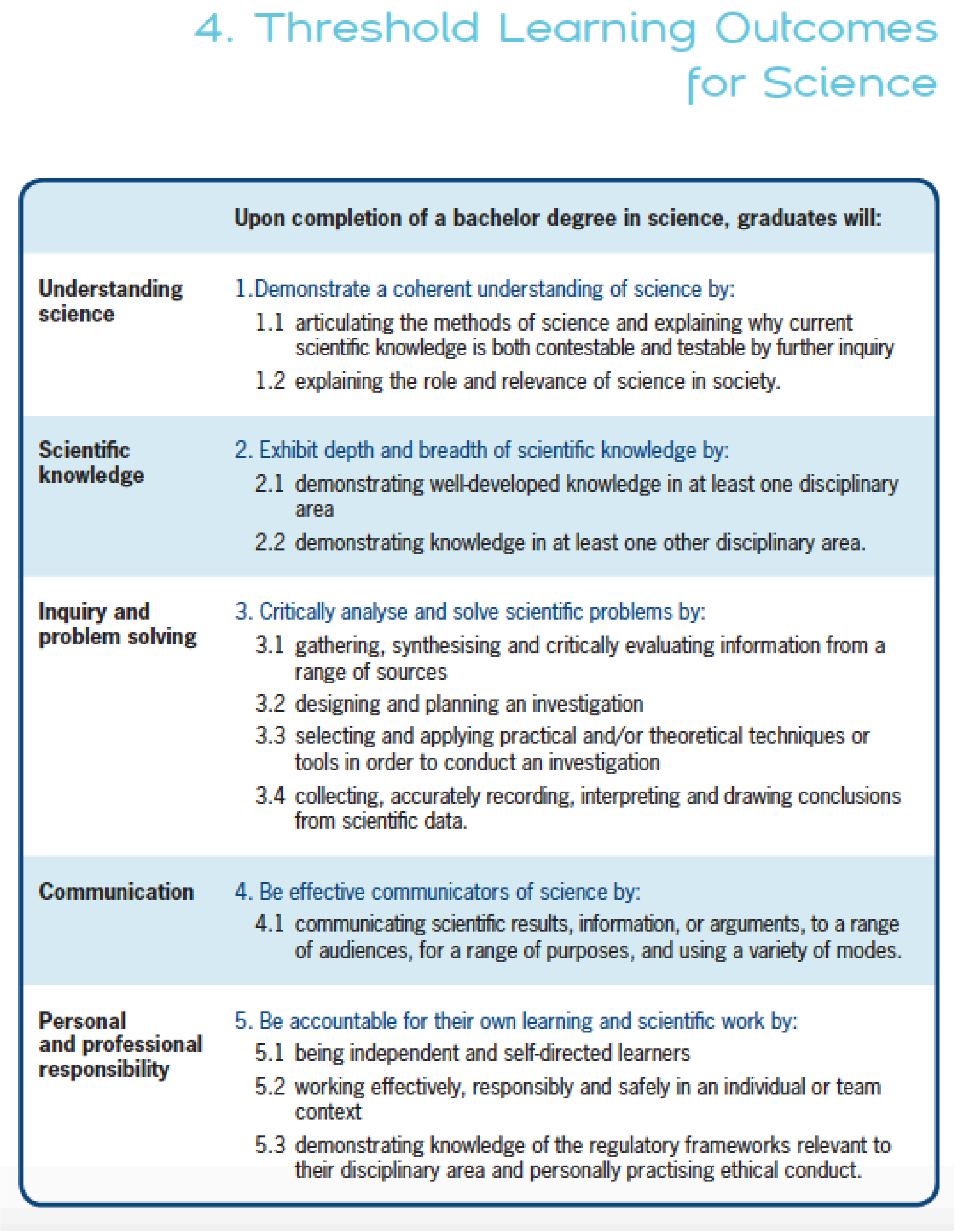
Science higher education is becoming increasingly focussed on ensuring graduates are able to tackle the novel, complex, unstructured problems they will face in the 21st century workplace (Bybee and Fuchs 2006). Inquiry-based curricula have been proposed as a mechanism to help students develop these advanced critical thinking skills, particularly in the evaluation of evidence in complex settings (Kuhn 2009). As a consequence, over the last 30 years, science education has gradually moved toward models of inquiry curricula in order to teach students these cognitive skills, and specifically, how to 'think like scientists' (Dunbar and Fugelsang 2005). However, the ways in which inquiry curricula are implemented in the classroom have significant impacts on the learning outcomes that students achieve (Kirschner, Sweller and

Clark 2006; Kuhn 2009). It is therefore important to understand the ways in which students engage with inquiry activities, and how their experiences shape their learning outcomes.

