UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Department of Chemistry

The Development, Implementation, and Evaluation of Labdog – A novel Web-Based Laboratory Response System for Practical Work in Science Education

by

Thomas Joseph Wilson

Thesis for the degree of Doctor of Philosophy

January 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
Department of Chemistry

Doctor of Philosophy

THE DEVELOPMENT, IMPLEMENTATION, AND EVALUATION OF LABDOG – A NOVEL WEB-BASED LABORATORY RESPONSE SYSTEM FOR PRACTICAL WORK IN SCIENCE EDUCATION

by Thomas Joseph Wilson

The teaching laboratory is common component of chemical, and science, education. However it is often an exercise in recipe-following, where students simply follow instructions. This research adopts a design-based research (DBR) approach to conceiving, developing, and piloting Labdog: a novel web-based technology for the teaching laboratory. Educators create a digitised version of practical instructions in Labdog, and can enhance instruction through the use of questions, and evidence recording. Students access Labdog before, during, and after the laboratory to answer questions, and record observations in text and photo. Labdog can be considered a Laboratory Response System (LaRS), a novel type of technology, which combines the pedagogical bases of formative assessment, classroom response systems, e-portfolios, and electronic lab notebooks. In alignment with DBR principles, this work adopts an iterative approach to generation and evaluation of such a novel technology. Specifically, this work details a series of three pilot studies, followed by a year-long investigation - each of which took place in Southampton between 2015-2017, with A-level equivalent chemistry students. DBR principles also focus on providing actionable advise for educators who may wish to use Labdog, or some future LaRS technology. The results from these pilots repeatedly suggest that Labdog helped students consciously engage with the relevant chemical or scientific principles of their actions. The evaluation revealed the importance of well-considered design of practicals in LaRS software, notably the need to space questions and steps in accordance with the flow of the practical activities themselves. Students should be allowed to immerse themselves in practical activities, without having to worry about managing the activity on Labdog simultaneously. These findings relate to the psychological concepts of cognitive overload, which should be avoided through question content and focus, and the flow-state, which can be encouraged by the spacing of questions and consideration of practical difficulty versus student ability. Ultimately the research presents a completely novel, technologically-enhanced approach to practical work. It produces both a new tool, and a series of heuristics for designing practical work with a LaRS, based on this, and previous, research. In doing so, this work represents a successful example of DBR, and identifies a number of avenues for future research. Namely, there is a need for more experimentally-designed investigation of the impact of LaRS use on student understanding, as well as more exploratory work on cognitive overload and question design.

Acknowledgments

Everybody who knows a doctoral student will realise, sooner or later, that completing the work is a group effort. It takes a village to raise a thesis. Never one for following tradition for its own sake, the simple truth is that I could not have completed this work alone. The following paragraphs are not empty platitudes, but rich and true expressions of gratitude.

I wish to start by thanking Prof. D. Read - the principle research supervisor. Without DR's faith in my ability and willingness to let me take risks this work would be immeasurably different. For allowing me to work with your students, especially when my software was more error than function, and for providing professional insight and guidance into the mysterious world of Academia. To Dr. P. Duckmanton, thank you for providing a fresh pair of eyes, and reassurance when required. Prof. S. Coles - thank you for your faith and understanding of my technical ability, and for your calm enthusiasm. I wish also to thank the patience, honesty, and time of every student who used Labdog and was generous enough to provide feedback.

To my parents, Micheal and Jane. Transitioning from an undergraduate to a young adult has been the most exciting part of my life, to date. Without your love, support, patience, and constant questioning it would have been much more terrifying. Thank you for every meal you have cooked me, every decision you let me make for myself, and always letting me know that you would be there if things went south. I love you both. Also thank you for reassuring Grandma that this is a legitimate career choice and that hopefully I won't be unemployed.

To my friends. You had to deal with the day-to-day ups, downs, sideways, whichever-ways that come with completing a thesis. Thank you for very convincingly pretending to care about my research, for accepting my peculiar love for software, and for eating the food I made when I needed a break. In no particular order, I wish to thank Mike, for persuading me to do a Ph.D. in the first place and for showing constant dedication and passion to your work; Ruth for being a stellar human being, for always taking me seriously, and for always being honest but never being judgmental; Matthew, for financial advice and optimistic nihilism; Hattie, for being fierce and constantly acting on how you feel; Rachel, for being the most honest and thoughtful human I have met to date; and Kathryn, for her calm words and unconditional friendship.

To the staff of Mettricks. Thank you for providing me with, conservatively, a thousand cups worth of coffee. I am certain this has lead to a caffeine dependence, but I am equally sure that I could not have spent so many hours at my other desks.

To God, or the universe. I'm rather afraid to admit that my spiritual development has been overshadowed by the work detailed over the next 200 pages. Nevertheless, I wish to acknowledge and thank the inner-peace which you find whenever you seek it. To the reassurance given to me by my cosmic insignificance. I offer thanks for every message I have received, even if I couldn't read it at the time or presently. I wish to thank the time given to me by those who have started their next great adventure: M. Moazami and M. Wilson.

On mental health. Over the three years it took to complete this work I have experienced depression, eating disorders, and anxiety. There have been moments where I have felt isolated, insignificant, and unable to complete the task at hand. With the possible exception of my problems relating to food, I have heard similar comments echoed by a great many Ph.D. students. I would even go so far as to say that the majority of those who I have met have made such comments. If you are a Ph.D. student, or know one such, I cannot express enough how important it is to vocalise these feelings. To friends, family, supervisors, or medical professionals - all of these people wish to see you succeed. Please take advantage of the help available. You are not making a big deal out of it, you are acting appropriately.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Even in the relatively new field of chemical education[1], there are already a broad range of focuses and methodologies from which a researcher could pick. This research, its intention, design, focus, and findings have all been influenced heavily by the context in which the work took place. It is therefore important to understand this context, in order to understand the work.

In this opening chapter I will briefly cover a number of early scoping projects conducted in the first 10 months of the research. Understanding these projects clarifies and rationalises how the doctoral work evolved. I then overview a number of relevant concepts and theories which emerged from this work, namely about the teaching and learning processes relevant to chemistry. I then provide a brief introduction to Labdog, the name given to the software developed as part of this research, and about which the rest of the work detailed is dedicated. The chapter closes by describing the research questions which guided all of the subsequent work detailed in the following chapters.

1.1 Early Scoping Work

A series of investigations which took place in 2014/15 heavily influenced subsequent work, introducing a number of recurring themes and concepts. The scoping work consisted of two projects: the summer vacation work (SVW) and a novel technology-based intervention with first year undergraduates (UG), working with an existing web-technology called Zaption[2]. The impact of each of these works is

---

[1] Used synonymously with ‘chemistry education’

[2] Zaption has since been acquired by Workday and is no longer available- `https://www.edsurge.com/news/2016-06-30-zap-zaption-sold-to-workday`

summarised graphically in Figure 1.1.

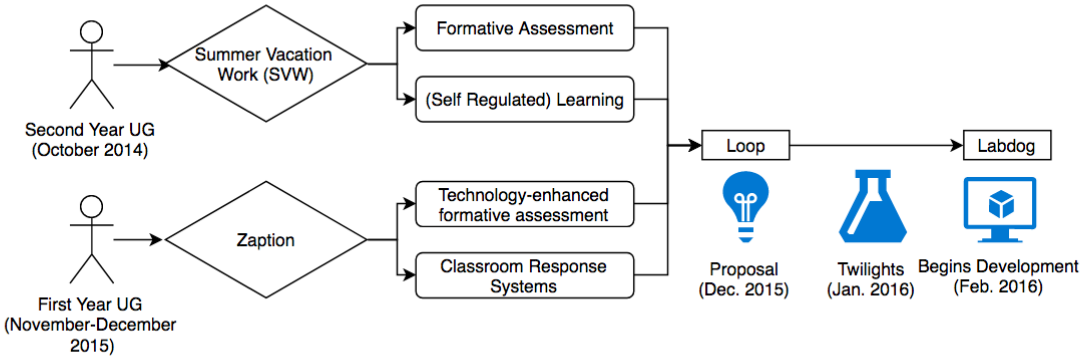


Figure 1.1: A schema and timeline of early scoping work.

The SVW is an exercise carried out by second year undergraduate chemistry students before they recommence full-time study after summer. Students complete a series of problem sets, created by their educators, which cover much of the first year syllabus. Students then mark their own work using video mark schemes created by the educators. As part of the scoping research, I looked at students' responses to reflective questions which were asked as they submitted their marks online. I also conducted a series of focus groups, and distributed two closed-answer surveys - in order to investigate the student-perceived impact of the SVW.

Zaption was used in a four-week period of first year undergraduate organic chemistry teaching. The lecturer, Jon Watts, would use a tablet-computer to make 10-30 minute-long screencasts which covered lecture content, or introduced new ideas. Using the Zaption software, Jon was able to place questions in the videos, which students would answer using a web browser, prior to arriving at lectures. Jon would then use students' responses to these questions to guide his lecture content ad pacing. I examined the qualitative and quantitative data produced as a result of this experience to examine student engagement.

1.1.1 Consequences to future research

Unfortunately, neither of the projects were logistically viable or suitable for continued doctoral-level research. Nevertheless, they provided invaluable guidance which shaped future research.

Both projects used technology to facilitate formative assessment, i.e. having students demonstrate their (mis)understanding in an un-assessed scenario. In the SVW this was used as a starting point for students to reflect on their knowledge of their understanding, i.e. metacognition (Thomas, 2012). Zaption used this information to help direct and inform lecture structure, the notion of just-in-time teaching (Novak, 2011).

As a result of this work, three central conceptual areas came to my attention from the research: the process of learning, specifically self-regulated learning; the potential of formative assessment; and technology-enhanced or blended learning. Each of these concepts is discussed briefly below, and also addressed in greater detail later in the thesis.

(Self regulated) learning

The SVW raised important questions about the learning process. From students' responses about their past and future learning practices, it appeared that they relied most heavily on simple memorisation, e.g. from notes or textbooks. This runs counter to the more conceptually advanced activities of applying one's understanding, e.g. in the answering of questions. Developing understanding requires students to apply and evaluate their ideas, rather than just memorise (Krathwohl, 2002). The learning tasks and processes required to develop deeper understanding require students to go beyond simply reading or consuming knowledge (Prince, 2004; Vrugt and Oort, 2008; Brown et al., 2014) - though such practices made up little of the student-reported study activities, according to the SVW.

This lead me to identify the broad concept of self-regulated learning (SRL). SRL characterises learning as the interaction between an individual's motivations, cognition, and metacognitions (Schraw et al., 2006), as well as interaction between an individual and their environment (Zimmerman and Moylan, 2009), not as an isolated series of activities.

Formative assessment

Both of the scoping investigations used formative assessment, i.e. asking students questions which would then be used to inform better learning and understanding. The intention was to give students both a reason to re-cover the material, and to help them understand the limits of their own knowledge.

The use of formative assessment has a strong and consistent research base behind it (Yorke, 2003; Black and Wiliam, 2009; Bennett, 2011) - and can be related to SRL

(Nicol and Macfarlane-Dick, 2006; Clark, 2012). Formative assessment is a powerful instructional tool which can help students across age, subject, and ability (Nicol and Macfarlane-Dick, 2004). Formative-based assessments therefore presented themselves as a valuable tool in promoting learning and understanding in students.

Blended learning

Both the SVW and Zaption projects featured the use of web-based technologies. The integration of technology and learning is referred to as either t'echnology-enhanced learning' or 'blended learning'. Blended learning can take many forms, depending on how it is integrated with non-technological instruction (Seery and O'Connor, 2015). The scoping research demonstrated how technology could be used as both a teaching and learning tool, while data which could be analysed as part of education research.

1.1.2 The decision to develop custom software

After the scoping projects, I became interested in how technology could be integrated with teaching and learning 'on the ground', i.e. in a typical classroom. I found that web-based technologies fitted this need well, as they are accessible regardless of operating system and location. During this period, I started examining a large range of existing educational web-based educational technologies. Specifically searching for those which had a focus on providing and facilitating formative feedback.

Many of the technologies were commercial, often requiring payment of a per-student, or year-long basis. This made little sense for the context of exploratory work - where the time and financial investment may ultimately outweigh the educational benefits.

I therefore became interested in the possibility of developing a custom piece of software. Initially I developed the proposal for a piece of software called 'Loop'. So named by the need to close the feedback loop, i.e. allow formative feedback to ultimately affect students' future actions (Sadler, 1989, 2010).

Loop would be a central repository for educators to give feedback to individual and grouped students. It would work with seminar or workshop environments, where students would complete the same problem sets. By relating the feedback to the questions, it would also be a research tool to identify misconceptions by specific concepts or question types. Educators would be saved time by providing feedback to groups, not individuals, and would benefit logistically from a paperless workflow. Students would benefit by receiving more and more specific feedback.

1.1.2.1 The evolution of Labdog

After initially proposing Loop, and getting constructive feedback about the logistics and purpose of such a technology, another opportunity presented itself. Every winter, the University of Southampton hosts a series of outreach events, so called ‘Twilights’. During the twilights, several hundred AS-level chemistry students (16-17 years old) complete a natural product extraction in the UG teaching laboratory. I was part of the demonstrating team for these events in 2015, which were co-ordinated by DR, the research supervisor.

Over the Twilights, myself and DR noted a significant difference in the quality and style of teaching delivered by other members of the demonstrating team, many of who are post-graduate (PG) students from the department. Many demonstrators simply aimed to facilitate students safely through the practical. This meant I would sometimes see students stand by as the demonstrator set up equipment, performed steps for them, or corrected mistakes. Where an educationally-focused demonstrator may see a chance for guided problem-solving, others were seeing a problem which would be easier and faster for them to fix themselves. This is not fault of the demonstrating staff, but nevertheless resulted in numerous missed opportunities for learning.

It became clear that certain points in a practical were more difficult, e.g. when students would have problems using a new piece of equipment, or performing some particular process. When helping students at these points, it was an excellent chance to start a conversation, e.g. why we use an ice-bath to aid crystallisation, or how a rotary-evaporator reduces the boiling point of the solute. At these points, the students became a captive, if not entirely willing, audience to Socratic style questioning on both the chemistry and the methodology at hand.

At these points, many students required a great deal of prompting and scaffolding to link the relevant concepts, e.g. temperature and solubility, or pressure and boiling point. In my experience, a great many students would attempt to answer questions with rote learning or ‘key words’ they might use in an exam, e.g. many used “surface area” to explain almost everything. However as soon as I would ask a follow-up question, e.g. why an increased surface area might increases the rate of evaporation, many students would stumble and not be able to answer. Students would require varying degrees of prompting to explain their answer, and very few students immediately explained a visual or observable phenomenon using particulate or scientific theory.

Unfortunately, it was impossible for myself and the small number of other educationally-aware demonstrators to provide valuable scaffolding and questioning to all students at the same point in the practical. Each session contained between 30-50 students from 1-3 schools, mostly working in pairs. It is simply not possible to see at a glance where all students are in a practical, to check their understanding, and to provide resources to students who need them.

Following this experience, myself and the rest of the research group started discussing this problem and potential solutions. We wanted some way to make sure all students were asked the same questions, at the same point in a practical. Or at very least, that students were given a chance to demonstrate their (mis)understanding at these points. Furthermore, this information should then be used to provide some kind of in-person feedback. After several days of thinking and talking the problem over, it lead to the following e-mail from DR to myself:

```
From: D.Read@soton.ac.uk
To: tw4g10@soton.ac.uk
Subject: Wagamamas
Date: 21/01/2015, 20:37

Have you ever been to Wagamamas, where the staff come to ask
you how your food was, and then mark on your placemat that
they've spoken to you so other colleagues don't come along
and trouble you? Maybe we could issue little coloured stickers
to demonstrators so they can stick them on the fume hoods
of the groups they've asked questions of at different
stages of the experiment. That way any demonstrator
will know who has been asked particular questions and
who hasn't.

Just an idea...
```

This e-mail formed the basis of a technology which could track students' progress through a practical activity, and facilitate educator-led answering of formative questions. No extant technology met our needs, and although existing technologies such as LabBuddy[3] or LabStep [4] are present, these were not accessible for teachers and students - requiring specialist-training and more heavy time investment.

This presented the perfect opportunity to develop a technological solution which could facilitate formative assessment in the laboratory context. The software I created in response to this problem is called Labdog. I developed Labdog into a fully functional piece of software[5], as it became the centre of the doctoral research. Educators are able to use Labdog to ask questions, provide feedback, and watch student progression in real-time. This research centres entirely around the development of Labdog, the subsequent piloting, and finally detailing a year-long case study application with the science foundation year (SFY) at the UoS.

The decision to write from a first person perspective

This is the first doctoral-level work to emerge from the UoS's Chemical Educaiton Research Group. The lack of pre-existing research in the department presented many possible paths that future research could take. It is this freedom for variation which justifies my, potentially unusual, decision to write in the first person. This research has been undeniably and continually shaped by personal experience with learners and educators - through conversations, observations, and investigation. By writing in the first person I hope to serve frequent reminder that this work was shaped by experiences 'on the ground', i.e. in the lecture hall and in the lab. Having made numerous attempts to present this research in the scientific third-person past-tense format, it became increasingly difficult to explain and justify the nature of the research I conducted.

1.2 The structure of this thesis

This first chapter provides an overview of research context. I give a literature-based discussion of the ideas of learning, both in general and in a chemistry-specific context. This discussion justifies the need for chemistry-specific education research, and introduces the ideas of 'learning' and 'understanding' as complex issues. I then provide a brief introduction to Labdog, to provide a better understanding of exactly what it is

---

[3] https://www.labbuddy.net

[4] https://www.labstep.com/

[5] http://soton.labdog.co.uk

and does. I close the chapter by formally stating the research questions which guided the rest of the research.

Chapter 2 provides an overview of the literature on two concepts: learning in the laboratory and blended learning. In examining the research on the teaching laboratory in science education I introduce a number of common and recurrent problems. Namely the problem of ‘cookbook chemistry’, where students simply follow a set of instructions, without demonstrating any understanding of the reasoning behind their actions. The literature on blended learning explores the impact that technology can have on teaching and learning. Specifically, I suggest that web- and cloud-based technologies can enhance the teaching laboratory environment.

The following chapter, Chapter 3, provides a detailed overview of the design of Labdog. I open by discussing the relevant methodological issues and decisions I have made in the research, namely the use of mixed-methods and design-based education research. I then detail a series of interviews I conducted with UG chemists at the UoS to determine the student-perceived purpose of the teaching laboratory, to compare with the literature-based findings in Chapter 2. The chapter then closes with a description of Labdog’s functionality from both an abstracted and grounded perspective, introducing and then illustrating its functionality.

I conducted a series of pilots to establish how well Labdog performed in practice. These are detailed in Chapter 4, where I discuss how Labodg was used in a eight-week period with science foundation year students in 2015/16, and in two years of outreach events (2016 and 2017). The results show an improvement in the usability over time, suggesting that Labdog is a valid and usable technology, but with areas for improvement.

Following these improvements, Chapter 5 then details how Labdog was used in a year-long trial with SFY students throughout the 2016/17 academic year. The results from this case study suggest that Labdog is a viable and usable technology which promoted students’ focus on chemical concepts in the laboratory, drawing their attention to the relevant scientific concepts within practical work.

In the penultimate chapter, Chapter 6, I draw together a collection of data to comment on Labdog as a tool for learning. I use a pre-existing survey instrument, a custom-built survey, and students’ responses to questions in Labdog to comment on the extent of meaningful student engagement with Labdog. The findings support the notion that Labdog promotes conceptually-focused learning.

The final chapter summarises the work conducted, and draws both data and literature together to answer the proposed research questions (given in Section 1.6). The focus of a design-based research approach is to apply what has been learned to the wider theory (Herrington et al., 2007), and the chapter therefore focuses on how the research informs what makes good instructional design in the practical environment when a technology such as Labdog is used. Specifically, I focus on the implication of cognitive load theory and 'flow-state', allowing students to engage in activities without being distracted or overloaded.

1.3 Self-Regulated Learning

As briefly introduced in the context of the exploratory work: the memorisation of information is distinct from understanding it. Yet what distinguishes memorisation and understanding? How, if at all, does an individual approach both of these processes differently?

Such questions are worthy of their own doctoral research. One of the most pragmatic, though expansive, answers to these questions is the concept of Self-Regulated Learning (SRL). SRL is a fuzzy concept, i.e. it is without a single unifying definition, but can be broadly understood as the ability of a learner to identify their goals and work towards them. A more precise working definition for SRL is “the processes whereby learners personally activate and sustain cognitions, affects, and behaviours that are systematically orientated toward the attainment of personal goals” (Zimmerman and Schunk, 2011, p.1).

Research into SRL is founded on two axioms (Zimmerman, 1986): i) non-controlled conditions, i.e. not through laboratory-based psychological experiments, and ii) that SRL is a product of the interaction between behaviours, metacognitions, and motivations. From its origins, SRL has been a purposeful departure from controlled-conditions in education research, instead focusing on developing “the more comprehensive picture of classroom acquisition” (Edwards, 1986, p.305). It is therefore a complex and holistic area of research, without predicable cause-and-effect one might expect of more positivist research (Zimmerman, 1986) instead relying on paradigms from the social sciences (Bandura, 1986).