Inquiry-based curricula commonly involve students "…identifying questions, attending to evidence, identifying patterns, making controlled comparisons, interpreting increasingly complex data, supporting claims, and drawing justified conclusions" (Kuhn and Pease 2008; p.512). All of these steps are essential to scientific method of investigation, and have been detailed as part of the threshold learning outcomes (TLO3) for science graduates (Jones, Yates and Kelder 2011; see Figure 1). It is well established that developing such procedural skills requires practice with appropriate guidance and feedback (Kirschner et al. 2006). However, when dealing with large cohorts, university educators who design the curricula question whether the curriculum is being implemented by their large teams of casual academics as intended, and the degree to which students are engaging in the experiential learning opportunities provided by the inquiry activities (Kirkup, Pizzica and Waite 2010). It is therefore important to utilise rare opportunities for lengthy objective observations of randomly selected students undertaking inquiry classes to understand how the students are likely to be engaging with the inquiry activities.

Beyond these process skills of TLO3, inquiry experiences may also provide students the opportunity to develop higher order understandings of the way knowledge is constructed in their discipline and the subsequent contestable and testable nature of scientific knowledge (Myers and Burgess 2003; Zimbardi et al. 2013). However, the developmental progression from believing in concrete facts to a mature understanding of how scientific knowledge as constructed and contestable, requires students to undergo several substantial shifts in their thinking (Kitchener and King 2002). An extensive longitudinal study has found that such shifts are most likely to occur after students graduate and have several years of experience in the workforce, but do sometimes occur if students engage in inquiry activities during their degrees which require them to investigate novel research questions where there are no known answers (Magolda and King 2004). In contrast, the use of assessment tasks that emphasise the recitation of known concepts have been shown to undermine even the most well designed inquiry-based modules (Kuhn 2009). Therefore it is important to understand how students think about and use scientific literature to determine if changes in their understanding of the nature of scientific knowledge are occurring as students work through the inquiry activities and associated assessment tasks.

This study set out to investigate and gain a more detailed understanding of how students engage with inquiry activities, through video observation of students in class and subsequent reflective interviews. In addition, we analysed these data sources in relation to students' assessment performance to investigate potential shifts in the way students relate to and use the scientific literature.



4. Threshold Learning Outcomes for Science

	Upon completion of a bachelor degree in science, graduates will:
Understanding science	1. Demonstrate a coherent understanding of science by: 1.1 articulating the methods of science and explaining why current scientific knowledge is both contestable and testable by further inquiry 1.2 explaining the role and relevance of science in society.
Scientific knowledge	2. Exhibit depth and breadth of scientific knowledge by: 2.1 demonstrating well-developed knowledge in at least one disciplinary area 2.2 demonstrating knowledge in at least one other disciplinary area.
Inquiry and problem solving	3. Critically analyse and solve scientific problems by: 3.1 gathering, synthesising and critically evaluating information from a range of sources 3.2 designing and planning an investigation 3.3 selecting and applying practical and/or theoretical techniques or tools in order to conduct an investigation 3.4 collecting, accurately recording, interpreting and drawing conclusions from scientific data.
Communication	4. Be effective communicators of science by: 4.1 communicating scientific results, information, or arguments, to a range of audiences, for a range of purposes, and using a variety of modes.
Personal and professional responsibility	5. Be accountable for their own learning and scientific work by: 5.1 being independent and self-directed learners 5.2 working effectively, responsibly and safely in an individual or team context 5.3 demonstrating knowledge of the regulatory frameworks relevant to their disciplinary area and personally practising ethical conduct.

Figure 1. The Learning and Teaching Academic Standards for Science Threshold Learning Outcomes (from Jones et al. 2011)

Methods

Ethical clearance and methodological approach

Ethical clearance was obtained from the UQ Behavioural and Social Sciences Ethical Review Committee prior to the commencement of this study (2012000846), and participating students provided informed consent for the videos of their classes and interviews, and their online annotations to be included in this study and to be made publically available online.

In this study, our aim was to capture students interacting during inquiry *as it was taking place* in order to capture a more nuanced and fine-grained picture of students undertaking inquiry and the changes taking place in their understanding of scientific knowledge. Microgenetic and phenomenographical qualitative research approaches were used as a framework to guide the analysis. The microgenetic method provides an opportunity to examine the changes in student thinking by looking for the evidence of the path, rate, breadth, variability and sources of change (Flynn and Siegler 2007). The phenomenographical approach allowed us to tease out the critical features of the variation in how a shared experience was conceptualised by a group of individuals. To use these methods, we needed a group of students who experienced the same phenomena in diverse ways and for whom we had extensive longitudinal data from multiple perspectives. The students described in this study were selected as follows.

Using convenience sampling (based on where we could place a camera in the classroom to capture student activity), the vast majority (92%) of undergraduate students who were asked to participate, agreed to be involved in the class videos (59 students) and in follow-up interviews (42 students). The majority of these students were scattered across groups, courses and disciplines. For one particular group from our second level biomedical science inquiry-based laboratory practicals, 6 out of 8 students agreed to be involved in 3 in-class videos, attended 2-3 interviews, and provided annotations on the videos of their classes. For these 6 students, the class context became a constant, and the rich, multifaceted data set provided the longitudinal detail necessary to explore critical features of variation in the ways in which these students learnt during that shared experience (Stenfors-Hayes, Hult and Dahlgren 2013). Specifically, we analysed the students' class and interview videos, their annotations and assessment items to understand how their experiences of the same curricula as they worked together in a single group, related to the changes in scientific thinking achieved by each student.

The assessment marks for these students (see Table 1) do provide a representative profile of the range of the course cohort (mean ± SD: 72±11%), but we make no other claims about the degree to which these students might be representative of the cohort. Based on the sampling and analysis methods used, this study is not intended to provide an exhaustive account of student diversity in an inquiry curriculum. It is very likely that other students experienced the curriculum in many more ways than the 6 students we describe here. Instead, this study aims to begin to unpack the variation in the ways in which undergraduate biomedical science students experience inquiry based practical curricula. With the class events, experimental design, data generated and assessment requirements all held constant, we as curriculum designers and course coordinators, wanted to understand what might underlie differences in student experiences of scientific inquiry and the associated learning outcomes – particularly in developing an understanding of the nature of scientific knowledge.

Curricular context

This study uses data from a group of six undergraduate students (*n* = 6) who worked together in the inquiry-based laboratory practicals within the second year human physiology course (Integrative Cell and Tissue Biology: BIOM2011), at the University of Queensland. This single semester (13 week) course consists of three 1-hour lectures each week and one 3-hour practical class each fortnight. The compulsory practical component of this course has been described in depth previously (Zimbardi, Bugarcic, Colthorpe, Good and Lluka 2013). Briefly, students undertake two modules, each consisting of three classes. In the first class, students are introduced to the experimental paradigm in a hands-on skill-building session, design their experiment and perform a trial run of the experiment, and submit a research proposal detailing their hypothesis and experimental methods. During module 1, students use a pithed toad as their experimental model to investigate the impact of a treatment, or combination of treatments, on the heart rate and contractility. In the second module, students act as the experimental subjects to investigate the impact of a treatment, or combination of treatments, on the electrical activity of a muscle or pair of muscles detected using electromyography (EMG). Both modules have been adapted from traditional recipe style practicals, but have been made into inquiry-based practicals by allowing students to choose the treatment(s) they are interested in investigating and requiring them to develop their own experimental approach.

In the second class of each module, students analyse the preliminary data they collected from the trial run of their experiment in the first class. In the third class, students perform their entire experiment and collect the data they will use in their four-page report, which is written to the conventions of a journal article and submitted one week after the third practical class. In the second class of module 2, students are also provided the opportunity to discuss feedback from their module 1 report with their teaching assistant (TA), who mentors the students throughout all six classes of the semester and marks their proposals and reports for each module (Good, Colthorpe, Zimbardi and Kafer, in press).

Data collection from classes, video annotations, interviews and assessment outcomes

All three classes of the module 2 were videotaped. Segments of the first class which captured students discussing their hypothesis and experimental design were spliced together into a 30 minute montage (Final Cut Pro 10.0.6, Apple). This montage was uploaded onto a university-hosted site where students were able to annotate the video with time-stamped comments (http://dev.ceit.uq.edu.au/vcop2/video/biom2011-p1-1st-class-module-2#t=0). The students also participated in three 30 minute interviews where they were introduced to LTAS TLO 3 and asked to describe the ways in which they were (or were not) developing each of the four sub skills (see Figure 1) in any of the courses they were currently undertaking, or had undertaken during their undergraduate degree. During the second interview, students were shown how to annotate the online video and asked to add at least three annotations that related their activities to the LTAS TLO 3 skills (see example in Figure 2). The individual reports submitted for both modules along with the marker’s feedback were also collected for analysis of learning outcomes.