Namely, it draws on Social Cognitive Theory (Bandura, 1986, SCT), a paradigm for social science research. In his original work, Bandura states that an individual's

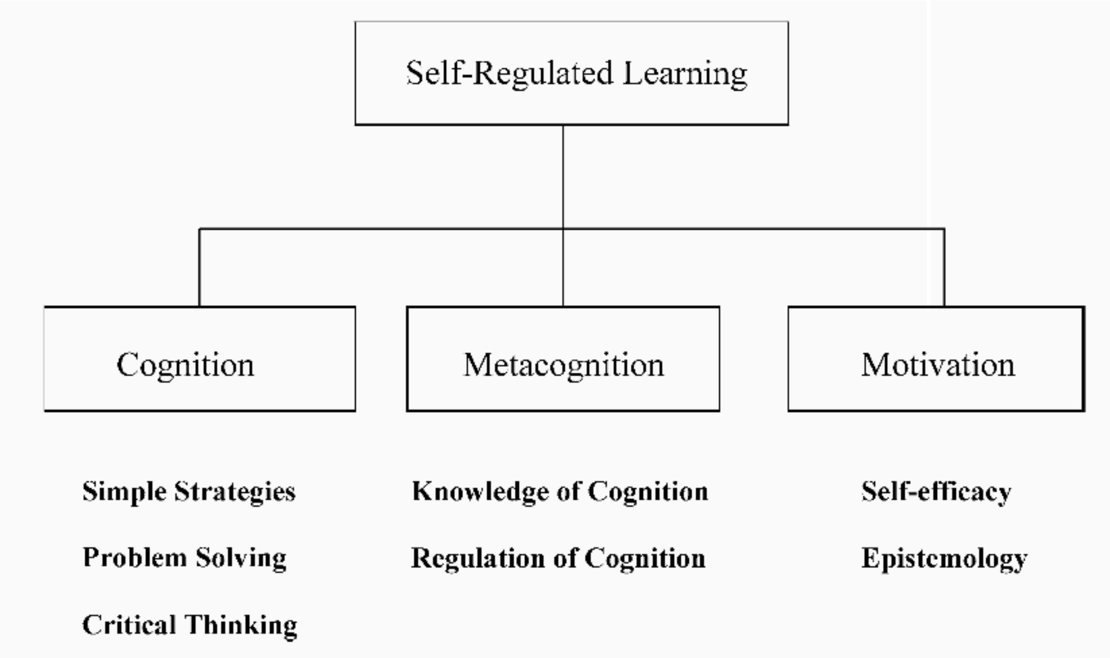


Figure 1.2: The components of Self-Regulated Learning, summarised by (Schraw et al., 2006).

behaviour, environment, and cognitions should be considered. The power of Bandura's theory comes from the diversity in possible interactions and the outcomes these can produce. Unlike simple cause-and-effect research, as one may expect in science, the SCT argues that we should not see elements of social-science research as uni-directional, predictable, or guaranteed:

> "Because of the complexity of interacting factors, events produce effects probabilistically rather than inevitably. In their transactions with the environment, people are not simply reactors to external stimulation. Most external influences affect behaviour through mediary cognitive processes. Cognitive factors party determine which external events will be observed, how they will be perceived, whether they have any lasting effects, what valence and efficacy they have, and how the information they convey will be organized for future use" (Bandura, 1978, p.345)

A model of Self-Regulated Learning

A number of researchers have provided conceptualisations and definitions of SRL. Schraw et al. (2006) present a model of SRL specific to science education, and their model is presented in Figure 1.2.

The authors attribute the model's influence of multi-directionality and non-determinism from Bandura's SCT. The authors also cite the work of Schunk (1996) and Zimmerman (2000) as influential in their definition of SRL as a process undertaken by students, and which progresses from comparison in a social context, e.g. comparison with and feedback from peers, into internal standards. Unlike an influential model proposed by Zimmerman and Campillo (2003), the authors do not explicitly show SRL as a process, but rather in the context of its three contributory factors: cognition, metacognition, and motivation.

Cognition. The authors characterise memorisation as the simplest form of cognition, and the application of knowledge to problems, and critical evaluation of knowledge as more cognitively advanced tasks. Each of these levels builds on each other, and can be combined to a diversity of learning learning activities and tasks. This is much a similar approach as given by Krathwohl (2002).

Being able so solve problems requires a high level of subject knowledge, encouraging students to learn quickly and effectively. Students must understand how to conduct relevant approaches to solving a problem. This is similar to the process of critical thinking, which involves identifying, analysing, and reflecting upon the source of information, i.e. their epistemic beliefs. The study of the relationship between epistemic beliefs and students' learning, specifically their goal setting (Muis, 2007), has become an increasingly well researched area of research in recent years (Muis and Franco, 2009). Developing an ability to critique knowledge based on these beliefs requires complex and long-term (e.g. several months) development. Despite the complexity and difficulty in acquisition, such skills are arguably quintessential to becoming a scientist, and also likely important in the development of metacognitive knowledge and awareness (Miri et al., 2007).

Metacognition. Metacognition is, simply, a student's knowledge of their knowledge. Metacognition has an established role science education (Thomas, 2012; Zohar and Barzilai, 2013), with Flavell (1979) proposing one of the first, and most ubiquitous, models of metacognition. The model has three components: metacognitive knowledge, experience, and, skill. Respectively, these refer to the theoretical ability of a student to assess their own understanding, the experience of doing so in a real-world scenario, and the efficacy when they do so.

Later foundational work by Brown (1987) divides metacognition into two types: knowledge of cognition, and regulation of cognition. Knowledge of cognition can be separated further into three kinds of knowledge: declarative knowledge, i.e. what can be said of our knowledge and ourselves as individuals; procedural knowledge, which is knowledge of strategies or processes, e.g. note-taking or answering problem sets; lastly conditional knowledge is an awareness of when a certain strategy should be used, and for what reason.

The process of regulating cognition involves planning, monitoring, and evaluation ones own performance or actions. This resembles the event-based models of SRL discussed by Zimmerman and Campillo (2003), wherein students must make ongoing and retrospective judgements of their learning and performance as part of problem solving. Regulation of cognition requires an understanding of the nature of each approach to a learning strategy, and it is therefore essential to consider the goal or purpose of an approach to problem solving. Some of these processes may develop subconsciously, and therefore be difficult for students to vocalise or attribute to any single situation or source - therefore producing a methodological difficulty or concern in measuring SRL.

Motivation. Motivation is "the relation of beliefs, values, and goals with action" (Eccles and Wigfield, 2002, p.110). The things which motivate a student to work have a significant impact on the nature and extent of the student's working activity (Pintrich, 2000; Zimmerman, 2000). Within the literature these motivators can be divided broadly into internal and external factors. Internal factors are those generated by the student or task, e.g. the chance to learn new knowledge or the satisfaction of problem solving. Contrastingly, external motivators are outside the learning task, and include summative grading or financial reward. A great deal of research exists on how motivation affects learners (Fiske et al., 2016), namely through intermediary processes such as the ability to utilise feedback (Mangels et al., 2006) or an increase in self-competence (Lavigne et al., 2007).

Zimmerman and Campillo (2003) identify how self-efficacy and epistemology can impact a student's motivation. Self-efficacy refers to a learner's perception of their ability to engage in, and persist with, challenging tasks. Development of self-efficacy is largely done through vicarious experiences, wherein a learner watches or discusses with another individual as they perform a task. Alternatively, students can use modelling, i.e. generating a step-by-step conceptual solution to a problem without actually solving the problem. Such an approach works in a similar vein, where focus is on showcasing the process of answering a problem as a skill in its own right. Additionally, generation and incorporation of feedback can improve motivation - as seen with the previous discussion of formative assessment.

Students' knowledge about the source, speed, simplicity, and certainty of knowledge (Schommer, 1990) may affect their motivations. These beliefs, so called epistemological beliefs, affect the development of critical thinking and the process of goal-setting (Muis, 2007; Muis and Franco, 2009) - another process central to SRL (Zimmerman and Campillo, 2003). However in this regard there is more need for work on context, e.g. chemistry, specific epistemological beliefs (DeBacker et al., 2008).

1.4 Teaching and Learning in Chemistry

Though the abstract definition of learning given by SRL provides valuable insights, it is important to consider the context-specific learning process. In the instance of this research: how are the teaching and learning of chemistry distinct from other fields, and is there a need for chemistry-specific education research?

Chemistry concerns the properties, behaviours, and changes of matter (Burrows et al., 2013), primarily at the atomic level - the atom being one of the fundamental building blocks of our universe. This includes the interactions between atoms, which can bond to form molecules, and the formation, destruction, and interaction of these molecules is the basis for all life and material on earth.

Interestingly, the study of chemistry long pre-dates the an ability observe, directly, the processes which happen at such a small spatial scale. Before 1900, thanks to the work of John Dalton, Amedeo Avagadro, Robert Brown, and J. J. Thomson (and many others) - the particulate nature of matter had become an increasingly accepted, though still contested, empirically-grounded scientific idea.

In the 1980s, over a century after the work of Brown and Avagadro, we were able to visualise the atom.[6] The work of these early scientists set the boundaries of what became modern chemistry, and established what it is to be a chemist. The exact boundaries of what is, and is not, chemistry are not a point for discussion here, though there are certainly overlaps with numerous other fields including maths, physics,

---

[6] using IBM's Scanning Tunneling Microscope, which was itself was a Nobel Prize-winning achievement.

biology, astrology, pharmacology, and medicine. Its application reaches even further into electronics, engineering, forensics, fashion, art, and countless other domains.

One of multiple sources of difficulty in the study of chemistry comes from the fact that even the most fundamental or ubiquitous chemical concepts do not resemble anything which we experience in our day-to-day lives. Yet often our real-world experiences act as the reference for our knowledge. For example, Treagust et al. (2003) recalls a teacher likening a cloud of electrons to a moving fan blade. Such a comparison implies that electron movement is predictable, uni-directional, and of a consistent speed. Unfortunately such characteristics are untrue of electrons, which adhere to quantum behaviour, being both wave and particle. The potentially irreconcilable differences between the classical and quantum nature of matter have been the subject of intense study, for the past century. To simplify the intricacies of such an argument, it would be deeply foolish to liken an electron to a fan blade.

The notion that something can be both a physical object with mass, while also being an abstracted wave, is not intuitively easy to understand. The discovery of quantum physics required delicate and careful investigation with nuanced models of thought, as well as empirical equipment. When a student looks at water in a glass they will not naturally intuit that it is made up of almost countless, practically identical, infinitesimal molecules, and the interactions thereof. This makes novel chemical concepts hard to introduce, and it often means that empirical observations or modern-characterisation techniques require a working knowledge of many chemical or physical processes.

Somewhat paradoxically then, interpretation of an observation requires an understanding of itself. Such difficulties extend through the classroom/lecture theatre, and the teaching laboratory. It is therefore unsurprising that, even in a Higher Education (HE) setting, students may arrive with chemical misconceptions which trace back to assumptions of the macro- and micro-scopic behaving in the same way (Taylor and Coll, 2001).

1.4.1 Johnstone's Triangle

Chemical educators therefore require a subject-specific framework to help them organise and present chemical information to students. Johnstone's Triange (or the *chemical triplet*), Figure 1.3, is one such framework. The framework was first proposed by Johnstone (1982), in a short, almost unnoticeable, article published by a chemistry teacher. Nevertheless, the model clearly resounded with teachers and researchers alike,

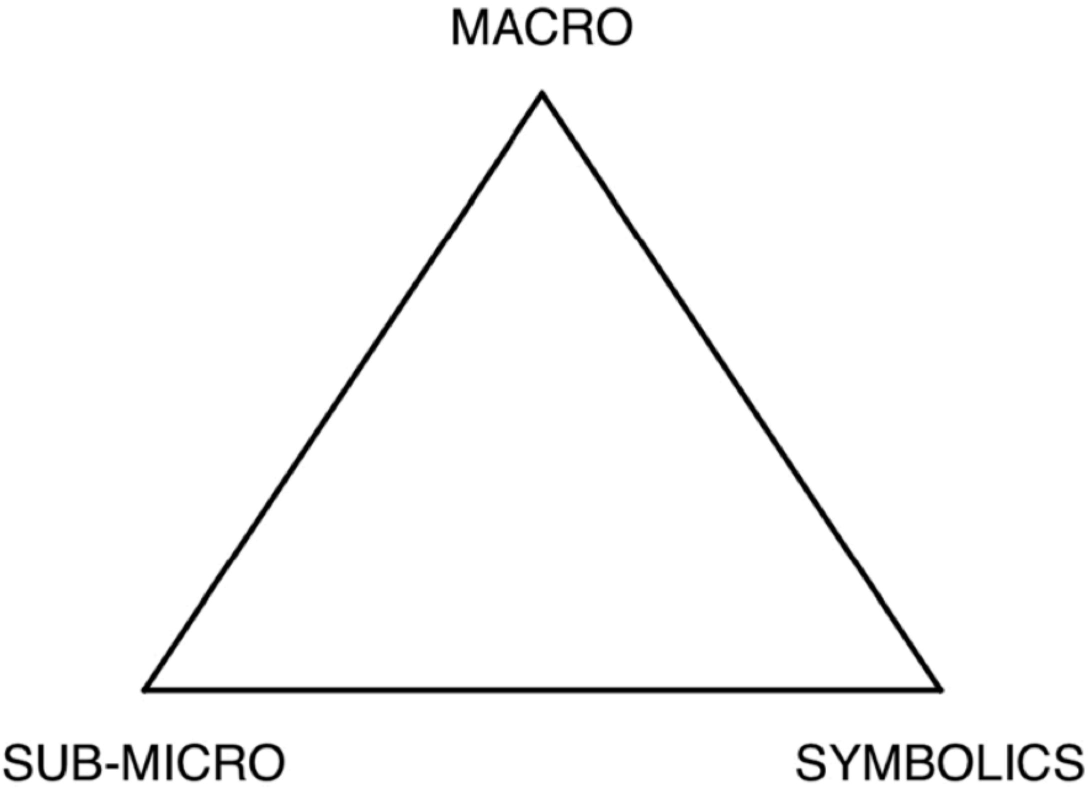


Figure 1.3: Johnstone's chemical triplet

and the simple idea catalysed a paradigm shift in the way educators and researchers think about the way we teach and present chemistry.

The chemical triplet proposes we consider chemistry as functioning simultaneously across three levels: macro, sub-micro, and symbolic. Taking water as an example: the macro would be the water we see in a glass or cup; the sub-micro is a water molecule, its shape, polarity, size, etc.; and the symbolic would be the way we represent these ideas visually, e.g. the structural formula H_2O, or the molecule in Figure 1.4.

Chemistry students must learn to move between and simultaneously handle these three distinct levels (Johnstone, 1991). Students must learn to conceptualise the idea of water as a fluid in a glass, as a single molecule's structure and physical properties, and to communicate these ideas through standardised nomenclature and convention. As Johnstone (2000) states, the need for simultaneous handling of three dimensions or levels is "at once the strength of our subject as an intellectual pursuit, and the weakness of our subject when we try to teach it, more importantly when beginners (students) try to learn it" (p.11).

In an analysis of teachers' body and verbal language, Stieff et al. (2013) found that when

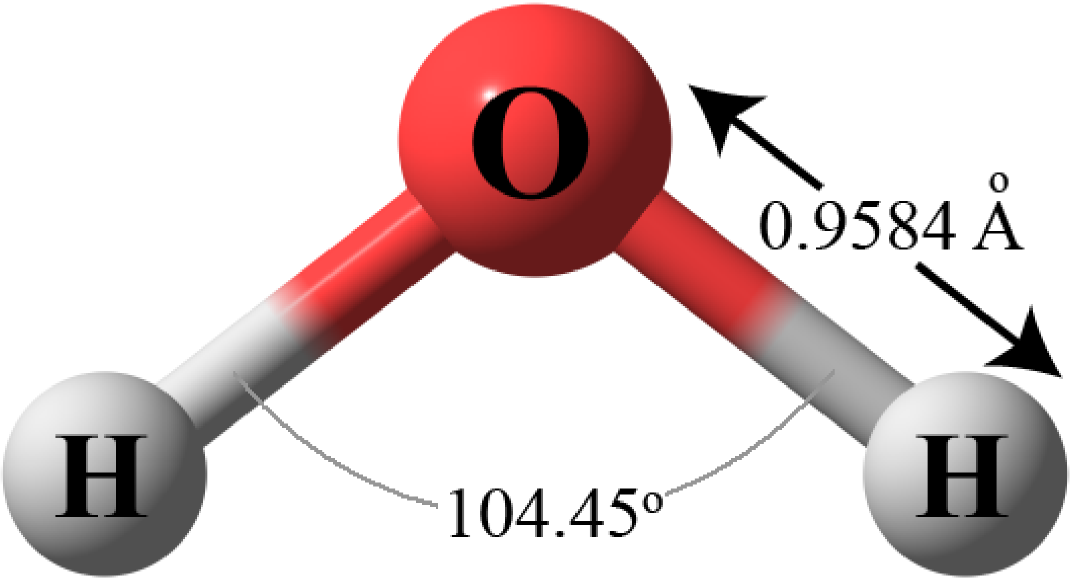


Figure 1.4: Water Molecule

teachers refer a chemical, e.g. water, they are non-explicit about which representation they are considering. They do not make it clear to students if they are talking about the macro water in a glass, the existent but almost invisible molecules of water, or a graphical representation (Figure 1.4). This non-explicit multidimensional nature of chemistry plagues students, despite their teacher's knowledge. As Bodner (2003) states:

> "Interviews with students struggling with organic chemistry has lead us to conclude that there is a fundamental difference between what the instructor writes on the blackboard and what students write in their notebooks, in spite of the fact that one seems to be a direct copy of the other." (p.13)

It is easy to imagine how this 'fundamental difference' between what the instructor is trying to communicate and what the students are interpreting disrupts effective teaching and learning. How then, can educators be assured that they are delivering the right material in an effective way to best communicate the relevant concepts?

1.4.2 Teaching Chemistry

An ability to move between the dimensions of Johnstone's triangle does not develop by accident, and must be the result of well-designed instruction. Good chemical instruction therefore requires educators to plainly present students with both a knowledge of concepts, and how they are presented. In turn, becoming a competent chemist requires that students are able to interpret the underlying chemical concepts, regardless of presentation (Bodner and Domin, 2000). Novice chemists must build an ability to identify what a problem is asking, and then quickly find the relevant approach and information required to solve it (Randles and Overton, 2015). In turn, students should use this knowledge and ability to solve and communicate effectively in chemical problems (Wheatley, 1984; Johnstone and El-Banna, 1986).

Educators may fall short in delivering this, by focusing too heavily on one specific representation, e.g. molecular or mathematical, and not introducing the links between them. Researchers from secondary education in Turkey presented students with chemical problems which contained either algorithmic, conceptual, or visual representations of the same chemical problem. The researchers found that a student's ability to answer a question with any one of the representations had no significant relationship with their ability to to answer either of the other types (Coştu, 2007).

In another case study, Treagust et al. (2003) observed numerous secondary chemistry lessons in Australian High Schools. The researchers suggest that teachers may underestimate, or simply not account for, the complexity of moving between the different chemical levels. The researchers recount an introductory organic class on structural diagrams, i.e. explicitly relating the sub-micro and the symbolic. The teacher introduces many additional concepts, e.g. energy, bonding, and movement, at the same time. They did not allow much time to communicate the very fundamental message that structural diagrams are a standardised medium for communicating the structure of a molecule. Such an idea is simple, in that it is a widely accepted and utilised convention in chemistry, much as a language would be. However this seeming simplicity of the concept being taught, more specifically the teacher's familiarity with it, means that fundamental messages are not being communicated to students effectively.

These arguments extend into practical instruction in the laboratory. As expert scientists, educators may implicitly assume that taking part in investigations will cause students to acquire an understanding of the concepts and practices through exposure of their actions (Abrahams and Millar, 2008). Even though educators may intend to develop students' conceptual and chemical understanding (Abraham et al., 1997; Bretz et al., 2013), such developments do not happen by accident, and require purposeful and targeted instructions (Zimbadi et al., 2015).

This relates to the idea of the Curse of Knowledge, a term which first arose in economics (Camerer et al., 1989). Since then the idea has been rediscovered under many names, including egocentrism, hindsight bias, or false consensus (Pinker, 2014). The Curse of

Knowledge is simply that one who holds knowledge finds it hard to imagine that another does not. Once we have learned something or acquired a new skill we underestimate the time it would take for someone else to get to the same understanding or ability that we are at (Kelley and Jacoby, 1996). This resembles the psychological concept of 'theory of mind', where an understanding of another's though processes, opinions, knowledge must be considered separate from their own (Frith and Frith, 2005).

Furthermore, assessment in HE chemistry has been criticised for relying heavily on a small number of unvarying questions that reward recall and pre-defined algorithmic methodologies (Bennett and Wilkins, 2004). Students are not being given unexpected or unfamiliar problems, or required to demonstrate conceptual understanding which underpins their actions. This is not only producing potentially unknowledgeable chemists, but may be contributing to the production of chemistry graduates with poor communication, leadership, innovation, and entrepreneurial skills, therein providing little value in the post-graduate workplace (Purcell et al., 2008).

As a result, chemistry teachers are facing a number of psychological or cognitive issues in themselves and their learners - which are exacerbated by the nature of chemistry itself. This difficulty rationalises the need to study chemistry teaching in specific, and to subsequently generate research which can aid both the teaching and learning process in chemistry.

1.5 A quick introduction to Labdog

Before presenting and discussing the research questions I wish to provide an overview of Labdog. As it forms such a central role in the research, it is important to have a functional understanding of the software before reading the questions.

Labdog as a web technology

Labdog is a piece of web software, meaning an application which is installed and runs on a web server. A web server is a computer[7] connected to the internet and, in the case of the vast majority of public services or applications, accessed using the HTTP protocol which make up the World Wide Web (www).

---

[7] There is an increasing popularity of 'virtual machines', wherein web servers are not defined by discrete physical computational devices, but rather as an elastic proportion of extensive networks of computers.

Users access the server using their own networked digital device, normally using a web browser, e.g. Microsoft's Edge, Google's Chrome, or Mozilla's Firefox. Some of the most popular technologies of the 21^{st} century are web technologies, e.g. Facebook, Google search engine, and AirBNB. Although each of these offers their own native applications on mobile operating systems, they can all be accessed through a web browser - which presents users with the same data and functionality.

There are numerous benefits to using web-based technology over traditional software, i.e. that which would have to be installed on a specific machine. By storing data in an online database, it can accessed across their devices and location, e.g. a home computer and classroom laptop. This is often referred to as 'cloud storage' (González-Martínez et al., 2015).

Secondly, web software is accessible across digital devices, e.g. smart phones or laptops, and operating systems, namely Microsoft's Windows and Apple's MacOS. As portable digital devices become more advanced, accessible, and distributed - traditional personal computers and laptops make up a decreasing volume of internet traffic. There are inherent caveats of usability, compatibility, and flexibility, but the trend cannot be ignored (Napoli and Obar, 2014). The use of a web-software means Labdog does not have to be concerned with compatibility to a wide range of operating systems, freeing up development time for user-facing features. Modern computers and digital devices increasingly come with pre-packaged web browsers, meaning Labdog is mostly accessible 'out-of-the-box'. This reduces the logistical and security implications involved with installing local software, of special concern in a school environment.

Labdog as a Laboratory Response System (LaRS)

Labdog combines influence from three types of educational technologies: learning management systems, e-portfolios, and classroom response systems. It is unique in doing so, and to the best of my knowledge is the first of its kind in the world. I therefore present Labdog as a Laboratory Response System (LaRS).

Broadly speaking, Labdog is a platform for digitising step-by-step practical activities, which are typically given to students on paper-based 'lab scripts'. Labdog is a platform for any step-by-step set of instructions, and is not a set of pre-made or lab-scripts in an existing system, e.g. a blog or spreadsheet. Labdog extends this functionality by hosting pre-, during-, and post-laboratory activities. In this sense, Labdog plays the role of a learning management system.

Labdog was designed to promote a continuum of learning, encouraging students to revisit concepts and ideas before, during, and after a laboratory session. This is where Labdog adds the majority of its pedagogical value, as it allows educators to pose questions throughout a practical activity. In pre-labs, these questions (re)introduce students to the relevant scientific concepts. Questions asked during a lab usually centre on some observation, methodological choice, or recording of data. Post-lab activities are a chance for students to reflect on their performance, or any methodological weaknesses or contradictions. Labdog facilitates both real-time and retrospective information on student performance and understanding. This data feeds into a framework for providing and reviewing formative feedback. This likens Labdog to a classroom response system (Beatty and Gerace, 2009).

All of this information and functionality is then stored in a single digital ecosystem, providing a logistical benefit for educators. This functionality resembles that of an e-portfolio (Ring and Ramirez, 2012), i.e. a digital record of students' actions and understanding.

The development of Labdog

Labdog was under near constant development from 2015-17, and a summary of the nature and timing of the the major changes and processes are shown in Figure 1.5. Labdog was first piloted in 2015/16, during which time a great deal of technical stability was added. A full academic year trial was run with science foundation year students in 2016/17. These pilots, and the resulting evaluation and improvements made to Labdog, make up the body of the rest of the doctoral research. This approach is termed a design-based research approach, and is discussed in Section 3.1.

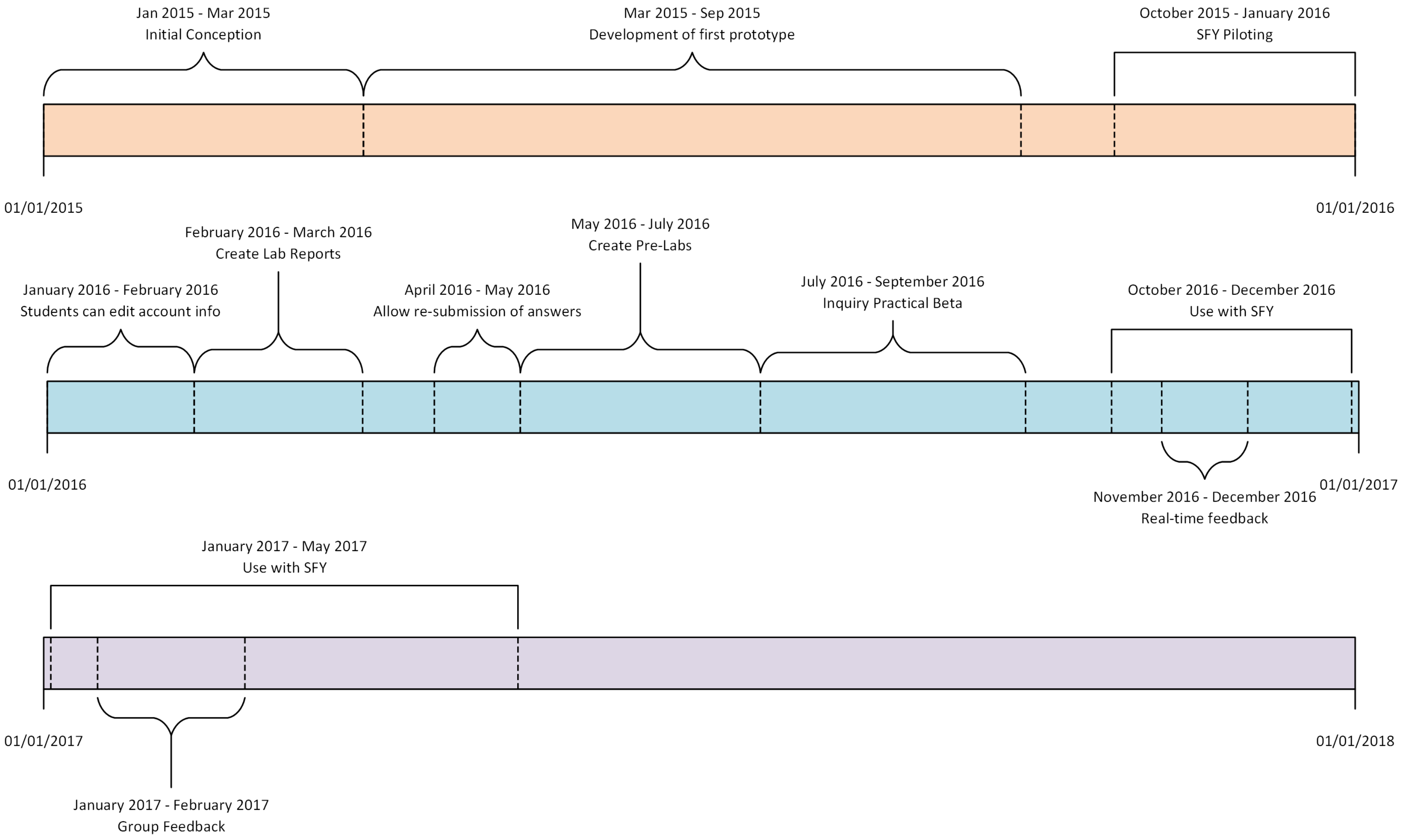


Figure 1.5: A summary timeline of the major development stages in Labdog's lifetime.

1.6 Research Questions

To accommodate for the exploratory nature of this work I adopted a design-based research (DBR) approach to conducting the research. I provide a full explanation of this in Section 3.1.2, but to summarise: DBR involves identifying a problem, proposing a solution, iterative cycles of improvement, and reflection (Herrington et al., 2007). DBR is intended to develop pragmatic evidence-based advice to improve teaching and learning in practice (Anderson and Shattuck, 2012). These considerations are captured in the following overarching research question:

> How does the integration of Labdog, a novel web-based Laboratory Response System (LaRS), affect the student experience and perception of learning in and from introductory-level practical-based chemistry education?

In turn, I generated a number of smaller, more specific, research questions, listed below:

1. RQ1: What is the student-perceived role of the laboratory in their chemical education?
2. RQ2: To what extent, measured through student-engagement, can Labdog be considered a technically stable and logistically viable educational technology?
3. RQ3: What is the student-perceived impact of adopting a LaRS-enhanced model of laboratory instruction?
4. RQ4: To what extent do students who use Labdog demonstrate engagement with, and understanding of, the relevant scientific concepts to a practical?
5. RQ5: What instructional design principles for a LaRS-enhanced laboratory environment can be drawn from the research?

1.6.1 Answering the research questions

As I will discuss in Section 3.1, I developed a mixed-methods methodology to answer these questions. This allowed me to capture and account for a broad range, and notable depth, in the data I used to evaluate each of the above questions fully. I rely notably on one-to-one interviews with a variety of university-level students, develop and deploy brief custom-made survey instruments, and adopt pre-existing and validated survey instruments.

When answering RQ1, I relied heavily on interviews with undergraduate students at the University of Southampton (Section 3.2). In a semi-structured interview protocol with 19 undergraduates across years one-four, I asked about their perceived purpose and benefits of the teaching laboratory in their current and previous education. The results revealed an inconsistent student-perceived image of the laboratory, suggesting that as many as half of the students explicitly did not see the teaching laboratory as beneficial to their understanding.

I answered RQ2 using usage and log data generated by Labdog. I used this across a variety of contexts, notably use in the Science Foundation Year bi-monthly laboratory sessions, and in two yearly outreach events at the university. I saw clearly that over time, student usage of Labdog increased, in accordance with ongoing improvements and design changes to the software. This data suggests that by the end of the research, Labdog was a stable and usable technology, as evidenced across multiple student cohorts.

To examine the impact of using Labdog, RQ3, I used a series of short surveys (Section 5.3) to provide rapid feedback from students. As the research progressed, I used longer-form surveys, and even interviews (Section 5.3.2) with Science Foundation Year students to assess the impact of using Labdog. The increased stability and functionality of Labdog (see above) meant that greater detail in analysis was possible later in the research. The results show that students report Labdog helped them focus on the chemical and molecular processes involved in their practical work, in particular valuing the in-person formative feedback from educators which could be facilitated by real-time data.

Using answers from students to question in Labdog, in RQ4, I found evidence that students were engaging with the relevant chemical concepts and processes when answering questions (Section 6.2). I found that inconsistent educator-given feedback in Labdog affected between a third and a half of student responses. However, it is likely that educators did not record their feedback in Labdog, and not that they did not give it in person. Additionally, I used a pre-existing and validated quantitative survey instrument (the meaningful learning in the laboratory instrument) to examine the differences in student-reported learning between foundation year and first-year undergraduate cohorts twice in the 2016/17 academic year (Section 6.1).

Lastly, I identify and discuss cognitive load theory and flow state (7.2.5) as relevant to designing good instruction with a LaRS technology (RQ5). It is important to strike the balance between challenging students conceptually, having them engage with the procedural processes at hand, and asking them to interact with novel software. Failure to do so may overload students' (extremely limited) working memory. In complement to this, students must also be encouraged to enter the flow state, where they are absorbed by the problem at hand, and allowed to progress through their actions. This must be balanced with the need to evidence understanding and completion of action, and also with inherently-demanding learning activities.

Chapter 2

Literature Reviews

The previously mentioned scoping work (Figure 1.1) raised two central themes: the role of the laboratory in science education, and the role of technology in education. This chapter provides an introductory literature-based discussion of these two central issues. Clearly, the full scope of both of these issues is outside the nature of this research, and the possibility of an introductory chapter. However, I firstly introduce the current understanding of the role, and related shortcomings, of laboratory work in science education. Reviewing this literature reveals that practical work has the potential to be a valuable tool in science education, but poor instructional design often prevents this from occurring in practice.

Secondly I discuss blended learning, i.e. how technology can be, and has been, integrated with HE instruction to provide improved quality of teaching and learning. The discussion shows that technology can be used to enhance instruction at any point from preparation for lectures/labs, to follow-up work. There is a growing movement to use technology in the laboratory context, however the evaluation in such research often does not touch on the pedagogical aspects of doing so.

2.1 The laboratory in science education

The popular image of a chemist brings to mind an individual working in the laboratory, using test-tubes, and wearing a lab coat. Just as a musician would tirelessly practice both their instrumental technique and performance - student chemists should be required to experience, and learn from, the laboratory. Justly, then, practical-based instruction[1] makes up a long-standing and significant portion of instruction in chemistry.

Practical work is the broad category of any instruction which involves a learner manipulating or observing objects and materials (Abrahams et al., 2013). Although computer simulations are increasingly being used within, or alongside, such activities (Kennepohl, 2007) - practical work has been, and remains a central component of science education since the 18^{th} century (Blick, 1955). Now, in the 21^{st} century, the UK government's stance on practical work is still strongly committed (House of Commons Science and Technology Committee, 2002):

> "practical work... is a vital part of science education. It helps students to develop their understanding of science, appreciate that science is based on evidence and acquire hands-on skills that are essential if students are to progress in science. Students should be given the opportunity to do exciting and varied experimental and investigative work." (para. 40)

However, orchestrating practical work takes a great deal of time, knowledge, space, and specialist equipment. Running a teaching laboratory may cost 15 times as much as lectures or classes (Reid and Shah, 2007), not to mention the costs of building and maintaining such a space. It is therefore important to have a more detailed understanding of the benefits which practical work brings, and if these benefits justify the proportionally large cost of practical work.

2.1.1 Why do practical work?

The literature on the purpose and benefits of practical-based work in science achieved moderate interest in the mid-late 20^{th} century. The field has recently undergone something of a renaissance, likely inspired by several researchers noting that previous research findings changed very little in practice (Abraham, 2011).

There is a great range of potential benefits from practical work. In a survey of 203 general chemistry teachers, Abraham et al. (1997) found that educators ranked 'chemical concepts' and 'scientific process' as more important than 'laboratory skills', as key benefits, with 'learning facts' and 'positive attitude' as less important benefits.

In a review of practical work in HE chemical education, Reid and Shah (2007) state that practical work provides "skills and insights which will be useful in numerous employment

---

[1] Throughout this section I will use practical- or laboratory-based instruction interchangeably

opportunities... where the student learner can be challenged to think, to argue, to weigh evidence, to explore chemical ideas" (p.183). Other authors also state that practical work can be used to develop students' understanding, as well as transferable skills (Carnduff and Reid, 2003). As a result, the range of potential benefits of the lab include the more abstract ideas such as interest in science, and creative and scientific thinking (Von Aufschnaiter and Von Aufschnaiter, 2007; Xian and King, 2017).

Some authors have provided more grounded investigations into the impact of increasing practical experience. For example, Samarapungavan et al. (2006), who themselves build on their previous work (Samarapungavan, 1992), examined the change in beliefs and attitudes towards science as individuals acquire more laboratory and scientific experience, as part of their ongoing chemical education. The authors interviewed 91 volunteers from an American university, using a semi-structured interview protocol.

The volunteers were classified as either 1) *Scientists* (N=13) who were members of the chemistry faculty involved in research, 2) *Grads* (N=22) who were Ph.D. chemistry students, 3) *Research trainees* (N=20) who were undergraduate chemistry majors involved in a research group, 4) *Undergrads* (N=17) who were chemistry students who attended a lab component to their degree, and 5) *High school* (N=19) who were students enrolled in chemistry at local high school.

The interviews lasted 50-60 minutes, were transcribed, and then analysed qualitatively. The interviews returned a wide range of results, however two stand out as important to this research context: the role of errors or anomalies, and the criteria for evaluating success.

Interviewees with more experience in the research laboratory (100% of scientists, 10% grads) were more likely to mention anomalies in results as a part of the research process. The same respondents also openly acknowledged the role of human and equipment error, as well as constraints from outside resources, e.g. funding, on the direction and focus of research. Not only were these individuals more tolerant of these influencing factors, but they also acknowledged their role in generating new unseen or unexpected areas of research. More commonly, those with slightly less research experience (90% grads, 50% research trainees) saw anomalies as common but as something negative or detrimental to research, and solely as something which should be removed or corrected. Lastly, undergrads and high-schools students all (100% each) saw anomalies as something to be ignored and covered up in write-ups, and that such mistakes rarely occur in the research setting.

Whereas all scientists evaluated their work in terms of personal cognitive criteria, e.g. difficulty or novelty, impact on humanity, and efficiency of the research process; most grads (68%) and some research trainees (10%) evaluated their work solely by cognitive factors, ignoring the wider implications of their work. The other grads and research trainees evaluated their performance by the timeliness of their performance and the satisfaction of their supervisor - an external source. Most undergrads (94%) and all high school students measured their success by ability to meet the assessments they were given, using their grades and teacher feedback as the sole important measures. It is interesting to note here the movement from internal to external criteria.

This provides valuable insight into the perceived role of the laboratory in the larger scientific process. Those with more laboratory and research experience report the laboratory as something which can generate new or novel ideas or interpretations, even if they are unforeseen. To this end, the laboratory is seen as more of an open tool, or source of information. Contrastingly, those with less experience see the laboratory as something to be completed and finished for its own sake. In these interviews, we begin to see the range of possible perceived roles of the laboratory to both the scientific process, and also the individual.

Many researchers have prioritised other benefits of practical work. In Table 2.1 I list several such benefits. I have divided these into three categories: i) hands-on benefits - the practical competencies related to working within a laboratory environment, ii) chemistry-specific benefits - the improvements to students' learning of chemical ideas, and iii) non-chemistry specific benefits, which are personal or affective aspects.

I have selected the authors featured in Table 2.1 because they have high citation count, a rough measure of their work's influence, and also because of a diversity in approaches:

- Trowbridge et al. (2000) adopt a pragmatic approach, seeking to improve the practice of secondary-school educators;
- Taber (2015) considers the pedagogical and theoretical literature;
- Grant and Jenkins (2011) report from a series of collaborative workshops with HE science and engineering educators about problems, experiences, and concerns with practical work.
- Hofstein et al. (2013) produce a "traditional list of objectives to be gained from including laboratory work into chemistry teaching" (p.155). This work most closely resembles a systematic literature review;

	Author				
Skill	Trowbridge et al. (2000)	Taber (2015)	Grant and Jenkins (2011)	Hofstein et al. (2013)	Bennett (2005)
Hands-on benefits					
Make and record observations/-data	X				X
Improve practical skills		X	X	X	
Demonstrate scientific theory		X			X
Chemistry specific benefits					
Organise knowledge and ideas	X		X		
Promote knowledge of nature of science and inquiry		X	X	X	X
Develop understanding of scientific concepts				X	X
Test or Apply knowledge		X			
Non-chemistry specific benefits					
Promote motivation, interest, or confidence		X	X	X	X
Problem solving and related skills (e.g. numericality)			X	X	X

Table 2.1: Various authors' proposed purpose of the teaching lab

- Bennett (2005) present a literature-based investigation intended more for policy-makers and educators in other fields, e.g. medical education.

There are clearly multiple potential benefits from using practical work in science education. Devising a structured list, e.g. an ontology or taxonomy, is outside the scope of this project - but would nonetheless be incredibly beneficial. By drawing this simple division in Table 2.1 I hope to stress that there are multiple distinct thematic types of benefits which can be gained from using the laboratory.

2.1.2 Cookbook Chemistry

Classifying the benefits, and then achieving them in practice are two distinct issues. Though educators may hold ambitious ideals about how practical work benefits students (Abraham et al., 1997; Reid and Shah, 2007), the realisation of these goals is often faulted. Students have become “accustomed to rote learning” (Bretz et al., 2013, p.281) in the laboratory context, simply seeking to memorise concepts without developing an understanding. Consequentially, practical work is not being used to its fullest potential (Carnduff and Reid, 2003) - as Hawkes (2004) comments, even “if labs are cleverly arranged and courageously taught, they can help in promoting interpretation and design of experiment but they are not useful in learning other aspects of chemistry.” (p.1257).