Data analysis

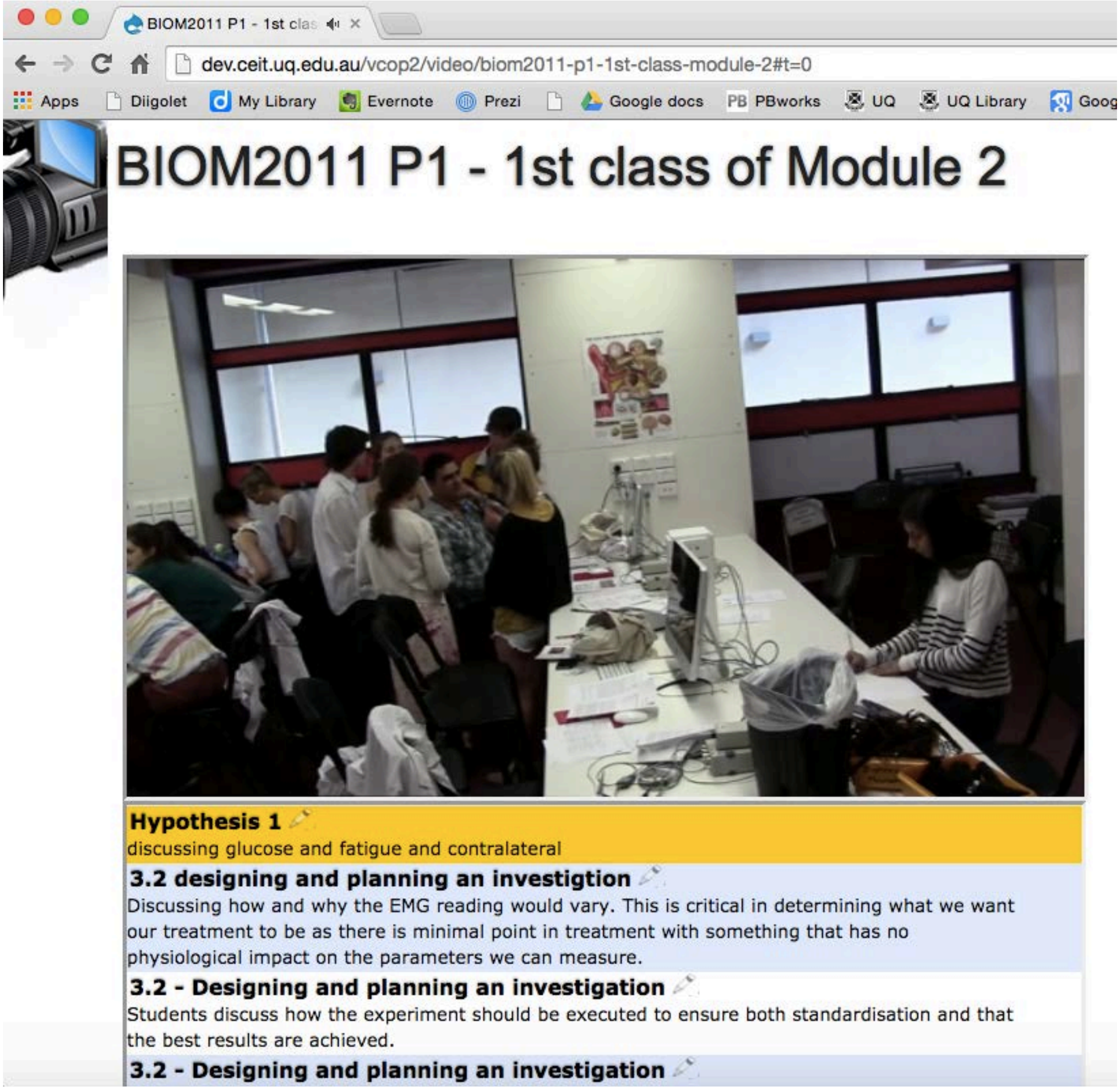
The videos from each of the three classes were reviewed and sections where the students were engaged in on-task discussions about their hypotheses, experimental design, data collection or analysis, interpretation of results, scientific literature or practical assessment tasks were highlighted using annotations in Final Cut Pro. The accuracy of the online annotations that students inserted in the video montage of their class activities were also confirmed by the authors. All of the annotations were used to determine whether the group was engaging in each of the inquiry activities embedded in the curriculum design.

For each student, we then created a summary document that collected together in chronological order the marks for each of the two reports students submitted as part of their assessment (details on assessment available in Colthorpe et al. in this special issue of IJISME), with sections from the interview transcripts and report feedback comments that related to the ways students used or viewed the scientific literature, or the experimental nature of scientific knowledge were extracted for analysis. This triangulation of data allowed us to put together a more complete description of the inquiry experience of each student, and thus use a microgenetic approach (Flynn and Siegler 2007) to characterise the changes and points of transition in their thinking over time. These profiles were then analysed using a phenomenographical lens (Stenfors-Hayes, et al. 2013) where several iterations of comparison between students were used to tease out the critical features underlying the variations in the ways in which students described how they experienced the inquiry classes and achieved the desired learning outcomes.