This is not an especially recent phenomenon: Goodlad (1983) examined the school science context, and reported that teachers intended to use practical work to help students develop argumentation skills, derive concepts from data, and test hypotheses. By observing school science teaching, the author found that, instead, teachers’ practice and activities would focus more heavily on having students recall information. Nearly 30 years later, and in the context of general chemistry, Abraham (2011) echoed this criticism, that educators “do not use instructional strategies that are most effective for teaching concepts” (p.1024).

In an early foundational piece of research, Domin (1999a) analysed the verbs present in 10 laboratory scripts. The results suggest that students are most often asked to utilise low-level cognitive functions such as recall, and are rarely asked to employ higher order functions, e.g. evaluation or creation. The authors note that such instructions produce a “‘cookbook’ in nature, with emphasis on following specific procedures to collect data and virtually no attention to planning the investigation or interpreting results” (p.109). Students who view instructions as straightforward recipes are more likely to have “overlooked the opportunity to consider the chemistry that afforded such simplicity in order to just adhere strictly to the procedure” (Galloway et al., 2015, p.235).

Domin (1999b) conducted one of the earliest investigations into the types of laboratory instruction, and documented four types:

- Expository: where students follow a detailed procedure,
- Open-inquiry: where students design a procedure to investigate a system,
- Guided-inquiry: where students receive a procedure but are required to collect and analyse their data,
- Problem-based: where students propose a solution to a problem using scientific knowledge.

Abraham et al. (1997) classified instruction differently: in verification versus inquiry. Students in verification labs attempt to verify or reproduce some previously known idea. Students in such labs reported that the activities helped develop their skills. Contrastingly, students in inquiry labs, i.e. where students must investigate some system, described more complex cognitive skills, such as the interpretation of data. In a later piece of work, Abraham (2011) expresses surprise over this finding given “the research history that has investigated traditional laboratory has come to the conclusion that there is very little evidence that traditional laboratory does anything very much except develop laboratory skills and teach factual information” (p.1021).

Despite the history of this research extending back some thirty years, e.g. (Hofstein and Lunetta, 1982), the disconnect between intended educational outcomes and actual instructional design appears to propagate. The work of Hofstein (Hofstein, 2004; Hofstein and Mamlok-Naaman, 2007; Hofstein et al., 2013), Abrahams (Abrahams and Millar, 2008), Abraham (Abraham, 2011), and many others (Galloway et al., 2015; Osborne, 2015; Walker et al., 2016) have maintained a research focus on the matter. Throughout nearly all of this research runs a similar narrative: that educators intend to provide complex and deep benefits to students, such as those seen in Table 2.1, but design instruction, and deliver learning activities which focus almost entirely on having students simply record and recall information.

In an investigation of 25 science lessons in eight English comprehensive secondary schools, Abrahams and Millar (2008) found evidence for this. The authors observed a mixture of biology, chemistry, and physics lessons - conducting interviews with teachers before and after the lessons, and interacting with students during them. In nearly all of the lessons observed, students were simply required to produce and observe a phenomenon. Some teachers used practical work as a chance to make something memorable, though the authors found little evidence that this resulted in it being better understood by students. Throughout the vast majority of lessons observed, the authors observed that students were not able to relate their observations to the underlying scientific theory, and that many educators simply were not encouraging or supporting this. The authors used students' language as a proxy for understanding: students who were able to explain a concept using colloquial or familiar language and then progress to using scientific terminology typically demonstrated a greater level of understanding.

In another recent study (Galloway and Bretz, 2016), 13 American undergraduate students taking an introductory chemistry lab module had a single laboratory session video recorded by a camera mounted by glasses to their face. The students were interviewed in the following 48 hours, and were asked to select times when they "had an 'aha' moment, [were] confused, felt lost, got stuck, [were] just going through the motions, and/or understood the majority of what was going on during the experiment" (p.6). When questioned about these moments, the vast majority of students paid little mind to the educational benefits of the lab, or the chemical principles behind their actions. For some of the interviewed students, watching the footage was the first time they had considered the rationale and underlying chemistry of their actions in the lab. The authors found many students focus simply on the actions themselves, e.g. setting up equipment, mixing, heating, etc. Furthermore, those students who did mention relevant chemical ideas often did so inaccurately or incorrectly. This suggests heavily that students will not intuitively, or un-prompted, consider the deeper conceptual elements of their actions or observations, even at an undergraduate level.

In another American study, Chopra et al. (2017) report from six interviews with undergraduate students in a General Chemistry (GC) 2 laboratory skills course. The prerequisite model (GC1) relied heavily on simply providing students with a series of instructions to follow. The authors redesigned the GC2 module, which subsequently contained small-group projects for which students had to design and conduct practical investigations, and then communicate their results orally or in writing. Phenomenological interviews revealed mixed feelings about the use of group work, specifically concerns about unequal effort between group members. However the redesigned course caused students to report a greater awareness and mastery of the underlying processes involved in the work, e.g. using a burette. The redesigned lab manuals also helped to make some students explicitly aware of the scientific process and method, which the educators paired with an instructional focus on ensuring understanding from students. Although this redesign represents a high time investment from the educators, the researchers found that "the overall experience of change is characterised by a transition from a learning environment that favoured mindless operation to one that nurtured mindful engagement" (p.122).

In a recent study, Xian and King (2017) examined the relationship between students' performance in exam questions and post-lab questions. Their results suggest that although students may be able to answer questions about their actions or experiment in particular, they often struggle to apply knowledge of the same concepts to new scenarios. The authors found that the variance in student performance between these two context was affected by both the conceptual area, e.g. kinetics or electrochemistry, and student ability. Specifically, low-mid performing students showed the lowest ability to cross-apply concepts. Student gender or their university major do not appear to relate to variance - suggesting that the impact of design and content of the laboratory work varies between student groups of different ability.

2.1.3 Correcting cookbook chemistry

Pragmatically, little more can be gained by bemoaning and evidencing cookbook chemistry. Clearly it can have a drastic impact on the educational outcomes of practical work. As Adams (2009) states, from a biochemistry perspective: "there is a pressing need to re-think the traditional approach to bioscience laboratory teaching... we must move away from 'spoon feeding' students during interminable, repetitive and boring practical classes that have highly predictable results" (p.1). Instead, I wish to examine several of the alternatives, or potential improvements, which can be used to address the shortcomings of cookbook chemistry.

Abraham (2011) distinguished two broad types of instructional design: inquiry and verification. In the above paragraphs I have described verification labs, which require students to simply confirm or demonstrate previously given chemical concepts. Inquiry laboratories, by contrast, require students to construct chemical concepts using investigation and associated findings. This shift into design, away from observation, addresses the problems caused by "too much emphasis on the *experiments to be performed* and not enough emphasis on what the *students should be gaining*" (Reid and Shah, 2007, p.177).

These authors suggest that clearly identified learning objectives and well-structured educational programmes are essential to delivering such benefits.

Millar and Abrahams (2009) clarify the need to have a ‘minds on’ approach to practical work, where students are cognitively engaged in their actions and observations. This complements the traditionally important ‘hands on’ approach. To do so, the authors state that educators need to carefully consider i) the proposed learning outcomes; ii) the related difficulty of completing the learning task, and iii) the presentation of the material and activity. The authors state clearly that educators need to consider the proposed learning objectives, and the measurable learning outcomes, so they can evidence what they believe students are achieving. Likewise, educators need to be aware if their planning activities are being presented, interpreted, and conducted in the way they intended them.

Other researchers have provided a vocabulary to help us explore such considerations. Von Aufschnaiter and Von Aufschnaiter (2007) contrasts teacher/content-centric and student/learner centric laboratory instruction. Here, teacher-centric activities involve activities which are familiar, easy, or simple for the educator. Student-centric activities focus on developing students’ knowledge and competencies. Similarly, Adams (2009) distinguishes student- and process-centric contexts. Student-centric activities aim to develop the competencies and knowledge of students; whereas a process-centric laboratory focuses on showcasing a scientific or chemical idea.

To better understand, and therefore promote, conceptual learning in the laboratory, instructional design needs to move away from simple verification and teacher-centric practices. The laboratory should be a tool for both improving and demonstrating practical competency and conceptual understanding, not just a chance to complete a practical activity (Von Aufschnaiter and Von Aufschnaiter, 2007; Millar and Abrahams, 2009). I therefore wish to briefly describe the differences between these two modes of instruction, and the consequences of their use.

Process-centric Laboratories

The benefits listed in Table 2.1 suggests that researchers consider practical work as a chance to promote students’ understanding of the nature of science, e.g. forming, testing, and then communicating a hypothesis (Hofstein and Lunetta, 1982, 2004). This need to demonstrate or model good scientific practices likely explains the choice to have students go through such an activity (Bretz et al., 2013).

Unfortunately, pre-university education focuses little on the design and conduct of practical activities, instead favouring observation and recording (Abrahams and Millar, 2008). Often educators are "focused on 'covering' knowledge of science topics and limited problem solving skills... laboratory activities have engaged students principally in following ritualistic procedures to verify conclusions previously presented by textbooks and teachers." (Lunetta et al., 2007).

Such ritualistic behaviour contrasts the nature of scientific research itself. It misrepresents both the broader scientific research process as devoid of creativity or innovation. It has little pedagogical benefit or foundation either, as Abrahams and Millar (2008) state:

> "The implicit assumption is that students will pick up a tacit understanding of what it means to plan and conduct an enquiry 'scientifically'. So their capability in science investigation can be tested at intervals, but does not have to be explicitly taught... this suggests that we still have some way to go in England... to develop models of practice in the use of practical work that more effectively integrate its roles developing substantive and procedural understanding" (p.1965)

Student-centric laboratories

Student-centric instruction focuses on the learner, and not the scientific phenomenon. Several researchers have argued for a more student-centric approach to laboratories, for a variety of reasons. These include the development of interest in (Grant and Jenkins, 2011; Hofstein et al., 2013) and motivation towards the scientific process (Bennett, 2005). Affective factors, e.g. emotions and motivations, are an important factor in developing and demonstrating meaningful learning (Bretz, 2001), including in the laboratory setting (Bretz et al., 2013).

A popular example of student-centric learning is project-based learning (Polman, 2000; Chopra et al., 2017, PBL). PBL is about presenting students with problems, for which they must then develop and propose solutions. This provides students with greater autonomy in the design of the processes carried out in the laboratory, while still involving the processes of design, analysis, and communication. As such, PBL can develop students' understanding of underlying scientific principles (Katchevich et al., 2013; Chopra et al., 2017).

Student-centric and inquiry-based practical work can also positively impact students' engagement with, and potentially conceptual gain from, practical work (Katchevich et al., 2013). There is a strong body of literature which advises the use of inquiry-based practicals to promote better student exploration of scientific ideas, as well as improve communication, co-operation, and understanding of the scientific process in chemistry (Hofstein, 2004).

Galloway et al. (2015) investigated students' enjoyment in their previously detailed thirteen interviews with American undergraduate chemistry students within 24 hours of completing an organic (n=5) or general (n=8) chemistry laboratory. The authors wished to investigate the affective, i.e. emotional, aspect of the laboratory experience. The students involved in the study were presented with a list of 18 affective words, split equally between positive (e.g. motivated and creative) and negative (e.g. nervous and frustrated). Students were asked to denote which related to their laboratory experience. The student sample reported heterogeneous results, suggesting that single experiments do not elicit single responses. The results suggest that the extent of a student's engagement goes on to affect their learning from practical work.

Abrahams (2009) completed a comprehensive study on motivation and interest in practical science education. The author distinguishes between motivation as 'an inner drive to action' (Bandura, 1986, p.243), and an interest as something which 'describes some preferences for objects' (Prenzel, 1992, p.73). The author used a mixture of observations, field notes, and tape-recorded interviews in eight non-selective secondary schools in England to examine the affective role of practical science lessons. Although the results found that students unanimously enjoyed doing practical work, this enjoyment was often limited to the practical lesson itself. As the author states: "for many pupils, practical work is perceived as distinct from, and separate to, science as a subject.... a preference for practical work *within* science did not always imply a preference for science over other subjects" (p.2345; emphasis in original).

This work highlights the nuance and complexity in linking two such broad factors as affect and learning. Nevertheless, it is clear that student-centric laboratory instruction is intrinsically linked to developing interest and motivation, though exactly how this affects the learning is unclear.

2.1.4 Pedagogy and Constructivism

It is important to ground this work's definition of the learning process in extant literature. I wish to briefly discuss the theory of constructivism, as the educational concept which underlies these processes.

Constructivism considers learning as a student-centric process, which involves an individual's active construction of knowledge, namely through interaction with process and people (Taylor and Coll, 2001). Constructivism stems from the early work of Jean Piaget and Lev Vygotsky. Today, it can be defined as:

> "constructivists believe that individuals re-structure the chaos of life to create meaning and order within their own worlds. In this model, teachers provide stimulation, guidance and extrinsic motivation in order to maximize the intrinsic motivation of our students and to help them to further their understandings and knowledge" (Goodman, 2014, p.10)

In a largely philosophical discussion, Ford (2008) comments heavily on the role of educators as authority in student learning. Ford argues that we should compare the processes of learning science to the construction of scientific knowledge itself. Under this paradigm, learners need to have an understanding of the methods behind the production of knowledge. We should therefore equip our learners with the tools to collect, analyse, interpret, and communicate data about phenomena - just as one would train a novice scientist. The opposite to this, and implied state in many cookbook, i.e. verification, laboratories, is the idea that the authority of the educator acts as an unquestioning source of knowledge, and that students are simply unquestioning recipients of these facts.

Other researchers have likened the nature of constructivism to that of chemistry itself. Namely as a subject where concepts and ideas are incredibly interdependent and related (Taylor and Coll, 2001). Such an approach must account for the complexity and diversity in linking observations to chemical principles and molecular processes, e.g. through simulation, video, and metaphor (Johnstone, 2000; Treagust et al., 2003).

Within the context of chemistry, constructivism can help students construct meaning of new concepts, using concepts they are already familiar with. Constructivism does not encourage students to simply recall a teacher-given definition or instruction. Instead, students should be able to explain novel concepts or specific examples using the underlying concepts which unify much of chemistry, e.g. the movement of electrons or the idea of charge. This idea of application has previously been identified as indicative of deeper understanding (Krathwohl, 2002).

Students with a constructivist view of knowledge are more likely to set intrinsic motivators, such as learning itself, rather than extrinsic goals such as summative performance (Muis, 2007). Intrinsic motivations are more likely to lead to improved academic performance (Vrugt and Oort, 2008), which is likely related to their strong link with more conceptually-focused learning (Glynn et al., 2011), and in creating self-regulated learners (Schraw et al., 2006, Section 1.3).

Despite these obvious benefits, many educators identify that students are unable to readily apply theory to practice (Grant and Jenkins, 2011). In order to promote this through constructivism, educators need to create environments which encourage individuals to interpret and confirm understanding on their own terms, and not just accept knowledge (Ford, 2008). This presents a need to strike a delicate balance between the didactics of telling, and the potentially misleading or over-simplified nature of open-ended enquiring or analogy:

> “If we wish [learners] to develop an understanding of the conventional concepts and principles of science, more is required than simply providing practical experiences. The theoretical models and scientific conventions cannot be ‘discovered’ by [learners] through practical work. They need to be presented. Guidance is then needed to help [learners] assimilate their practical experiences into what is possibly a new way of thinking about them. ” (Levinson, 2005, p.47).

Labdog’s principles of instruction are highly informed by constructivism. Constructivist instruction facilitates students in becoming active and engaged in their process of meaning making, instead on relying on what is easiest or *status quo* (Eilks and Byers, 2010). To this degree, the fundamental purpose of Labdog was conceived to a constructivist paradigm:

> “The issue is not whether student laboratory work has a place in chemistry teaching - it certainly does - but when the commitment of resources (in particular class time) to this type of activity is going to be post productive. Given the difficulty of learning to perceive what can be observe in the laboratory... it is important to ensure that what is observed during practical work can be actively linked in students’ minds with the theoretical descriptions and explanations” (Taber, 2015, p.81)

Student-centric generation of meaning is essential to effective constructivist instruction in the laboratory. The fact that such practice is often missing from practical work has been the source of criticism from numerous authors (Hawkes, 2004; Abrahams and Millar, 2008). Labdog was built around that idea that we need to encourage, and clearly evidence, the practice of actively linking representations, explanations, and observations. It centres around providing opportunities for students to construct accurate knowledge of chemistry trough using the tools, data collection techniques, models, and theories of science (Walker et al., 2016).

These statements about Labdog's technical capacity raises questions about the accepted or expected uses of technology in an education setting. In the rest of this chapter I describe and discuss how technology can, and has, been used to enhance teaching and learning, specifically in the contexts of HE, science, and the laboratory.

2.2 Blended Learning

Digital devices are becoming ever more ubiquitous and better connected. Digital technologies are already inseparable from day-to-day life in the developed world's economic, leisure, and social systems. Universities across the UK and US are creating and implementing institutional technological initiatives (Lewis, 2002) to promote effective integration of technology into the education, and research, sectors.

Integrating technology with teaching and learning is called e-learning. There is a spectrum of e-learning: at one end technology is used to facilitate distance or remote learning, i.e. decentralised from a single location. The most well-known example of distance education in the UK is the Open University. Distance learning is not a new idea, correspondence-education through textbooks or the radio emerged during the 20^{th} century (Guri-Rosenblit, 2005). Faster and more accessible computers, as well as the widespread adoption of the internet, have increased the potential of distance learning many fold (Lockwood, 2013).

Contrastingly, *blended learning* is “the thoughtful integration of classroom face-to-face learning experiences with online learning experiences” (Garrison and Kanuka, 2004, p.98). Bliuc et al. (2007) define blended learning as “learning activities that involve a systematic combination of co-present (face-to-face) interactions and technologically mediated interactions between students, teachers and learning resources” (p.234).

Blended learning there complements common HE practices, e.g. lectures or workshops. Therefore, there are innumerable ways to implement blended learning. Bonk and Graham (2012) draws the analogy of a book: to say that you have tried blended learning based on a single intervention is like saying that you have read and understand all books based on reading just one.

This does not mean that technology seeks to replace existing instruction, or teacher-student contact (Zare, 2000). Thanks to contemporary research, "the affordances blended learning offers are now well understood, and flexibility, ease of access, and the integration of sophisticated multimedia and technologies are high among the list of appeals" (Adams Becker et al., 2014, p.9). This is not to say that geographical, financial, and technological barriers do not exist to implementation and uptake (Lockwood, 2013).

Blended learning can improve instruction, assessment, and feedback - helping to address concerns over the dominance of lectures in HE, or the ever-growing demands on academics to both teach and research (Berrett, 2012). Both of these concerns are prevalent within chemistry at the University of Southampton, where blended learning has previously been used to increase student retention of information while using staff time efficiently (Brown et al., 2012).

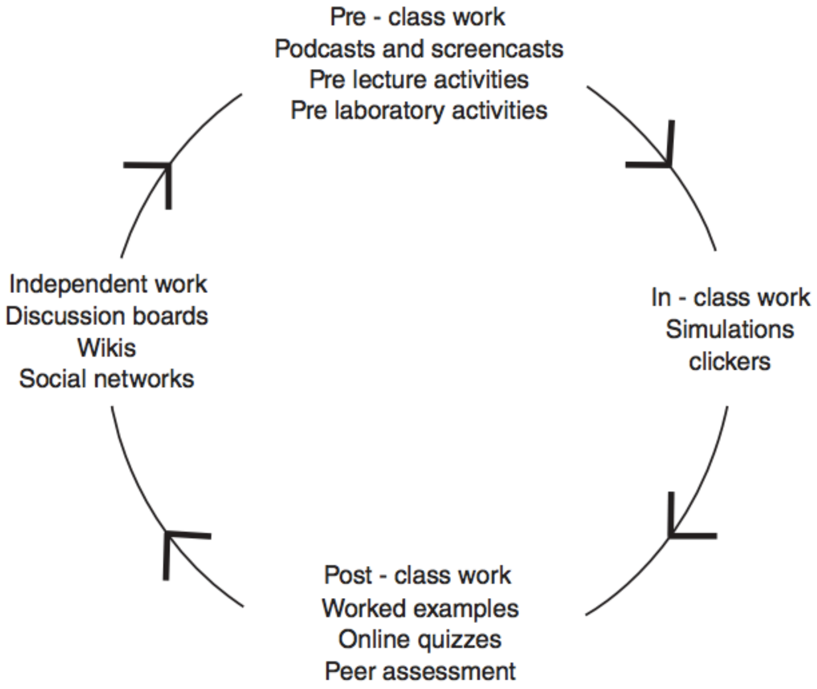


Figure 2.1: The blended learning cycle, a model for integrating technology and instruction, proposed by Seery and O'Connor (2015)

.

This diversity in approaches, represented in Figure 2.1, is a result of the many stages, pre- to post-session, where technology can enhance learning (Seery and O'Connor, 2015). Any approach should allow "the classroom [to extend] to the online space, and in turn the online space informs the classroom" (Seery and O'Connor, 2015, p.652).

Flipped-learning or -teaching is a particularly common example of blended learning. Flipped learning was pioneered in the early 2000s (Bergmann and Sams, 2012), and arose from concerns that i) student absence from lessons was having a disproportionately negative impact on the students' learning, and ii) students were not able to ask for help when they were struggling with homework problem-sheets. To address these concerns, the authors started video-recording themselves delivering portions of their class, e.g. explaining a single chemical concept. Students were required to watch these recordings as their homework, which liberated a period of the class where students could do traditionally homework-based activities, such as problem sheets.

Flipping has since become very popular, across subjects and educational levels (Fulton, 2012; Herreid and Schiller, 2013). This is understandable, given the the potential to increase learning in contact hours, and provide greater freedom to help students come to manage new concepts outside of the classroom. This can potentially increase amount of material that students can cover in a single chemistry course (Seery, 2013). Additionally the use of unobtrusive metrics, e.g. whether a student visited a web page or answered a question, provides educators with information on student engagement. In one chemical education study, these metrics have suggested consistent student engagement with digital resources (Seery, 2014).

In a cross-domain review of 24 research articles on flipped teaching, Bishop and Verleger (2013) found student and teacher perceptions of flipped teaching were generally positive. However there were often a minority of students who opposed it, instead preferring more traditional instruction methods. Dissatisfaction is often due to the use of technology (Seery and O'Connor, 2015). It is therefore important to provide high-quality and pedagogically-valuable face-to-face time to complement the use of technology (Williams et al., 2008).

In order to use blended learning to most effectively, it is therefore important to understand both the underlying pedagogical and logistical mechanisms. These are research areas which I cannot reasonably cover in full detail, so I therefore focus my discussion on literature which most closely resembles the Labdog use case. I examine how technology can facilitate formative assessment, question-based learning, and how previous educators have attempted to blend the laboratory environment.

2.2.1 Classroom Response Systems

The use of Classroom Response Systems (CRS) is a specific example of blended learning. CRSs are hand-held individual voting pads, colloquially known as *clickers* or *zappers* - the of CRS therefore involves the presentation and answering of questions, often multiple choice, in real-time during lectures. The digital nature of CRS means that a cohort's understanding can be quickly and easily quantified.

The use of CRS has expanded rapidly in the HE context, including science subjects (MacArthur and Jones, 2008; Beatty and Gerace, 2009), having notable roots in HE physics, though it has also been examined in chemistry (MacArthur and Jones, 2008). Physics education benefits from a series of concept-inventories, e.g. the forces concept inventory (Hestenes et al., 1992). Concept inventories are a set of published and validated multiple-choice questions which probe students' knowledge of one particular concept, and allow for accurate and specific measurement of understanding. This reduces the burden associated with writing specific and accurate questions, lowering the barrier for entry for physics educators, as well as providing better quality of information on students' understanding.

The assessing-to-learn (A2L) model is an early model of CRS-enhanced education, developed by physics educators at Harvard University (Dufresne and Gerace, 2004). In A2L, a class of students are posed a multiple-choice question which they then consider individually or in small groups. After an appropriate amount of time, the students submit their answer using CRS, and the results are presented immediately to both class and teacher. If necessary, student responses then direct classroom discussion of possible answers and afterwards, students can be asked to reflect on their answer or the discussion as a whole.

The peer instruction (Beatty and Gerace, 2009, PI) model of CRS evolved from A2L, and furthered the role of student-student and student-teacher interaction (Crouch and Mazur, 2001). Just as with A2L, students are given a question, discuss it in small groups, submit their answers using a CRS, and then discuss the responses. Unlike, A2L, students are then asked to find someone who submitted a different answer to them and to discuss the reasons why they answered as they did. Students must then discuss their responses to find an agreed answer between them. Students are then represented the same question and asked to submit their agreed answer.

There is no clear relationship between the use of CRS and an increase in student summative performance. Several authors report an association between the use of CRS, notably PI, and an increase in summative performance. In a piece reporting on universities and colleges who used the forces concept inventory and PI to teach physics (Fagen et al., 2002), 27 of 30 (90%) of educators reported a normalised learning gains of $\langle g \rangle = 0.39 \pm 0.09$ between the beginning and end of semester, which the authors report as a 'moderate' gain. [2]

The creators of the PI method report on 10 years (1990/01 - 2000/01) of using PI to teach a physics course at Harvard university (Crouch and Mazur, 2001). Over this period, PI was used with seven cohorts. The authors report a higher normalised gain in the forces concept inventory for students students who experienced teaching with PI ($\langle g \rangle = 0.48$) than those who do not ($\langle g \rangle = 0.23$).

Morgan and Wakefield (2012) conducted a similar analysis of PI within an introductory physics course for non-science majors in an American university. The authors examined individual students' responses to questions and associated them with exam marks. The researchers found no noteworthy correlation ($R = 0.04$) between students' course grades and the percentage of PI questions where they answered correctly when being given the forces concept inventory for the second time. The authors noted a much stronger ($R = 0.39$) correlation when students answered correctly on the first asking, regardless of their second answer. Interestingly there was a stronger correlation for females ($R = 0.48$) than males ($R = 0.3$) in this first asking. Unlike Crouch and Mazur (2001), these authors suggest that the use of PI does not necessarily affect summative performance. It is possible that differences in educational context account for this, i.e. comparing physics students at Harvard to non-science majors in a less high performing university.

Though such quantitative approaches help to argue the value of CRS, it is important to understand how they promote learning, qualitatively and experientially. This can help explain why studies with very similar teaching and assessment methods, e.g. the forces concept inventory, can yield notably different results.

Beatty and Gerace (2009) propose that CRS can facilitate Technology Enhanced Formative Assessment (TEFA), i.e. instruction where "the teacher works with the whole

---

[2] $\langle g \rangle$ is an accepted quantitative measure of normalised change in correct response rates. Here, $\langle g \rangle = 0.39$ represents a 39% increase in student correct responses between the beginning of the semester and the end.

class to help them make sense of new content material” (p.157). TEFA follows the pattern, which the authors claim is in-line with radical constructivism (Von Glasersfeld, 1984):

1. Pose a question to the students;
2. Students answer the question, alone or in groups;
3. Collect responses, usually through some technological tool;
4. Select and present a variety of answers to students;
5. Facilitate a student-centric discussion on the assumptions, perceptions, ideas, and arguments involved in each type of answer or explanation;
6. Use the information / misconceptions from students’ answers and explanations to feedback through a ‘micro-lecture’, or segue to another question.

TEFA is distinct from A2L and PI because of its focus on formative assessment (FA), the practice of assessing students in order to provide feedback on their understanding. This practice positively affects learning across ages, subjects, and abilities (Black and Wiliam, 1998, 2009), and can help develop metacognition and SRL (Nicol and Macfarlane-Dick, 2006; Clark, 2012). An increasing volumes of research is examining how educators can achieve these benefits Beatty and Gerace (2009). TEFA is a one such framework, presenting a technologically-enhanced, constructivist, and student-centric approach, relevant to this work (Bransford et al., 2000).

There are a number of more specific theories applicable to TEFA (Beatty et al., 2008), notably question-driven instruction (QDI), dialogical discourse, formative assessment, and meta-level communication. QDI places questions as the centre of instructional activities, promoting active learning and agile teaching Beatty et al. (2006). QDI encourages students to interact with, and therefore potentially change, their conceptions and understanding (Bransford et al., 2000), similar to the benefits and purpose of metacognition (Thomas, 2012).

The idea of ‘agile’, similar to just-in-time teaching (Novak, 2011). Just-in-time teaching is the idea that information on student understanding or progress should be gathered as close to class-time, or instructional delivery, as possible. This means that information is able to affect the focus, speed and nature of the teacher’s instructional practice. I have taken the term agile from modern software development practices (Martin, 2003), where similar evidence-based flexibility is highly valued.

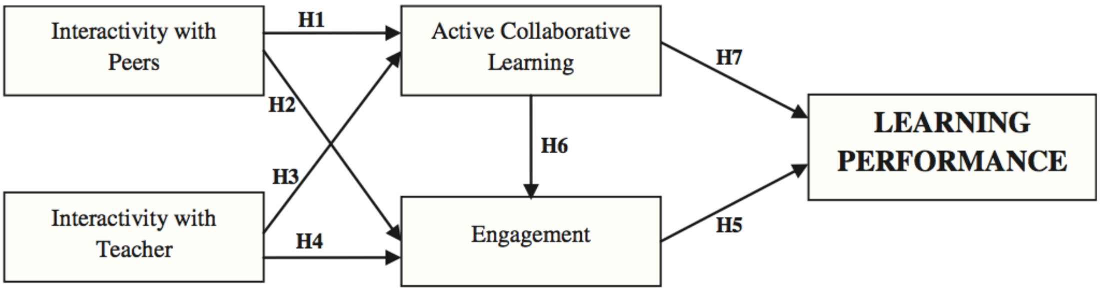


Figure 2.2: Conceptual framework for the mechanistic benefits of CRS-based teaching practice, proposed by Blasco-Arcas et al. (2013)

Blasco-Arcas et al. (2013) provide a quantitative model for the benefits of TEFA, shown in Figure 2.2. The authors suggest that learning from CRS comes from a combination of increased interactivity, engagement, and active collaboration. Interaction between and within student and educator is an essential component of the learning process (Mayer et al., 2009). Peer-peer interaction can help active processing of course material and promote higher order cognitive tasks (Crouch and Mazur, 2001). Such interactions lead to greater student engagement, or involvement, with the course material (Fredricks et al., 2004).

Active collaborative learning is about students participating, writing, reflecting, and otherwise being involved in the material, and is generally considered beneficial and essential in science education (Freeman et al., 2014). Collaborative active learning is a natural extension of active learning, where students work together in small groups towards a common of developing an agreed understanding.

Research by Blasco-Arcas et al. (2013) sought to support their theoretical model through a quantitative questionnaire to examine the existence and relationships between these factors. The method was used by the researchers in 2010/11 with 280 undergraduate marketing students in Spain. The group was taught using a CRS-enhanced model where, over a semester, students were given seven multiple-choice tests of 10 questions. Instruction followed an A2L model, where students were shown each question in the test, given 60-90 seconds to discuss the question, and were then answered the question using their individual voting pad. The class would then discuss the reasons for each answer given, before presenting the final answer.

The authors also quantitatively validated their model (Figure 2.2), by constructing an 80-item survey, with responses ranked on a seven point likert-scale. The survey was piloted with 15 students and was revised slightly. It was then administered to the entire cohort in January 2011, with a response rate of 71% (N=198). The authors conducted a confirmatory factor analysis and structural analysis. The results confirmed that each of the four factors discussed above were independent and related, which suggests their model was valid.

These findings develop our understanding of how formative and social aspects of CRS encourage student learning. There are also a few caveats to the research. Firstly, it is unclear how representative the 71% sample is by demographic, academic, or socioeconomic factors. Secondly, the authors claim to measure learning performance, but the items that they cite in their survey are about perceived understanding, e.g. "the use of clickers has allowed me to better understand the concepts in this model". Such phrasing is problematic as it requires an excellent metacognitive ability, specifically judgments-of-learning (Dunlosky and Nelson, 1992), and the accuracy of such estimations varies by student knowledge (Zell and Krizan, 2014) - presenting a complex feedback mechanism. While this may not affect the model of interaction, engagement, and active construction being linked to CRS-enhanced teaching, it does limit the strength of the findings on to academic / learning performance. In fact, the authors present no information on student performance or engagement with the A2L intervention.

Though there is no universal relationship to summative performance, the use of CRS can help students construct understanding of material, through increased social engagement and interaction (Carlsen et al., 2007). This principle resembles social constructivism, wherein students construct meaning through social interaction, reaching a shared interpretation and understanding (Adams, 2006).

2.2.2 Blended Learning and the Teaching Lab

Though there are no reported instances of using CRS-like technologies in the lab, others have blended the environment. This ranges from simple sharing of resources (Nilsson et al., 2010), to the use of virtual reality in the laboratory environment (Alkhaldi et al., 2016). In this section I discuss a number of such case studies, and how they have improved both teaching and learning practice.

Gaynor and Brown (2012) present an example of how technology can facilitate and enhance face-to-face instruction in the teaching laboratory. The authors present CHOBS, an online booking system, which students can use to indicate attendance at particular laboratory or feedback sessions - allowing educators to plan and use their time most efficiently. Although use of CHOBS was not mandatory, the authors found a high and consistent use throughout the academic year, also finding a weak correlation between its use and students' final grade. Though limited in pedagogical scope, such a case-study highlights how technology can help students and teachers use their time more efficiently.

Educators can also blend pre-lab instruction. Jolley et al. (2016) present a case-study from a second-year analytical laboratory in a compulsory course in Australia. The authors used a web-based system to deliver the lab-script, instructional videos, and pre-lab quizzes. Students were required to complete the quiz at least 24h before the lab session, and must achieve 80% within their first two attempts. The authors evaluated the intervention using a voluntary survey at the beginning (week 1) and the end (week 12) of the semester. $>\sim$ 90% of respondents reported that the videos and quizzes were useful their study. By week 12, 89% of students felt 'very' or 'somewhat' prepared for the laboratory, compared to other subjects where they did not have the pre-lab activities. The authors were unable to demonstrate a significant difference in student summative achievement in the laboratory reports, or drop-out rates.

To be educationally beneficial, pre-lab activities must contain learning activities focused on conceptual gain (Johnstone and Al-Shuaili, 2001). Unfortunately, the authors do not mention using information on students' understanding for formative feedback, or to shape the contents or direction of laboratory instruction. It is also important to acknowledge the issues caused by anonymous and voluntary surveys, lack of non-self-reported data on student usage of software, and little experiential or qualitative information from students. As a result, it is unclear how representative the data is of the student cohort, and how accurately it reflects reality - making it almost impossible to draw conclusions on software usage and impact.

Another case study is the creation of YouTestTube (McClean et al., 2016) - a video sharing web application similar to YouTube[3] in a UK HE bioscience context. The researchers required students to record, edit, and upload a video documentary for each of five laboratory sessions over 12 weeks. Students were also required to engage with other groups' videos on the web platform, e.g. through commenting and rating. The intention was to help students engage with, and reflect on, the chemical concepts involved in the particular laboratory, and the actions they undertook during the session. Secondly, the intervention was designed to promote social cohesion - with students working in small groups, and interacting on the website.

---

[3] The popular online video sharing platform: `https://wwww.youtube.com`

The authors relied largely on a pre- and post-semester survey to measure the impact of YouTestTube on students' experience and learning. They also interviewed three (of $N = 154$, 1.9%) students in 2012/13. Of the 98 (63% of cohort) students who completed the post-semester survey, the majority (62%) reported enjoying the creation process of the videos, and nearly three quarters (73%) reported that it helped them make links between the practical and theoretical aspects of the course.

Interviews suggest that storyboarding, i.e. planning the content and structure of the video, caused students to revisit their actions and the underlying theory. Re-covering previously seen or learned material is a strongly evidenced learning practice (Brown et al., 2014). The multimedia and project-based nature of the work also benefited students' digital literacies, communication, and social skills - such benefits are often touted by advocates of project-based learning (ChanLin, 2008), and resemble the holistic benefits which result from, and are desired, from the laboratory (Section 2.1).

Electronic Lab Notebooks

Other educators have integrated technology more completely with in-lab work. Notably, educators use electronic lab notebook (ELN), i.e. the digitisation of traditionally paper-based lab-books or -logs. Here, pre- , during- and post-lab work are all digitised, and often stored on external servers. This makes the data accessible from multiple digital devices, by both teacher and student.

ELNs allow educators to easily gather, compare, and provide feedback to students' work from the laboratory setting (Sinex and Chambers, 2013; Weibel, 2016). ELNs a tool which can be used in a variety of ways throughout the instructional process, as introduced in Figure 2.1. Furthermore, ELNs are gaining attention from researchers, due to the transparency, reproducibility, and interactivity they can provide (Shen, 2014). With additional considerations on the design of instruction, this approach can promote technological literacy, including code literacy, which are increasingly important for the STEM workforce (García-Peñalvo, 2016; Scaramozzino, 2010).

Many of the technical infrastructures which can facilitate ELNs are cloud-based, i.e. based on a collection of externally-hosted and -managed servers. The use of a unified platform can make student collaboration easier, by providing an agreed and accessible system, allowing students to focus on content over method (Denton, 2012). Cloud-based storage and computing are becoming increasingly popular in both academic and business settings, and provides a central flexible solution for data, e.g. results, lab-reports, feedback, or reading material (Pence, 2016). I discuss cloud-based technology in more detain in Section 2.2.4

In a case study of using Google Drive[4] as an ELN in a undergraduate physical chemistry laboratory, Weibel (2016) details how ELNs can be integrated throughout the laboratory instruction in pre- to post-lab activities. This highlights the diverse functionality available to educators: educators used the system as a repository to store and disseminate pre-readings and risk assessments, and for hosting an online word processing document for the lab report and results. The system allowed Weibel to easily monitor student progress during the days-long practical, whereas before they would require students to physically hand their work in at the end of a lab.

A similar case-study is presented by Amick and Cross (2014), who report on using Apple iPad tablets and the Notability app.[5] The 12 students in the study used a custom-built document template in Notability to create paper-like notes, using a mixture of photos, hand-drawn, and typed information to report on their work in the undergraduate chemistry laboratory at an American institute. Students would upload their notes (as a PDF) to Dropbox, an online document hosting service, at the end of each laboratory session - preventing students from editing their work after the fact. Educators would also use Notability to mark and provide feedback on students' work, which would then also be uploaded to Dropbox and made accessible to individual students. The author found very minor difference in summative performance between the 12 students who used Apple iPad tablets compared to the 60 students in four other undergraduate laboratory groups who did not. This difference existed in both lecture (average summative score 79% compared to 76%) and laboratory (77% and 74%). The differences in sample size, and lack of pre- versus post- intervention performance are clear caveats to the strength of this study's findings.

Van Dyke and Smith-Carpenter (2017) describe how they used the web-based note platform Evernote[6] as a ELN in biochemistry labs. The authors adopted a bring your own device (BYOD) model of instruction, meaning that students used their own devices. The authors praise the software's availability across devices and operating systems, and allowing students to link multiple forms of information, e.g. text reports, hand-drawn schemes, pictures, and empirically generated data (e.g. spectra files). The system allowed educators to access / preview students' work, which allowed them to provide

---

[4] `https://drive.google.com`

[5] `http://gingerlabs.com/`

[6] `https://evernote.com/`

ongoing and retrospective feedback, while removing the need to hand-in and carry paper lab notebooks.

The authors provide very little quantitative evaluation. In a voluntary survey they found 93% of students found the ELN easy to use outside of the laboratory, and that 70% of students would choose a ELN over the traditional paper-based approach. The rich features and cloud-stored data made it easier to input and relate multiple sources of data, which suited the needs of teacher and educator. While nearly 90% of students found the entry of qualitative information easier through the ELN than in paper-based notebooks, only 50% found entry of quantitative data easier, though this was easier with laptops. Lastly, the authors acknowledge that financial and infrastructural barriers can prevent widespread adoption for institutions and individuals.

2.2.3 E-portfolios and Badging in the Lab

The case-studies above have largely focused on reporting the perceived impact of an ELN system from either a student or teacher perspective. Many authors detail their use without a pedagogical mechanism or rationale for these works.

Fortunately, a subset of recent work has begun to investigate this problem. This work is concerned with the use of multi-media records to facilitate more representative and accurate assessment of student knowledge and competencies. This focus on assessment is essential to informing instructional and student practices (Seery et al., 2017). Traditional assessment practice is for students to submit a lab report or quantitative results at some point after the laboratory session has ended (Graham et al., 2008).

In order to improve both the validity and relevance of the assessed material, Seery et al. (2017) identify a method of assessment focused on practical skills and competencies. The use of multimedia, e.g. videos, pictures, and written accounts - are vastly more appropriate for accurately representing students' abilities. Gathering and presenting such evidence is broadly termed an e-portfolio. E-portfolios can be the basis for both formative and summative assessment (Ring and Ramirez, 2012), including self-assessment and reflection (Roberts et al., 2013). E-portfolios can provide students with a sense of ownership over their learning and encourage more active learning in and from their work (Nguyen, 2013). Through these processes, E-portfolios can be help develop the skills necessary to be a self-regulated learner (Zimmerman, 2000; Sandars, 2009).

Some researchers have developed the idea of e-portfolios into the related idea digital badges. Digital badges are online evidence-based verification that an individual has demonstrated some competency or completed some action. This seemingly simple notion comes with the underlying educational problems of clear definitions and transparency, and the technical problems of availability, reliability, transparency, and metadata (Gibson et al., 2015). The idea of digital badging has received serious attention for its potential in an open- and ongoing-learning context, garnering interest form the likes of Mozilla (2017).

Badges simplify the outcomes of a potentially expansive e-portfolio of data into a concise and standardised vocabulary. This means that badges act simultaneously as a goal or achievement, and also as the reward or acknowledgment (McDaniel and Fanfarelli, 2015). As a result, digital badges can motivate students to work towards these goals (Gibson et al., 2015), which includes elements of gamification - i.e. the use of incentives and rewards common to video games in the educational-setting (Deterding et al., 2011).