Results

It is clear, from the videos of this group of students working through all the three classes and their annotations of the online video montage of the first class, that students were actively engaged in the inquiry activities intended in the curriculum design. The first class was dedicated to skill building with the new experimental paradigm, formulating a testable hypothesis and a method to test that hypothesis that had to be written in full in a formal proposal for summative assessment. During this class, there were many long discussions where students proposed, countered, adjusted and refined their hypothesis and method. These discussions demonstrated that students were considering control variables, measurable outcomes, underlying physiological mechanisms and the prevalence of scientific literature relevant to their investigation.

Importantly, the annotations that students inserted into the video montage from this class went beyond merely identifying instances of each TLO3 sub-skills, instead providing detailed explanations and critiques of their discussions (for examples see Figure 2). This demonstrated that the students understood what they were learning, could identify specific events that evidenced all four sub-skills of TLO3, and for the hypothesis formulation and experimental design skills in particular, were working at quite an advanced level of skill.



BIOM2011 P1 - 1st class of Module 2

Hypothesis 1
discussing glucose and fatigue and contralateral

3.2 designing and planning an investigation
Discussing how and why the EMG reading would vary. This is critical in determining what we want our treatment to be as there is minimal point in treatment with something that has no physiological impact on the parameters we can measure.

3.2 - Designing and planning an investigation
Students discuss how the experiment should be executed to ensure both standardisation and that the best results are achieved.

3.2 - Designing and planning an investigation

Figure 2. Examples of the annotations students added to the online video montage of their first class in module 2. Each annotation is time-stamped so that as the video plays, the relevant annotation scrolls to the top and is highlighted in orange

During the second class, the students read through the feedback they received from their TA on their module 1 reports, and discussed this feedback with each other and their TA. Each student seemed to focus on the areas in the criteria where they had scored what they considered to be low marks, and then asked their peers and the TA to figure out what the specific weaknesses were and how to improve. These discussions revealed the students' anger and disappointment, but also their quick adjustment to focusing on fixing the 'problematic areas'. There was a large degree of consistency in these problems across the group, with most students being prompted to increase the depth of the physiological knowledge incorporated in their reports by relying more on primary literature and less on textbooks. When the students asked how they should reference and how many references they should include, their teaching associate (TA) countered with an explanation of the importance of using literature that is relevant to their findings and making sure the depth of their reading was evident in their writing.

During the third class, the group adjusted key aspects of their experimental design during the early stages of data collection. For example, their discussions showed they were considering how to standardise the experimental protocol across human participants, how to ensure they were applying the treatment (inducing muscle fatigue) to the particular muscle they were taking measurements from, and were able see detectable changes in the EMG recordings as they were being collected. Once the group was satisfied with their revised protocol and their method for labelling the data as it was being collected, they ran all of the participants through the experiment and then distributed a copy of the full data set to each group member to complete the analysis individually after class.

Overall, the videos and annotations provided sustained evidence of each of the students engaging collaboratively in the discussions as they identified a research question of interest, formulated a testable hypothesis, designed their experiment and ensured they were executing it effectively and collecting reliable data. How each student worked individually through the subsequent steps of identifying patterns in the data, analysing the data and interpreting their findings in relation to the scientific literature as they prepared their reports was elicited through the interviews.

Statements made by students during the reflective interviews were compared to the report feedback provided by the TA. Through this comparative analysis we gained insights into how students interpreted their feedback and how this impacted their approach to the second report. The module 1 report feedback urged all of the students in the study to use more primary literature when framing their hypothesis and interpreting their experimental findings. Therefore, when each student began constructing their module 2 reports, they encountered the problem of situating their experiment within the current literature. Interestingly, there was a large degree of variation in how the students approached this process. The following section provides excerpts from the reflective interviews and TA feedback on the module 2 reports to illustrate this variation in the ways students used, and related to, the scientific literature. We have organised these examples into a series of three categories, beginning with students who expected the literature to contain all of the well-established ‘facts’ they were used to finding in textbooks, through students who described important shifts in how they used and viewed the literature, to students who understood that their report was expected to build upon the current literature and contribute to the construction of scientific knowledge.