An early implementation of digital badging into the chemistry laboratory was the creation of a single badge to demonstrate student competence in handling and using pipettes in the laboratory (Towns et al., 2015). This early trial acted as a proof-of-concept that digital badging could be used in the practical teaching environment. Working with a large American cohort of 965 students, $\sim$ 90% of students were awarded the badge, of which the majority of students reported increased knowledge and confidence in handling pipettes as a result of the exercise.

Later work by many of the same authors (Hensiek et al., 2017) describe a scaled-up version. The authors wanted to “improve hands-on skills through more authentic and evidence-based assessment, to improve data analysis skills, and to lower the cost of laboratory equipment replacement” (p.30). To do so, the authors trialled the deployment of four badges: for handling a pipette, volumetric flask, and burrette, and also for effective calibration of data. The authors implemented badges in their non-major chemistry undergraduate labs at Purdue University, in the United States. To earn a badge, students were given a clear set of criteria and instructions, to which they had to submit a narrated video of themselves performing the related task. In the video, students had to state what they were doing and why.

The authors used a retrospective pre- and post-lab surveys, which they refer to as a ‘now-and-then’ survey, to measure student perception of the badging system. Specifically the survey examined how the badges affected student knowledge, confidence, and experience. The results showed that the students who used digital badging reported feeling more knowledgeable and confident, and reported a better experience in using the equipment. Additionally, during the period of investigation, the authors saw a reduction of approximately $3,200 in money spent replacing glassware, suggesting tangible improvements in their handling of glassware.

Similar work is now taking palce across to the UK context, e.g. Seery et al. (2017) at the University of Edinburgh. These authors looked at how digital badging could be combined with peer assessment, a progression from the previously-used staff assessment of videos. This brings the concerns about logistical and time commitments required to watch and mark up to hundreds of student videos, which can take 25-45 minutes to assess a group of 24 students (Hensiek et al., 2017). Furthermore, it incorporates the benefits of formative assessment into digital badging, bonding formative and summative assessment.

Seery et al. (2017) examined this in the context of three laboratory skills: titration, distillation, and making a standard solution. The authors describe a three-step process: prior to the laboratory session, each student would watch an example video of the technique being conducted. During the laboratory, students would demonstrate the technique to each other, while a peer used an observation sheet to check each step was done correctly. Students would then video themselves completing the technique (which was not done for the standard solution), and then review the video alongside the observation sheet, and then chose to re-shoot the video if they wished. After the laboratory, students upload their video anywhere online, and then linked them to the university's VLE. The authors used the same instrument as Hensiek et al. (2017) to measure students' knowledge, confidence, and experience. The findings further support the idea that digital badging and portfolio-based assessment can benefit each of these three elements. Students' responses suggest that the intervention was most effective at promoting knowledge of the distillation and burette procedures - which students were required to film and submit.

2.2.4 Cloud-Based Technologies

While some researchers have used technology in relative isolation to teaching and learning (Gaynor and Brown, 2012), others have used it to exclusively in pre-labs (Jolley et al., 2016). A growing number are integrating it throughout the teaching and learning process (McClean et al., 2016). The use of ongoing evidence-focused multimedia assessment tools into laboratory work seems to offer numerous valuable educational and logistical benefits. Such a broad approach fits well with the on-going nature of blended learning (Seery and O'Connor, 2015, Figure 2.1).

Yet this approach also presents a unique technological challenge, namely in letting educators and students upload, modify, delete, and persistently store a large volume of data which needs to be easily accessible. Technology needs to support creating educators and students, while maintaining a simple and secure workflow.

I see web-based technologies as the best response to this challenge. Specifically in this context, it is important to consider how such technologies store data across locations, e.g. school and home. This is referred to as a cloud-based approach, which simply means that data is stored on an external server, and not locally, e.g. on a specific computer or USB drive.

As testament to the popularity and potential of cloud systems, Microsoft[7], Google[8], and Amazon[9] each offer their set series of cloud services. Working on the economies of scale, these companies provide a small percentage of their considerable computational resource to users, in the form of 'virtual machine', i.e. an isolated segment of a single computer. This also means the resource are 'elastic', i.e. they can expand and shrink to fit the needs of the individual, meaning users pay for what they use and can easily expand infrastructure. The ability to expand and contract reactively, as well as outsourcing the maintenance and management of physical infrastructure, makes cloud services an incredibly enticing platform for developers, entrepreneurs, and users (e.g. educators) who wish to focus on a product or service.

Though cloud-based systems do not directly mean better, or even different, learning or instruction - they are affecting the business workflow in the UK HE context. Cloud-based systems are increasingly adopted by institutes for e-mail service, data storage, and platforms for VLEs or Massive Open Online Courses (MOOCs) (McDonald et al., 2010). González-Martínez et al. (2015) conducted an extensive systematic review of the use of cloud services by HE institutes. The authors report a number of institutional and educational benefits, namely the fact that the huge resources available in cloud infrastructure can host VLEs, MOOCS, or learning analytics systems.

Cloud technologies play a critical role in the student- and teacher-facing context. They are the basis for both the Zaption and SVW work which founded this research. Modern

---

[7] `https://azure.microsoft.com`

[8] `https://cloud.google.com/`

[9] `https://aws.amazon.com/`

VLEs provide rich and detailed (meta)data on students action and other metrics from third-parties, which can be used to inform and improve instruction design (Adams Becker et al., 2014). Other cloud-based systems focus on the storage of information, e.g. Google Drive or Dropbox, while more still enable individuals to create, access, edit, and share documents or information, such as Google Docs. Such tools can enhance educator (Rienzo and Han, 2009) and learner (Mehlenbacher et al., 2015) capabilities and actions in an educational context.

Chapter 3

Designing Labdog

Several educators and researchers have reported examples of how web-based technologies can ameliorate both the logistical and pedagogical concerns of laboratory instruction. However, many educators have applied existing general-purpose technologies, e.g. Evernote or institutional VLEs, to the laboratory context.

In response to this, I propose Labdog: a novel, purpose-built, web-based educational technology. This chapter details the design principles and process behind both the development and functioning of Labdog, describing how it was created and what it does. I open this chapter by grounding the work's methodology in existing, valid, and appropriate research methods. Most principally I introduce and explain the idea of design-based research, and explain the advantages of mixed methods research.

I then detail a series of interviews I conducted with UG chemists at the UoS. These interviews acted as a chance to examine the existing literature-based problems within an immediate educational context. The interviews examined the UG-perceived role of the teaching laboratory. In response, students offered varying, often shallow, descriptions of learning in the laboratory - justifying the need for a technology such as Labdog.

I close with a technical, and then an illustrated, description of Labdog's functionality. In the technical description I describe how Labdog provides pre-, during-, and post-laboratory facilities for both teacher and student. In the illustrated examples, I provide a series of screenshots from the software itself to help better visualise and demonstrate the functionality.

3.1 Methodological Considerations

Selecting a methodology appropriate for answering the research questions (Section 1.6), presents a staggering number of possible methodologies. It is important to select a methodology which fits the nature of the work, and is able to validly and reliably record the relevant data (Cohen et al., 2013).

There are two foundational and contrasting philosophies in social science research: positivism and post-positivism. Positivism is characteristic of the natural sciences, and centres around the manipulation of tightly controlled conditions. Positivism therefore relies heavily on quantitative data, and the construction of mathematical models to represent and reproduce phenomenon. Positivist research attempts to negate the influence of the researcher, i.e. a system should be behave consistently, regardless of who observes it.

In contrast, post-positivism acknowledges the (sub)conscious bias of a researcher, and the holistic or unquantifiable aspects of experience. Post-positivist methods, e.g. the use of long, potentially unstructured, interviews has recently become prominent in social science research.

These camps should not be considered ‘at war’, but as presenting different methods for approaching a problem. The idea of pragmatic research dictates researchers should use methods which are appropriate to the problem and context at hand (Dewey, 1938), and it is therefore impossible and unhelpful to label one approach as ‘better’ than another. Pragmatism is needed within this research context especially, which consists of a number of different aspects: the development, implementation, and evaluation of a novel educational technology. As such, there are no specific methods which can be easily adopted or applied universally across the research.

Instead, a research methodology must be created, one which is:

1. **Flexible.** To account for the development, implementation, and evaluation aspects, while providing appropriate and relevant evidence;
2. **Rigorous, reliable, and valid.** The data must accurately represent what is being investigated, and should be reproducible in similar research contexts;
3. **Unbiased.** The research must not assume that certain educational/pedagogical outcomes are inevitable, and allow for unexpected or null results to emerge;

4. **Realistic.** It must be logistically feasible, for all major stakeholders: researcher, educators, and students.

3.1.1 Mixed-methods research

Increasingly, education researchers are drawing together data from a mixture of methods, ultimately telling a cohesive narrative. This is the paradigm of mixed methods research: any research design which combines qualitative and quantitative tools (Johnson and Onwuegbuzie, 2004).

Quantitative research employs empirical methods, often relying on numerical data, e.g. observations or surveys (Cohen et al., 2013). Quantitative data can be used for analysis and modeling (Creswell, 2013), e.g. the measurement of student opinion through closed-answer survey items, analysis of computer log data, or summative performance (Boudah, 2010).

In Section 2.2.2, I described the work of researchers who wanted to assess the impact of using novel technological interventions on their teaching. The majority of these papers, (Jolley et al., 2016; Van Dyke and Smith-Carpenter, 2017; Hensiek et al., 2017) used quantitative surveys, often completed at the end of a semester.

Such surveys are frequently used in social science research (Bryman, 2006), and can be one-off, pre-post, or ongoing (McClean et al., 2016). Quantitative surveys allow rapid and concise measurement along pre-defined criteria (Cohen et al., 2013), which can provide researchers with specific information and allow them to better understand the impact of teaching, e.g. Galloway and Bretz (2015).

However, these approaches are constrained to pre-defined criteria, which does not allow unexpected themes to emerge (Creswell and Clark, 2007). Additionally they rely on student-reported factors, which may not always be accurate (Samuelstuen and Bråten, 2007). Despite these shortcomings, quantitative surveys represent a central component of this, and other, research projects. They require little time from respondents, and represent a standardised comparison within and between groups.

Contrastingly, qualitative research relies on experiential, personal, and openly subjective information. Open-response survey items, one-on-one interviews, or focus group are all examples of qualitative research methods (Lichtman, 2013). Qualitative research accounts for, and incorporates, the subjective and experiential nature of the researcher and the researched. Qualitative data acknowledges the context of the research, e.g. the time, content, and participants (Grbich, 2012).

Several extremely relevant research articles adopt a qualitative methodology. For example, Samarapungavan et al. (2006) conducted a series of open-ended interviews with individuals to examine how attitudes towards science change with experience. In another research project, Galloway and Bretz (2016) used interviews and video footage of undergraduate chemists as the basis for a series of interviews on student-perceived learning from practical work in the lab. Both sets of authors carefully designed their studies to examine specific and pre-defined aspects of learners' behaviours or cognitions. Although more broad, and likely complex, the resultant research data is less structured than numerical data, but can provide unique, rich, and unforeseen insights into the lived experience of individuals (Grbich, 2012).

Both of these approaches present merits to researchers. It is therefore unsurprising that some researchers design a methodology which allows them to integrate aspects of each, so called mixed-methods research. More specifically, a mixed-methods design is one which:

> "focuses on collecting, analysing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone." (Creswell and Clark, 2007, p.5)

This idea dates back to the 1850s (Erzberger and Prein, 1997), however it was contested strongly during the 'paradigm wars' of the 1970-80s, which saw a polarisation between qualitative and quantitative researchers (Gorard and Taylor, 2004). Some researchers, e.g. Sieber (1973), touted an irreconcilable difference between the two approaches, stating the two should and could not be mixed. Those who argued for quantitative data valued the standardisation and comparability it provided. Qualitative researchers argued for the need of nuance, idiosyncrasies, and subtleties in research. Almost ironically, those who argued for segregation from an epistemological standpoint, provided little evidence of understanding the underpinning philosophical and theoretical distinctions between the approaches (Sale et al., 2002).

Current consensus holds that the two approaches can co-exist, both providing valid data for research (Howe, 1988). There has been a reduction in the polarisation of approaches, with fewer researchers adopting extreme approaches at either end of the spectrum (Miles and Huberman, 1994). Ultimately, research methods must be guided by both the nature of the problem and the research context (Johnson and Onwuegbuzie, 2004), the previously-mentioned pragmatic paradigm (Dewey, 1938; Rossman and Wilson, 1985).

Mixed-methods allow researchers to observe a complex processes through many perspectives (Johnson and Christensen, 2008). This allows them to account for personal (e.g. affective) and contextual (e.g. social) information (Grbich, 2012). Balancing these factors with accurately collecting valid quantitative data is difficult, but possible (Johnson and Onwuegbuzie, 2004).

The nature of this doctoral project, as one which was conducted in a series of real-world trials and evaluations, benefits immensely from the flexibility of mixed-methods. Where quantitative data can provide a breadth of data, the qualitative insights can provide a depth of information on context and experience. In a review of the extant literature on blended learning, Bliuc et al. (2007) commented that research examining the effects of blended learning has historically neglected, and needs to improve focus on, being “comparatively more holistic or systemic in its focus” (p.242). Given the nature of the research project, I have adopted the principles of mixed-methods into my research design.

3.1.2 Design-Based Research

This research exists outside of controlled conditions, affected by the unpredictable and uncontrollable factors of day-to-day life (Hutchins, 1995). This applied nature is at the centre of the emerging research field of ‘learning science’, which is characterised as the applied study, of real-world learning (Yoon and Hmelo-Silver, 2017).

The development, nature, and focus of Labdog were all informed continually throughout the research. Specifically through three short-term pilots and a year-long investigation. Over these four scenarios I worked with two distinct student groups (A-level and SFY), and each pilot used an increasingly mature and stable version of Labodg, in turn providing targeted feedback for improving Labdog for the next (Collins, 1992).

A classical pre-post experimental design would be impossible in such a situation, because conditions would not be kept equal over time or across conditions. Further inconsistencies in student demographics, as well as both the purpose and functionality of Labdog made each context fairly incomparable and distinct.

This context it not unique within education research. Design based research (DBR) is defined by Wang and Hannafin (2005) as "a systematic but flexible methodology aimed to improve educational practices through iterative analysis, design, development, and implementation, based on collaboration among researchers and practitioners in real-world settings, and leading to contextually-sensitive design principles and theories" (p.6-7).

Similar to action research, in that it seeks to improve education through research (Hickman and Alexander, 1998), DBR is not just conducted by a teacher (Anderson and Shattuck, 2012), but with a researcher as well. Other similar paradigms include design-experiments, design research, development research, curricular studies, or instructional design (Wang and Hannafin, 2005; Reimann, 2011). DBR is differentiated from each of these methods through its use of mixed-methods, and its focus on understanding how research can be used to affect real world teaching and learning (Reimann, 2011).

DBR's distinction can be summarised in its need to produce 'theories that do real work' (Tiberghien et al., 2009, p.2277). These authors distinguish between grand theories and specific theories: where grand theories include overarching philosophies, e.g. epistemological concerns, specific theories can be directly applied to affect teaching and learning 'on the ground'. In discussing specific theories, Tiberghien et al. (2009) raise the 'two-worlds' theory for discussing physics, wherein students link theory to experiment through constructing models. Similarly, chemical education has the chemical triplet (Johnstone, 1982). The triplet is an example of a specific theory - because it informs the way instruction is designed, and therefore the learning which takes place.

This grounding in real-world settings means that DBR focuses on collecting and analysing a variety of data types (Brown, 1992). To do so, DBR relies on cooperation between researchers and practitioners so that changes to teaching practice can take place, and be evaluated (Collins, 1992). As such, DBR projects are often focused around a specific education context or intervention, adopting an iterative cycle of design and evaluation (Anderson and Shattuck, 2012). Wang and Hannafin (2005) clarifies this definition with five principles:

- **Pragmatic.** DBR should inform both theory and practice. Theory generated from the research is more valuable if it informs and improves practice.

- **Grounded.** DBR is conducted in real-world settings and should be driven by, and grounded in, relevant research, theory, and practice.

- **Interactive, iterative, and flexible.** DBR consists of an iterative cycle between analysis, design, implementation, and redesign. The designer(s) of both intervention and research should work with participants, namely teachers and students, and should allow freedom of changes from both designers and practitioners.
- **Integrative.** Research should be credible and rigorous, carefully designing the data and analysis for different stages, and following emergent issues or challenges.
- **Contextual.** The research process, findings, and changes should all be documented, and explicitly connected with the setting, results, and process. This includes providing guidance for applying theory into practice.

A DBR methodology is well suited to the proposed work with Labdog, given that it is a ‘significant intervention’ (Anderson and Shattuck, 2012, p.16) that seeks to improve teaching and learning practice, and justifies the use of mixed-methods. DBR should focus on interventions which “should be able to migrate from our experimental classroom to average classrooms operated by and for average students and teachers, supported by realistic technological and personal support” (Brown, 1992, p.143). By the end of the research, Labdog could be said to be a stable and usable technology, though far from a finished product. It would be more accurate to consider it as a prototype technology, distinct from a more complete intervention (Easterday et al., 2016).

The iterative nature of DBR is in-line with what Anderson and Shattuck (2012) call “research through mistake” (p.17), as well as the software development principles of ‘agile’ and ‘lean’ (Martin, 2003; Ries, 2011), which I followed throughout the development process. This is to say Labdog was continually developed and evaluated, in contrast to ‘waterfall’ development, where a software product is rigidly and complexly conceived and then produced, with little prototyping or evaluation, until a market-ready product is produced.

DBR has been advocated as appropriate for doctoral-level educational research, where the researcher wishes to improve teaching and learning in practice (Herrington et al., 2007). DBR suits the simultaneous design and implementation of Labdog as an educational technology which features “systematically engineering [the learning context] in ways that allow us to improve and generate evidence-based claims about learning” (Barab and Squire, 2004, p.2). Furthermore it complements the use of mixed-methods research, which I have previously justified (Cobb et al., 2003).

DBR is evidently appropriate to the context at hand, so how does one use DBR to produce theories that 'really work' (Tiberghien et al., 2009)? Herrington et al. (2007) propose a four stage framework:

1. **Analysis of the problem.** In which researchers identify and gain a detailed understanding of their particular problem. This was detailed in Chapters 1 and 2, where I identified this research as concerning the ubiquity of cookbook chemistry, and poor understanding which evolve from this. The analysis of the problem culminated with research questions (Section 1.6).

2. **Development of a solution.** Researchers should then identify a theoretical framework which guides them to address the problem. In this context, I have identified a number of relevant frameworks: SRL (Zimmerman and Schunk, 2011), technology-enhanced formative assessment (Beatty and Gerace, 2009), the chemical triplet (Johnstone, 1982), and DBR itself. At the broader level the work has been informed by formative assessment (Yorke, 2003), constructivism (Taylor and Coll, 2001), and psychological literature on the learning process (Cassidy, 2011; Freeman et al., 2014; Kornell et al., 2015). These frameworks construct a narrative of learning in which students engage with, and get feedback on, the scientific and molecular processes involved in their practical work. These frameworks culminate in the proposal and design of Labdog, detailed in this chapter.

3. **Iterative cycles of testing and refinement.** Researchers build and iteratively refine their intervention. I detail this process in both a series of one-off pilots (Chapter 4), and a year-long trial (Chapter 5).

4. **Reflection.** The process of producing design principles from the work. In Chapter 6 I discuss evidence on the quality and nature of learning, and in Chapter 7 I answer the previously-identified research questions, and consider how a Labdog-enhanced model of practical work can promote better teaching and learning.

3.1.2.1 DBR and educational technology

DBR is suited to work with educational technologies in particular. Working with educators to adopt educational technology as part of their regular teaching or assessment practice has proven difficult, despite extensive investment in providing equipment (Romiszowski, 2004; Chen, 2008).

To an extent, the hesitant adoption of technology is valuable, as Roschelle et al. (2010) state: “new technology without new curriculum isn’t worth the silicon it’s written on” (p.6). There are ethical and competitive advantages to adopting technologies in such a way that the benefits, to learners and educators, outweigh the costs. However, as I have discussed in the literature review on educational technology, the focus of much published research is shallow. Perhaps the most tangible impediment in overcoming this in the HE environment is the ‘publish or perish’ nature of academia. Research staff who are expected to teach will focus on simply producing a publication over novel or rigorous educational research (Reeves et al., 2005).

This ultimately means research cannot generate guidelines for better implementation of technology in classrooms. This problem necessitated and distinguished DBR (Brown, 1992) from previous research methodologies. Given the innovative nature of technology, in particular Labdog, it can be used to achieve tangible benefits to teachers and learners, as Reeves (2006) argues:

> “Educational technology is first and foremost a design field... thus our paramount research goal should be solving teaching, learning, and performance problems, and deriving design principles that can inform future development and implementation decisions. Our goal should not be to develop esoteric theoretical knowledge.... this has not worked for more than 50 years, and it will not work in the future” (pp.100-101).