The first category of students appeared to be convinced that the literature contained the ‘facts’ they needed to support their claims and justify their results, if only they could find the articles or had the time and impetus to understand them. The first student admitted, albeit guiltily, to writing first his report first and then searching for the literature to support his assertions, but then found he struggled to find references for everything he has assumed would be present in the literature:

> Student 1: *what we usually end up doing is finding one* [reference] *that agrees with everything we’re saying so we… we’re not really analysing the source itself…we are just using it to back up what we’re doing… but I just don't think I'm good at it* [finding the articles]*…like I don't know what to research for and what literature to look at… I know that other people can find stuff so easily…so they are really good with like research and finding things that they need and I guess that I'm not 100% sure what I am after either… but I guess that other people know exactly what they are looking for so maybe I need to be more strategic…*

In the end, this student appears to have used the articles he found in a superficial way to support the claims in his report, which resulted in feedback from the TA marking this report that highlighted the absence of critical evaluation:

> TA: *there is still an absence of key research - simply saying your findings agree with others without discussing those studies is not really critically evaluating your experiment or the theory.*

In this same category, another student described a similar conviction that even the most unexpected results could be explained by the literature, if he had enough time to work through the jargon-rich articles that contained the answers:

> Student 2: *one* [result] *we had was unexpected… stretching the heart in all the literature I could find always increased heart rate. And ours noticeably always decreased. Also in literature the contractile force was supposed to increase with treatment with adrenaline. And ours like less than halved. So we got yeah. We got the complete opposite of the literature for our results.*
> Interviewer: *How did you cope with that?*
> Student 2: *Trying to come up with some reasons for it and justify them.*
> Interviewer: *What did you come up with?*
> Student 2: *Well I sort of said my general consensus was with the contractile strength when the heart wasn't fully relaxing before the next contraction… but I had no other reasons for… yeah I read some literature and managed to find a reason why it could happen, not 100% sure what is was. Some technical term. It was one of those situations where I didn't fully understand that so…*
> Interviewer: *Time wise or interest, or review of the literature or…*
> Student 2: *Time wise yeah. Time wise and review of the literature. It was a bit over my head. Like I could have obviously spent another couple of days and probably understood it, but at that point it wasn't worth it.*

Based on the feedback from the TA on both reports, Student 2 appears to have consistently focussed his discussions on the limitations of his experiments but provided no data and limited literature to support his criticisms of his experimental findings.

In the second category, two students reported shifts in their use of literature and their understanding of how scientific knowledge is continually undergoing construction and revision:

> Student 3: *Well for the first* [report] *I made my conclusions based on like scattered research and then I went to the literature to support that, which is clearly the wrong way to go about it! But for the next two* [reports in two different courses] *I went and looked up all of the literature first, so … I looked at the research first before I wrote up the final hypothesis, and then I wrote my introduction based on the literature which was much easier. So definitely will do that from now on!*
> Student 4: *I think that what surprised me the most when I came to write it was … realising that actually there's a lot we don't know and there's a lot where you know it should happen but we don't actually know if it does… I suppose in a way, when I think about it,* [it] *can make me more nervous to think that there's so much that you don't know…*

Student 4 described gaps in the current literature that were specific to experimental contexts, saying that there were plenty of articles reporting the impact of the treatment (glucose) "*on animals and...you know...just isolated muscles…* [but not on muscles in] *a live human subject.*" Thus, this category contained an additional level of variation, from Student 3 who simply became aware that their hypothesis needed to be formulated from the current literature, to Student 4 who had come to understand that the limits imposed by each experimental approach creates important gaps in the current literature, and that the purpose of her report was to add novel insights to fill a gap in current body of scientific knowledge.

In the final category, a pair of students who were interviewed together demonstrated a clear understanding that their experiment should address a novel research question. However, in contrast to Student 3 who realised he needed to reformulate his "*final hypothesis*" once he had become familiar with the literature, Students 5 and 6 had become stuck because the original hypothesis they developed before reading the relevant literature, was now turning out to be neither novel nor supported by the literature they were reading. When prompted by the interviewer (an academic previously involved in the course) to change their hypothesis and approach to the data analysis, the students immediately linked in additional relevant literature they had read recently that supported the alternative approach:

> Student 1: *Our report is on GI [glycemic index] and heart rate, and there is no research to suggest there is any correlation at all. So… why would we bother even doing the experiment…It's very difficult to angle it in a way that makes sense. I wouldn't even bother testing it*
> Student 2: *It was more of like a group thought.*
> Student 1: [we thought] *"Yeah, that sounds right… yeah definitely like heart rate…" but then everyone went ahead and did it...[cross talk] and now we are researching… why did we even test that, there is no association…*
> Interviewer: *Ah that's interesting. Because what you could do is, if your heart rate isn't changing, then that is actually a control measure, showing that in terms of cardiovascular performance …it was the same…and then if you show that people who had all these weirdo perceptions of how hard it was…*
> Student 2: *Does glycemia have an effect on the central nervous system? Fatigue of muscle? There is [sic] studies that if you are looking at central and peripheral muscle fatigue…then glycemia has a lot of effect on the central part.*
> Student 1: *That's good, we should do that. That would actually be an interesting read. As opposed to…everyone does heart rate because we can and "Oh look…there's no change!"* (laughter)

As shown in Table 1, the marks that these students received for their module 2 reports overall, and the criterion for critical integration of scientific literature specifically, were loosely consistent with this separation into three levels of maturity in understanding the constructed and contestable nature of scientific knowledge.