Unsurprisingly, the pairing of DBR and technology is considered to lead to both the production of better theory, and the adoption of valuable educational technology practices (Wang and Hannafin, 2005). For example Roschelle et al. (2010) detail the development, deployment, and evaluation of a series of technology-based curricular tools in mathematics. The authors praise a DBR approach in the early stages of software development, primarily for its ability to consider the problem, tools, and context that the research takes place in. This allowed the researchers to develop software which capitalised on the most potentially beneficial areas, while pruning cumbersome or useless features.

DBR is favourable over alternative approaches, such as integrating extant technologies, because it views the use of technology as a process, rather than a goal in itself. As such, DBR can facilitate innovative research in educational technology, while promoting its use by practicing educators (Amiel et al., 2008).

DBR must therefore address the logistical concerns associated with its implications: issues of scalability, usability, and sustainability of the technology (Fishman et al., 2004). As a result, DBR can help produce more widely used educational technology, ultimately resulting in social good, in that they offer better education to more students Reeves et al. (2005). As such, DBR is a justified and valuable research methodology to adopt in an educational technology-focused research project such as this.

3.2 The undergraduate-perceived role of practical work

In Section 2.1 I discussed the shortcomings of practical work. The literature suggests that novice chemists are often exposed to cookbook-style instructions, with limited experience in scientific investigation (Domin, 1999b), resulting in practicals only illustrate previously-introduced material (Hofstein et al., 2013).

Much of the existing work has examined teacher- (Bruck et al., 2010; Bretz et al., 2013) or researcher- (Hofstein, 2004; Lunetta et al., 2007) held beliefs about the purpose and role of the laboratory. Little work has examined the student perspective, though I have discussed both Samarapungavan et al. (2006) and Galloway and Bretz (2016), who both do so. These authors primarily used qualitative tools to assess the impact of laboratory, and general chemistry, experience on students' learning, and broader attitudes about science.

To provide a focused, and in-context investigation into students' attitude on the purpose of the teaching laboratory in specific, I decided to conduct a series of interviews. In reviewing the previously described literature, I identified the purpose, motivations, and self-reported SRL of students to be of most relevance to the research.

I decided to interview UG students, not SFY students, as they represent a more typical chemistry student. Many of the SFY students progress to further study, though only 5-10% of the cohort will study chemistry. Furthermore, UG students attend chemistry laboratory sessions once a week in the first two years of their degree, compared to fortnightly chemistry labs in the SFY[1]. Additionally, I did not wish to have the use of Labdog interfere with student-held beliefs about the laboratory.

---

[1] In alternate weeks SFY students attend biology laboratories

3.2.1 Methodology

I designed a one-on-one, semi-structured interview protocol, which lasted between 20-60 minutes. Questions were divided into three sections:

- Section A: The student-perceived purpose of the teaching laboratory (4 questions).
- Section B: Conceptual and metacognitive benefits of the teaching laboratory (3 questions).
- Section C: Student motivations surrounding the laboratory.

Students from all UG years were invited to attend interviews by cohort-wide e-mail from senior members of staff within the department. I organised a time and place for each interview on an individual basis.

3.2.1.1 Generating theory from interviews: Grounded Theory

Qualitative data holds a valuable role in education research (Section 3.1.1). In order to generate theory from qualitative data I adopted Grounded Theory (Grbich, 2012, GT) - a method of interpreting personal experiences to develop wider theories or models, first posed by Glaser and Strauss (1967). These models can help generalise results or apply findings between contexts, e.g. from this work to the school classroom, or biology HE labs.

Glaser and Strauss proposed GT to address a perceived incompatibility between the dominant quantitative research methods at the time, and the sensitive and human nature of experience. By adopting aspects from both positivism and post-positivism, Haig (1995) argues that GT is in agreement with the scientific method:

> “grounded theory research should meet the accepted canons for doing good science (consistency, reproducibility, generalizability, etc.), although these methodological notions are not to be understood in a positivist sense... suitably reconstructed, grounded theory offers us an attractive conception of scientific method.” (p.1-2)

GT revolves heavily on the practice of coding, i.e. identifying, describing, and quantifying themes or patterns which emerge from qualitative data (Savin-Baden and Major, 2013). I coded the interviews in three stages: open, axial, and selective (Lichtman, 2013). Open coding is the process of identifying the main issues; axial coding the process of detailing a concept, or showing its complexity or various aspects; and selective coding is used when the researchers believed they have identified the core concepts to an issue, and wish to relate various codes.

Existing chemical education researchers have adopted GT. Markic and Eilks (2008) interviewed students and used GT to contrast student-teachers' beliefs about chemistry and other science fields. Similarly, Grove and Bretz (2012) used interviews and reflective essays with GT to examine how students approached learning organic chemistry. Both of these researchers describe the time required to conduct a thorough GT analysis, from the iterative processes of re-reading and adjusting codes (Glaser and Strauss, 2009).

3.2.1.2 Analysis

Between October-November 2016, I conducted 18 interviews with UGs, which lasted between 20-60 minutes. A full list of students interviewed can be found in Table 3.1. All interviews were recorded digitally and transcribed into a separate Microsoft Word document. The transcripts were imported into NVivo 11 for Mac, where I analysed them using GT. I started the analysis by examining each of the questions individually, and then all of the responses within each section of the protocol. This allowed me flexibility in circumstances when students would inconsistently move between certain themes within or between questions.

During the analysis I developed a series of hierarchical codes, which evolved through 'constant comparison' (Lichtman, 2013) of reading and re-reading. All themes were allowed to emerge from the data, and were not the result of a previously-made schema or theory, so called 'inductive qualitative analysis' (Burnard et al., 2008).

3.2.2 Results

The findings from the interviews revolved around two central themes:

1. The student-perceived benefits of the undergraduate teaching laboratory;
2. The nautre of learning in and from the teaching laboratory.

I discuss both of these areas in the following section. During the discussion I draw heavily on extracts from interviews, which I attribute to individual students (Table 3.1). In these extracts my words as the interviewer, are placed in square brackets: [for example this would be the words of the interviewer].

Student	**Gender**	**Year**	**Notes**
Alice	Female	4	MChem; Conducting organic research project
Chapman	Male	1	
Charlie	Female	1	
Charlotte	Female	2	
Elizabeth	Female	2	
Hayley	Female	2	Repeated her first year
Hemant	Male	3	Conducting analytical research project
John Joseph	Male	3	
Josh	Male	1	
Kahlan	Male	1	
Miraj	Male	2	
Nick	Male	2	Transfer from Warwick University
Nickhil	Male	1	
Par	Male	1	Former SFY student
Ruairi	Male	2	Transfer from Bangor; Previous voluntary industrial experience
Thomas	Male	1	
Toby	Male	1	
Will	Male	1	

Table 3.1: Details of undergraduates who were interviewed in Semester 1 2016.

Finding: Student-perceived benefits of the teaching laboratory

I used responses to questions in section A to identify five distinct types of benefits, listed below. None of the students interviewed believed there were no benefits to the teaching laboratory, though several expressed a dislike for it. The difference between the most and least popular type of benefit is small (4 students, 22%), suggesting that the themes which emerged were consistent and representative.

- **Procedural skills** (16): Developing skills related to the procedures or motor actions within a laboratory context, e.g. handling equipment or chemicals;
- **Chemical processes** (15): Seeing or contextualising chemical or molecular processes;
- **Conceptual learning** (14): Conceptual or cognitive tasks, from understanding and memorisation;
- **Personal experience** (13): Personal or experiential aspects;

Procedural Skills	
Develop practical skills	11
Learn how to use equipment	8
Handle chemicals	2
Knowledge of transferable practical skills	2
Work with equipment	2
Total number of students	16

Table 3.2: Students' mention of procedural skills developed by the laboratory

- **Professional skills** (12): Developing skills useful to future professional contexts.

The data suggests that the laboratory provides a relatively consistent series of benefits to chemistry undergraduates. The rest of this section rationalises and describes each of these five categories.

Procedural skills

The most common benefit of the teaching lab was the opportunity to use and interact with materials or equipment. This can be broken down into a number of more specific benefits, seen in Table 3.2.

The students' definition of equipment encompassed sophisticated analytical machines, e.g. NMR, as well as more simple equipment such as hazardous materials or simple glassware. Many students gave vague statements, e.g. that the laboratory "gives you practical skills" (Chapman), or "I'd say the main purpose behind the teaching labs is just to give you the lab skills" (John-Joseph). Several students provided more detailed responses, however, for example:

> "And I guess it's mostly just little things, not anything big, just little bits of knowledge here and there that you just retain I guess. [Do you feel that's just theory knowledge, or more practical knowledge?] I'd say mostly practical [could you give me some example?] Cleaning equipment, using equipment, how to set up a reflux condenser or whatever, how to do a lot of the techniques, was really helpful" (Ruairi)

Several (8) students stated that they benefited from learning how to use the equipment. These students clarified that the laboratory didn't just let them use equipment, but rather gain some skill in it. For example:

Chemical Processes	
Place knowledge in context	7
Put knowledge into practice	7
Visualise a chemical process	4
Collect and gather data	2
See a process happen	2
Complete Procedures safely and effectively	1
Proof that a process happens	1
Total number of students	15

Table 3.3: How students mentioned the benefits of chemical processes in interviews.

> “It’s helped quite a bit with knowing how to use the equipment, and getting an understanding of that and the risks and stuff, like how careful I can be with some stuff. Before we did labs I didn’t know what most of this stuff was, I’d never really used it, but now I kind of get a better understanding of it” (Miraj)

Two students also mentioned that they had the opportunity to handle equipment and chemicals. While this is doubtless a foundational skill for a practising chemist, the pedagogical basis is limited. Only two students mentioned that physical experiences were transferable - suggesting perhaps that the majority of students see the benefits of the lab in isolation to particular investigations or events.

Chemical Processes

Many students mentioned the laboratory as a chance to interact with chemical experiences in the real-world, i.e. outside of theory or lectures. A frequency table of how students interacted with these processes can be found in Table 3.3.

A distinction between putting knowledge into context, and putting it into practice, became apparent. Putting knowledge into context is a passive action, where students better understand an action or concept in relatinoship to the wider chemical theory. Many students described that they used the laboratory to observe something, e.g. “you need to see how they work in real-life” (Charlotte), “It let me see the ideas that we learned in theory” (Josh), or “seeing it in a different light, kind of seeing it in a practical aspect” (Thomas).

In contrast, other students mentioned actively putting knowledge into practice. For example, Hayley stated “it gives me a chance to put that into practice and see ‘okay

I've learned about this and now I can actually do it' and go back and look at the theory behind it", or Josh's statement that "Rather than just being able to know a fact and reel it off in an exam, you'd actually be able to apply it - know what it really means rather than just knowing the first level".

These two ideas are related, and are not mutually exclusive: three students (Josh, Ho, and Nick) used both active and passive language during their interviews. Nevertheless, it draws an important distinction about the student's perceived autonomy, engagement, or control over the chemistry during a laboratory. A similar finding was found by Galloway et al. (2015), who identified varying levels of 'control and responsibility' even within the same teacher-led instructional environment. The authors found several students feeling either responsible or engaged in their work, while others felt this inhibited the control they could exert over their learning and actions. Students who engage with the material actively are demonstrating a better learning practice (Freeman et al., 2014), which is associated with more complex and desirable cognitive skills (Krathwohl, 2002). Unfortunately, interviews did not provide enough information to construct a narrative around why some students felt actively and engaged in the laboratory, while other were more passive.

Related to this idea of engaging with the concepts, a small number of students (4) mention visualising a molecular theory or process, i.e. the submicroscopic (Johnstone, 1982). For example:

> "It helps me visualise the theory that otherwise I actually have a lot of trouble doing. [What do you mean by that?] Sometimes the lecturers talk so fast about different ideas and they paddle through them, that it's good to take one out of context and analyse it and understand it. I wish I could do it to all the idea" (Ruairi)

It is clear that Ruairi benefits from the self-paced nature of laboratory instruction, as it allows him to construct a mental model or visualisation of the chemistry at his own pace. Ruairi suggests that the laboratory environment allows him the opportunity to focus less on abstract theory, and more on the application. This is supported by Charlotte, who stated "I think the idea was that it will help students to see it happen or to make it happen rather than having to make a table [of results]". This may suggest at a difference in cognitive demand on students in lectures compared to the laboratory (Paas and Ayres, 2014), where the laboratory is not so cognitively demanding, potentially allowing students to dedicate cognitive resources to learning activities (Seery, 2014).

Conceptual Learning	
Advance understanding	8
Reinforce previously-taught material	4
Make material more memorable	3
Advantages for assessment	2
Problem solving	2
	14

Table 3.4: How students mention using the laboratory to benefit them conceptually.

Conceptual learning

A distinction in the underlying cognitive processes emerged as a theme in its own right, and detailed in Table 3.4. Several students reported using the laboratory as a chance to further their understanding of chemical concepts and ideas (‘advance understanding’). This advancement in understanding may be attributed to students’ direct observation of the chemistry, the idea that it was “easy to understand and quite cool because you could see the chemistry happening in front of you” (John-Joseph). Several students furthered this idea, stating that they learned not just from being able to see it, but to interact and participate:

> “but if you’ve done it by hand, following a book or whatever, then you do know what’s going on, so I think practical work is a human way of learning” (Hemant)

> “I’m one of learners who learns when they do things instead of reading things, so practicals are really useful for me.” (Par)

These comments are in contrast to several others which reported the laboratory is a useful tool to help them reinforce, or better remember, previously-learned material. Similar to the above, for some students this is strongly related to the observability of the chemistry, for example:

> “and it helped in the exams to remember a practical that we’d did based on that piece of theory... [What do you mean by helped you remember in exams?] We did oxidation of primary alcohols, we did a practical based on that, so we oxidised an alcohol to carboxylic acid, and if that came up in the exam we could remember the colour-changes that would happen because we’d seen it first-hand.”

Personal Experience	
Gave personal satisfaction	5
Group work	4
Enjoyable	2
Experience unplanned or outcomes or events	2
Gave a break from lectures	2
Increased confidence	2
Increased initiative	2
Number of students	13

Table 3.5: Students' mention of personal or experiential factors in interview

One specific subset of responses identified remembering for an exam scenario. One student referred to pre-university A-level coursework assignments (Thomas), while another referred to it in the university context, stating "it helps when you're sat in an exam if you've made it yourself" (Toby).

Very few students mentioned using the laboratory for higher-order cognitive skills (Krathwohl, 2002). Two students mentioned that they enjoyed the problem solving aspect: "it's sort of like a puzzle: get an unknown and then you do all these little tests to see which one is actually your solvent" (Nikhil). Individual students mentioned aspects of reflection or metacognition, however these were one-off and so not included in the table above.

Personal Experiences

As well as benefits to learning and cognition, the laboratory has a strong experiential and affective benefits, listed in Table 3.5. About a quarter of students mentioned that the laboratory provided them with some degree of personal satisfaction:

> "It feels good when you get good results or like even if it goes wrong when you kinda figure out when it goes wrong, why you figure out why it goes wrong because it's sort of an accomplishment." (Alice)

And several others appreciated the small-group work:

> "I think human interaction makes people pay more attention rather than just writing on paper, or filling something in on a computer... You're working more with your classmates than in your lectures, you're more independent, so you get more social skills." (Par)

Professional Skills	
Experience the practice of a professional chemist	5
Develop employable skills	4
Improve organisational skills	3
See what would be required of a professional	3
Number of students	12

Table 3.6: Students' mentions of professional skills in the laboratory.

The laboratory reportedly helped a number of students develop autonomy, confidence, or initiative, e.g. "labs actually forced me to embrace that side of me, and say 'you can do this' so that especially... helped me to gain more confidence" (Elizabeth). Others students simply enjoyed the labs, or found them a welcome break from other contact hours, such as lectures: "It's nice to have a bit of a change from being in a lecture. So it splits up my week a bit" (Toby)

While the total number of students who felt the laboratory-based component of their education benefited them on a personal level is not much lower than the above cognitive and procedural elements, and there is less overlap. Where previous results (Tables 3.3 and 3.4) have featured fewer thematic categories with higher number of students reporting each, experiential elements appear to be less unanimous and have smaller overlap, i.e. students who report more than one theme.

The role of affect in education is a point of great discussion, which simply cannot be addressed in this research (Galloway et al., 2015). However to simplify: students who find the laboratory personally rewarding or worthwhile may become more motivated (Abrahams, 2009), an essential component of SRL, and therefore potentially increasing summative performance (Zimmerman and Schunk, 2011). However, the effect of motivation on learning practices and summative outcomes vary immensely, and so such suppositions merit much more detailed further work (Schraw et al., 2006; Salta and Koulougliotis, 2015).

Professional Skills

The least-mentioned benefit of the laboratory comes from the skills relevant to future professional work. The types of benefits can be seen in Table 3.6. This category captures students who reported they used the laboratory as a chance to do what a professional chemist would do:

> “Yes, I feel like going forward if you’re going to do anything with chemistry you need to have that practical element because if you’re going into research or working in a lab at any point then it’s kind of essential to be able to do everything that you need to in a lab, and if you haven’t done lab work then it’s quite a bit harder.” (Kahlan)

With other students stating that they would develop skills which would benefit them in vocations other than a practising scientist:

> “I think that, apart from theoretical scientists, everyone else who’s doing chemistry to work in science will likely, at some point in their career, actually be in a lab doing experiments and you need to be prepared for that, you need to have the practical skills and knowledge to do that, and I don’t see how you could do that without actually practising in a lab.” (Nick)

As with experiential benefits, there was very little overlap within these categories. Only one student mentioned the laboratory would help them both see and experience what a professional chemist would do. Likewise with experiencing and developing the skills required of a practical chemist. All other students in this category were divided between categories, mentioning only one.

The quality of graduate chemists is essential to providing value to the skilled workforce (Purcell et al., 2008). It is therefore good that students recognise practical skills as relevant part in their future career. However, it was not valued widely by students. These results suggest that more explicit connection between laboratory activities and post-graduate employment or skills may help students to explicitly connect the two. This relates to wider issues about curriculum design and the role of HE in general, as well as the interplay between laboratory and lectures, workshops, and tutorials. It is therefore a difficult conversation to have in isolation.

Finding: The relationship between learning and the teaching laboratory

Similar to extant related work (Bretz et al., 2013; Sinapuelas and Stacy, 2015; Galloway and Bretz, 2016), the interviews did not reveal a single, simple relationship between the laboratory and learning. Instead, three broad themes emerged:

- The majority of students report engaging with the relevant concepts in a laboratory.

- Approximately half of students recognised the laboratory as an opportunity to contextualise their chemical knowledge, or put their knowledge into practice.
- Half of the students stated laboratory sessions were not beneficial to their conceptual understanding, instead preferring pre- or post-laboratory activities.

Students use the laboratory to engage with chemical concepts and ideas.

Most (15; 83%) students recognised the laboratory as a chance to engage with the relevant chemical concepts and ideas (Table 3.3). This is an important point to discuss, especially as educators may be overestimating the educational impact of laboratory work (Osborne, 2015). The following quotes are examples of student engagement with the relevant chemical processes:

> "you could see the chemistry happening in front of you [what benefits did that bring you?] Easier to understand and a better understanding, and also to understand what's going on, and why you're doing what you're doing... it developed new understanding, and the processes behind it. Also if you did the practical first you could then, once you had the lecture about it, then it'd hit you "oh that's what was happening"" (John-Joseph)

> "There's also certain experiments you do in the lab, the you'll learn theory in the lab that isn't necessarily covered in the lectures, you may have to do an experiment, and then figure out what's happening with the experiment, so there's a certain aspect of problem solving there." (Nikhil)

Before progressing to more critical analysis, I wish to draw attention to the fact that many students reported linking the laboratory and the chemistry that happens in it. To a good extent, the laboratory was viewed as tool for learning This is a hopeful finding, which means Labdog has to improve an existing relationship, not construct something new entirely.

Several students mentioned that the laboratory was a useful way to apply previously-learned theory. For example one student stated that the laboratory "made me think a bit more about the chemical reactions than just reading about them" (Par). Another student provided further detail:

> "I get to see what we predict with theory... if something doesn't go according to plan, in theory, you can explain it using the other bits of theory

> that you know [So how does that help you learn? What benefits does it bring to your understanding, would probably be the better question]... It's mostly reconfirming, going over what we've learned, it's almost a revision exercise in some places." (Kahlan)

The distinction between learning new, or reinforcing old, material was mentioned by eight students. A number of students also mentioned interacting with their peers (3) or the post-graduate demonstrating staff (8) during a laboratory session. One student summarises the role of demonstrators as providing the opportunity to "ask what's going on... especially if you've got spare time and you want to answer the [post-lab] questions" (John-Joseph), and another that:

> "They're there for reinforcement, it's kind of like the scaffolding technique they use in teaching, they partially hold your hand when you want them to hold your hand, but then they'll let go of said hand when you're more confident and you can get on with it again" (Charlotte)

It is good that students feel they can approach the demonstrating staff with both conceptual and procedural questions. It is important, however, to consider the quality and content of the student-demonstrator interaction. Such interactions are complex, and do not always contain consistent or beneficial pedagogical underpinning (Velasco et al., 2016). Recent research has looked at the competencies and practices of these teaching staff, with these concerns in mind (Deacon et al., 2017). It cannot therefore be assumed that such interactions always lead to positive outcomes.