Table 1. Scores that each student achieved in their module 2 report as an overall total score and for the criterion* related to the use of scientific literature, in relation to variations in the level of epistemic maturity evident in their descriptions of their use of scientific literature

Student	Use of literature score (out of 10 marks*)	Total score (%)	Level of epistemological maturity
Student 1	3.5	54	1
Student 2	6.5	57	1
Student 3	5	59	1 - 2
Student 4	6.5	83	2 - 3
Student 5	6.5	61	3
Student 6	8	79	3

* the full criteria-referenced standards rubric for this report can be found in Colthorpe et al., in this special issue of IJISME

Discussion

This study used publically available, video evidence of students undertaking inquiry-based practical classes to illustrate how students engage in experiential learning of the TLO3 skills in scientific inquiry. Furthermore, students annotated the video evidence and commented on their ability to use the TLO3 framework and provided detailed explanations of instances where they demonstrated these skills. We explored the variation in how students used and related to the scientific literature when they wrote their reports, and found critical features that aligned with different levels of maturity in understanding the way scientific knowledge is constructed through experimental findings.

There was a large degree of consistency in the way the students in this study worked together in class. Each contributed to the development of a research question that was interesting to them, to the formulation of a testable hypothesis, and to the design of an experiment to test their hypothesis. In the first and third classes, the students frequently challenged each other with questions to identify and address issues in the wording of the hypothesis and the design and execution of the experiment. This collaborative banter has been studied intensively in middle and high school inquiry settings (e.g., Sampson and Clark 2009) where the act of critiquing and putting forward alternative solutions has been shown to produce the greatest learning gains. It is clear from the class discussions, and the level of detail that students included in the annotations, that this group was engaged in the inquiry process and in reflecting on their learning experiences.

This group of students made reasonably similar mistakes in their module 1 report, relying too heavily on textbooks, which lacked the depth of physiological mechanism required for the course. The feedback they received on this initial report indicates that these students entered the course with the expectation that they should focus their reports on the regurgitation of well established facts. This suggests that all six students held a somewhat immature epistemological perspective, believing that science was concerned with concrete facts handed down from textbooks and lecturers (Kitchener and King 2002). With consistent and detailed feedback and what the students thought were worryingly low marks, several students were prompted to develop more mature ways of using and understanding the scientific literature. The realisation that there were numerous gaps in the current body of scientific knowledge, and that these gaps arise from the limitations imposed by each experimental context, signifies an important step along the development from novice to expert scientist (Duschl 2007). However, the variation in student explanation of their use of literature, taken together with the performance and feedback on their subsequent reports, suggests this understanding of the importance of experimental context seems to develop somewhat separately from simply understanding that the purpose of the scientific articles is to identify, and fill, a gap in the literature. Indeed, in their comparison of undergraduate, graduate, and expert researchers in chemistry, and Samarapungavan and Westby (2006) found that an understanding of the rules a discipline uses to judge the relative significance of findings from particular experimental approaches only comes after many years of research experience.

Still, there was evidence in the interviews and feedback on the second report that some students maintained a fixed mindset; they held onto the belief that the scientific literature is used to substantiate overarching facts and concrete truths. From a psychological and philosophical perspective, the thinking involved in evaluating the scientific literature is seen as the behavioural outcome of a set of dispositions (e.g., open-mindedness and curiosity) rather than a set of skills (Facione, Sánchez and Facione 1995). For students who appear stuck in this way of thinking, additional scaffolding and exercises may be necessary to help them progress to a more flexible mindset (Dweck 2006). As presented here, the continued use of superficial, sweeping statements of consensus between the literature and a student’s experimental findings may be the clue to identifying students who need additional assistance to developing their understanding of the contestable nature of scientific literature.

The need to revisit and alter a hypothesis in light of unexpected experimental findings and new insights into the relevant literature is also a critical point at which students become stuck. Detailed longitudinal observations of research groups in molecular biology have shown that important conceptual changes and discoveries often arise from collaborative debate around unexpected findings (Dunbar 1995). We have shown that some students move easily to the realisation that it is both necessary and appropriate to adjust a hypothesis when new information comes to light. However, as is the case with expert scientists and non-scientists, students will often struggle to consider alternative hypotheses when they have a single approach in mind (for review see Dunbar and Fugelsang 2005). Previously we have seen students struggle with this process of re-iterative data analysis and hypothesis formulation when dealing with complex data sets (Zimbardi et al. 2013). As a result, we adjusted that inquiry curriculum to include an additional class dedicated to the collaborative data analysis with TA support, greatly improving students’ experience of the data analysis, their learning outcomes and the quality of their final reports.

This study has provided rare insights into how early stage university students engage with inquiry activities in class during an extended experimental investigation. It is clear that these students are working at an advanced level to formulate testable hypotheses, develop relevant, controlled and rigorous experimental designs appropriate to their context, and to ensure that they collect meaningful data as they execute the experiments. Both observations of the students, and their own critical annotation of their class activities, provide encouraging evidence of their experiential learning of the TLO3 inquiry skills, although the retention of this learning remains to be investigated. Notably, the high expectations set in this assessment and the feedback students received on integrating their experimental findings with the primary scientific literature, prompted several students to change the way they used, and more significantly, the ways they viewed the scientific literature. We have also revealed critical stages at which students may get stuck in developing their understanding of the way scientific knowledge is constructed and its contestable and testable nature. Importantly, students who overgeneralise in their use of the literature, or who provide unsupported justifications for their findings may be struggling to develop beyond a concrete epistemological perspective. Lastly, the difficulty of considering alternative approaches is a well-known danger that threatens to hinder the progress of scientific discovery. Prompting students to engage in open critical debates over their approach to their data and findings toward the end of the inquiry cycle may not only produce better outcomes in the short term, but also instil a culture of constant revision that challenges dogmas and fosters future discoveries.

Acknowledgements

We are grateful to the students who participated in this study and generously gave permission for their videos and annotations to be made publically available (http://dev.ceit.uq.edu.au/vcop2/course/inquiring-minds). We would also like to thank Phil Long and John Zornig for their support in developing and maintaining the online video capture of practice site. This work was funded by a University of Queensland Teaching and Learning Fellowship.

References

Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching*, *43*(4), 349–352.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. *The Nature of Insight*, *18*, 365-395.

Dunbar, K., & Fugelsang, J. (2005). Scientific thinking and reasoning. In: *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press.

Duschl, R. A. (2007). Quality Argumentation and Epistemic Criteria. In: S. Erduran & M.P. Jiménez-Aleixandre (Eds.), *Argumentation in Science Education* (pp.159–175). Springer Science & Business Media.

Dweck, C. (2006). *Mindset: The new psychology of success*. RandomHouse.

Facione, P. A., Sánchez, C. A. & Facione, N.C. (1995). The disposition toward critical thinking. *The Journal of General Education*, *44*(1), 1–25.

Flynn, E., & Siegler, R. (2007). Measuring change: current trends and future directions in microgenetic research. *Infant and Child Development*, *16*(1), 135–149.

Good, J. P., Colthorpe, K. L., Zimbardi, K., & Kafer, G. N. D. (In press). The roles of mentoring and motivation in student-teaching assistant interactions, and in improving experience in first year biology laboratory classes. *Journal of College Science Teaching*.

Jones, S. M., Yates, B. F., & Kelder, J. (2011). *Learning and Teaching Academic Standards Project: Science Learning and Teaching Academic Standards Statement*. Sydney, NSW: Australian Learning and Teaching Council.

Kirkup, L., Pizzica, J., & Waite, K. (2010). Realizing a framework for enhancing the laboratory experiences of non-physics majors: from pilot to large-scale implementation. *European Journal of Physics*, *31*, 1061–1070.

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.

Kitchener, K. S., & King, P. M. (2002). The reflective judgment model: Twenty years of research on epistemic cognition. In: B.K. Hofer & P.R. Pintrich (Eds.), *Personal Epistemology* (pp.37–61). Psychology Press.

Kuhn, D. (2009). *Education for Thinking*. Harvard University Press.

Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction*, *26*(4), 512–559.

Magolda, M. B. B., & King, P. M. (2004). *Learning Partnerships*. Stylus Publishing, LLC.

Myers, M. J., & Burgess, A. B. (2003). Inquiry-based laboratory course improves students' ability to design experiments and interpret data. *Advances in Physiology Education*, *27*(1), 26–33.

Samarapungavan, A., & Westby, E. L. (2006). Contextual epistemic development in science: A comparison of chemistry students and research chemists. *Science Education*, *90*, 468–495.

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, *93*, 448–484.

Stenfors-Hayes, T., Hult, H., & Dahlgren, M. A. (2013). A phenomenographic approach to research in medical education. *Medical Education*, *47*(3), 261–270.

Zimbardi, K., Bugarcic, A., Colthorpe, K. L., Good, J. P., & Lluka, L. J. (2013). A set of vertically integrated inquiry-based practical curricula that develop scientific thinking skills for large cohorts of undergraduate students. *Advances in Physiology Education*, *37*(4), 303–315.