Interestingly, students who mentioned the benefits of interacting with their peers also framed the laboratory as a more collaborative activity:

> "If I want to ask a team-mate I can ask 'what do you think about this question', there's discussion time, it's quite relaxed, the atmosphere, it's not really tedious or quiet" (Elizabeth)

> "Everyone is talking and trying to figure out any bits of theory they don't understand. And one part of learning that works especially for me is teaching, so if I can explain it to someone else I know that I know it well enough" (Nikhil)

> "In a lab you've got people around you and you can be like 'do you know exactly this is? Am I doing this right?' And I think that's helpful to me" (Charlotte)

It is interesting that peer-peer interaction was less mentioned than student-demonstrator. It is possible that many students took peer-to-peer interaction for granted, not mentioning it explicitly. While peer-pper interactions are potentially pedagogically valuable (Freeman et al., 2014), this does not guarantee that every interaction between students leads to an improvement of one or both parties' understanding. This may help explain why some of the students mentioned the interactions with peers as a cause for anxiety on their current position or activity during a practical - which is discussed in more detail later.

Contextualising chemical knowledge does not equate to applying it

I spoke previously about the distinction between active and passive language when discussing the relationship between the laboratory and the relevant theory. Two distinctions emerged during an analysis of the 11 responses which linked laboratory practice to the relevant chemical concepts and ideas. One subset of students (n=7) detailed how the laboratory could help students better contextualise chemical knowledge either to their own knowledge or practice, or to the outside world.

An example of a student who saw the laboratory as a chance to contextualise their knowledge is as follows:

> "the whole point of chemistry is not only just learning theories, you need to see how they work in real-life. [So what do you mean by that?] When you start doing your research project and you decided you wanted to go down the synthesis route, you need to know what to expect from a practical point of view, you need to know what colour should it be, should it be a solid, should it be a liquid, should it be hydroscopic so are you expecting it to stick to the sides of the glassware?" (Charlotte)

Here, Charlotte implies that the laboratory allowed her to "see how [chemical theories] work in real-life". She expresses that her past laboratory experience would inform her future practice. Specifically, her present experience would inform future expectations of what particular types of chemical reactions might look like, so as to have an idea about what to expect in practice. Compare this to Hemant's response:

> "A lab I've just done, it's reinforced a lot of what I've just done in first year, it gives me a chance to put that into practice and say "okay I've learned about this and now I can actually do it" and go back and look at the theory behind it." (Hemant)

Hemant states a more immediate benefit: to assess the quality of his current understanding of previously-learned material. Furthermore, Hemant suggests personal satisfaction from implementing these previously theoretical chemical concepts into his real-life actions. In this instance, the laboratory confirms previously-learned material - just as Charlotte's quote suggests that she will use current experience to (in)validate future experience or results.

Another student, Josh, spoke about how...

> "rather than just being able to know a fact and reel it off in an exam, you'd actually be able to apply it, know what it really means, rather than just knowing the first level"

Both active engagement or passive contextualising are important. A chemist should know what to expect when performing a procedure, and also aware of the chemistry taking place. However both are distinct and serve different purposes, and may lead to different educational outcomes. Being able to identify when either is happening, and develop instructional practices that can incur both when needed/desired would be beneficial. Future research could consider both the factors which lead students to either passive or active consideration of chemical concepts, and the subsequent impacts to understanding or performance.

Finding: Several students report the laboratory as non-beneficial to them

I have largely focused on the benefits of the teaching laboratory to students' understanding. Alongside these benefits, a notable number of students report the teaching laboratory to have little to no influence on their learning. Instead, students seem to use pre- or post-laboratory activities, or other contact hours (e.g. lectures) to engage with and further their understanding of chemical ideas.

Half (9) of all students interviewed stated that the laboratory (not including pre- and post- activities) were not beneficial or useful for their conceptual development. The three quotes below typify these kind of statements:

> “Whereas actually, the actual teaching lab itself, I don’t think it didn’t really help much with the understanding more with just the practical.” (Alice)

> “I didn’t really think too much about the theoretical elements of what was happening... I might be interested in what the mechanics might be. You’d perform the script and you learn how to make observations and recordings of data [but you don’t feel like it helped you develop your conceptual understanding?] Yeah, I’d say sitting in the class and learning, that’s where the theory came in.” (Chapman)

> “There wasn’t really much of a learning style in the labs, it was more “you have this script, do it, ask us questions if you’re stuck or anything” so we didn’t really learn too much” (Miraj)

These responses suggest that students did not always understand why or what they were doing during a laboratory. Such findings are typical of cookbook chemistry (Domin, 1999a), and provide some basis for the fears that students would “blindly continue from one step to the next” (Galloway and Bretz, 2016). Unfortunately, students provided little insight as to why this might be. It is possible that the time pressure or constraints on the laboratory, mentioned by 6 students in total, draws student attention away from the conceptual aspect of the laboratory. This link is made explicitly by Charlie:

> “I bear the pre-labs in mind while working in the lab, it is kind of harder because obviously you are working within a time constraint”

Another student, Hayley, mentioned the stress involved in group work:

> “I tend to want to learn on my own as opposed to going in a lab with everyone else, and having to take that into account as well. I feel I just get overwhelmed in the labs, it’s just a bit too much”

A number of other students suggest that the laboratory may not lead directly to learning. For example:

> “I like to be able to do the theory at my own pace, and then obviously the experiment separate, well not completely separate, but separate from my theory notes?” (Josh)

> "You're under timed pressure, and you know you have to get it done before the deadline...but in lectures and independent study it's your choice as to how you plan it out, and how much you want to do per day" (Nikhil)

These underlying notions or opinions about the laboratory are likely much harder to address, due in part to their varied causes, and interpersonal differences. While it may indeed be beneficial to remove all time-pressures from the students, universities are limited in space and time, and so must enforce such limits. Furthermore, how does one change the relationship within and between students, especially when some students (e.g. Elizabeth) revel in the social aspect, while others (e.g. Hayley) clearly do not.

Perhaps the most important aspect to consider is the assessment of practical work (Bennett and Wilkins, 2004; Abrahams et al., 2013), which can be seen as a proxy for the educator's perceived purpose of the work. Presently, UG assessment at the UoS consists of a pre-lab quiz, the lab book which students have to have signed by 5pm on the day of the practical, and a post-lab report. This assessment only requires demonstration of conceptual understanding after and before the laboratory, with focus on simply completing a practical activity during the lab.

3.2.3 Conclusions

I conducted interviews with 18 UG chemistry students in 2016 to better understand the student-perceived role of the laboratory. I analysed the interview transcripts using grounded theory. Previous research suggests that the laboratory does not help students attain the educational goals set out for them by optimistic educators (Hawkes, 2004; Abraham, 2011).

The interviews revealed a variety of benefits including procedural, conceptual, experiential, and professional. A noteworthy and subtle difference emerged between students who use the laboratory to develop their understanding, compared to those who applied their knowledge. Additionally, one half of students stated that the laboratory was not useful in helping them develop their understanding. This was tied to both the nature of instruction of the laboratory, i.e. as 'cookbook chemistry' (Domin, 1999b), and also to the social environment which could induce anxiety in some students.

These interviews present a further range of supporting evidence that the laboratory environment does not necessarily benefit students' learning, though it does present a range of potential benefits. The design of instruction and assessment can be altered to shift student focus more towards the chemical processes *as* they happen.

3.3 An Overview of Labdog

From extant literature and the above interviews is clear that two hypothetical students who finish a practical activity may not benefit in the same ways, or to the same extent. Although the teaching laboratory may be designed by educators with valuable or consistent pedagogical goals in mind, such benefits may not materialise (Hawkes, 2004).

Labdog seeks to prevent practical work from becoming an exercise in 'cookbook chemistry' (Domin, 1999b), and facilitate pedagogically-founded teaching and learning practices. One problem is that laboratory instruction does not require students to demonstrate conceptual understanding (Von Aufschnaiter and Von Aufschnaiter, 2007). Labdog addresses this concern directly, providing a tool for educators to design and deliver practical activities while collecting evidence of action and understanding. The rationale for such a function is clearly present in extant research, and exemplified by Galloway and Bretz (2016):

> "There is value in step by step procedures to teach new techniques, but students should not be permitted to blindly continue from one step to the next without purposeful checks on their understanding... Asking students to explain as they go is one step to teach the value of thinking about the experiment." (p.153)

Labdog is a web-based, i.e. cloud (González-Martínez et al., 2015), platform for educators to create, store, and disseminate stepwise practical activities to students. At any step in a practical, educators are able to pose questions, request photos from students, and have students submit written observations or data. Additionally, educators are able to create pre- and post-lab activities, e.g. questions and online resources. A student is then able to access any of these activities form any web-connected device. During the laboratory as they progress through the practical activity, step by step, a student can submit photos, data, and answers wherever required. This explains the name "Labdog" - a portmanteau of 'lap dog' and 'laboratory': a companion for students in the lab.

Labdog was influenced by the concepts of retrieval-based learning (Karpicke and Grimaldi, 2012), active learning (Prince, 2004), and formative feedback (Black and

Wiliam, 2009). At a basic technological level, Labdog acts as a learning management system (LMS), similar to Blackboard or Moodle, as it stores resources and activities for students to access at any time prior to, or after, a lab. Additionally it acts as a CRS (Blasco-Arcas et al., 2013). This similarity to a CRS, dedicated to the laboratory environment, is the basis for referring to Labdog as a Laboratory Response System (LaRS).

In a traditional lab environment, students are expected to record their actions, observations, and results in their paper lab books. Electronic lab notebooks (ELN) have become an increasingly popular alternative, as previously discussed (Amick and Cross, 2014; Weibel, 2016). Given the simplicity of handling and creating text data, Labdog acts as a simple ELN. This facilitates Labdog being used as an e-portfolio system (Roberts et al., 2013). A student could demonstrate their competence in the synthesis of aspirin by using Labdog to store photos and observation, and evidenc etheir unserstanding by responding to questions. All of the evidence is collated in a single location, attributable to student, time, and activity.

3.3.1 Introducing Labdog's functionality

In the following section I present three schema to illustrate Labdog's structure and functionality (Figures 3.1, 3.2, and 3.3). Before progressing, I wish to briefly introduce the purpose of these diagrams, and how they should be interpreted.

Schemas represent a series of entities and the relationships which exist between them. An entity can be considered as a single type of data, e.g. a question or a response. For readability and simplicity I have renamed and reduced the number of entities presented compared to the state of the databases.[2]

The shape and colour of the entity both convey information. The shape of the item indicates the mutability and origin of the data: square objects can be considered as an individual record which is consciously and purposefully created or edited by a user, e.g. a practical or a step in a practical. Rounded shapes are trace data – they are created when an user visits or interacts with the software, e.g. students leave trace data when they progress past a step during a practical activity. Lastly, objects which have a wavy shape represent deliberately created data which is not always present or required, e.g. student-submitted photos, as photos are not required on every step.

---

[2] In total, Labdog contains around 48 separate tables in the database, and 53 unique relationships between them.

The colour of each object indicates the ownership. Red objects are owned by an educator - for example a teacher can create a new practical and then edit its name. Blue objects belong to, are created by, and can be (sometimes) edited by the students – for example when a student answers a particular question in a practical. Grey objects are without direct ownership. This includes the teacher and student accounts, but also data which exists to organise other data, e.g. a 'lab report' object which allows an individual student's responses to questions, photos, and observations to a particular practical on a particular date to be gathered on to the same web page.

Lastly, arrows between entities indicate ownership. Arrows with a direction represent a simple 'belongs to' relationship, where the origin of the arrow shows the owned, and the side with the arrow head represents the owner. This is how the software is able to collate or filter data, e.g. showing a teacher all of the responses to a particular question during a particular lab session. I have simplified these relationships, and a more detailed schema may distinguish between 'BELONGS TO', 'HAS ONE', 'HAS MANY', or 'HAS MANY THROUGH' relationships. I have not included these as it does not add meaningfully to the schema, and I do not intend to provide an overly technical account of Labdog.

3.3.1.1 Pre-labs

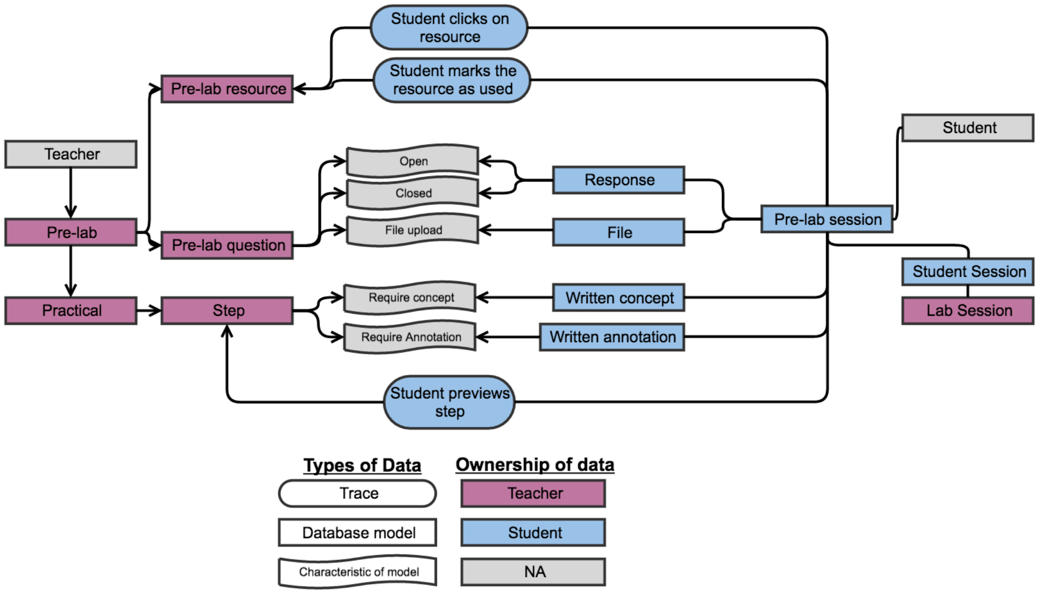


Figure 3.1: A summary of Labdog's pre-lab features, explained through a simplified data schema.

Labdog is a platform for pre-lab activities, allowing educators to create, distribute, and monitor educational activities, and students to view and complete them. Figure 3.1

showcases the prominent types of data related to pre-lab activities, and the relationships which exist between them.

Educators create a ‘pre lab’, which in turn belongs to a ‘practical’. Educators can make a particular practical available to a group of student accounts through a ‘lab session’, which individual students can access through their own ‘student session’. When students log in to Labdog they are able to access the pre-lab activities separately to actually completing a practical, which they do through a ‘pre lab session’.

Teachers are able to create two kinds of material to put into a pre-lab: ‘questions’ and ‘resources’. Questions can be open- or closed-answer, and educators can also allow students to attach a file (e.g. an image) to their response. ‘Resources’ are links to externally-hosted resources, e.g. videos on YouTube.

Teachers are also able to edit an attribute on particular steps (of the practical that a pre-lab is associated with) to require students to submit ‘annotations’ to steps and/or note the related scientific ‘concepts’ to a particular step.

When a student logs on to Labdog they can access their ‘pre lab session’, where they may see each of these items in a single interface. They are able to click through to externally hosted URLs, and also mark these resources as complete or used. Labdog records when a student does both of these activities through trace data.

Students are also able to answer questions and submit files in response to the questions posed by teachers, and preview all of the steps in a practical. Through this previewing interface, students can submit concepts and annotations to steps which are required.

3.3.1.2 During labs

The details of during-lab functionality are showcased in Figure 3.2. As mentioned in the previous section, a teacher allows a ‘group’ of ‘students’ the chance to complete a ‘practical’ (and associated ‘pre labs’), by creating a ‘lab session’. Students can access the information associated with a practical through their own ‘student session’.

A ‘practical’ is simply a series of ordered ‘steps’ for a student to complete. Teachers can create a ‘question’ on an individual step, which are presented to students as they complete that particular step in Labdog. Students then submit a ‘response’ to these questions. Furthermore, educators can request that students provide an ‘observation’ for a particular step, and request for a photo in a similar manner.

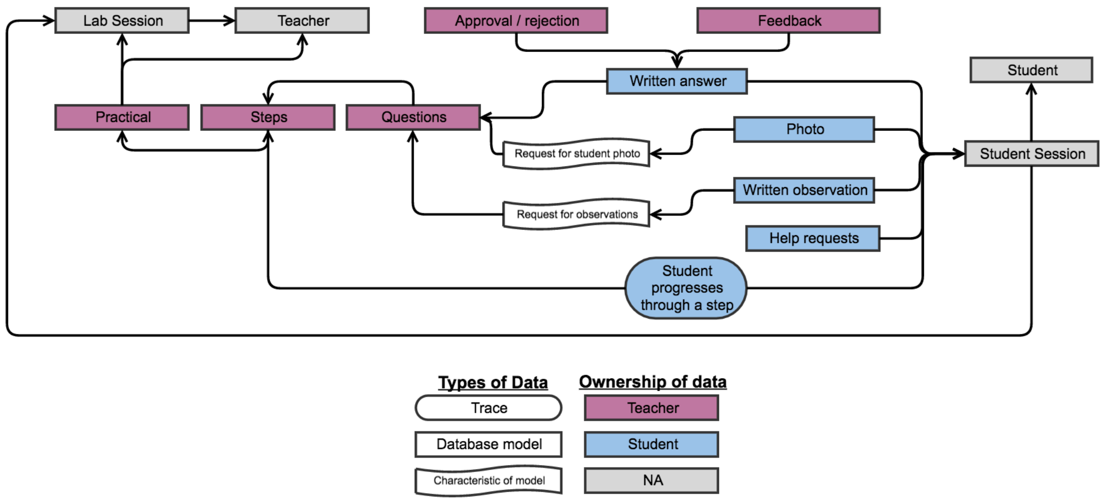


Figure 3.2: A summary of Labdog's features which students and educators can capitalise on during a laboratory session, explained through a simplified data schema.

All responses that students make for a particular lab session can be viewed by the educator, who can 'approve' or 'reject' (formally 'answer needs revising') an individual response and provide feedack (formally 'step question feedback') which can either be textual or verbal. In the latter case, students summarise the verbal feedback the educator gives.

Students are also able to submit 'help requests': small messages sent to the educator in real-time, if they need help during a session. Additionally, whenever a student progresses from one step to the next, Labdog creates and stores a piece of trace data.

3.3.1.3 Post-labs

Figure 3.3 showcases the basic functionality of Labdog in the post-lab context. This information builds on data submitted by students during the lab session (above; Figure 3.2), particularly an understanding of the relationship between practicals, lab sessions, and student sessions.

All the responses to questions given during a lab session are collected into a 'group feedback' page, where educators can provide 'post lab feedback' to both the entire lab session, and also individual questions. Labdog automatically generates types of feedback - one each for good, neutral, and bad points about a particular lab session. Educators

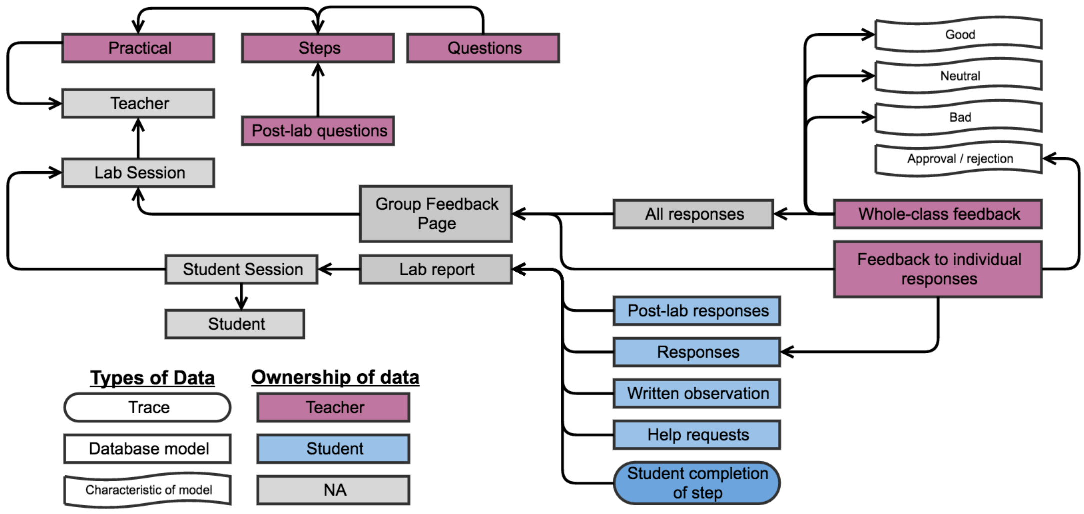


Figure 3.3: A summary of Labdog's post-lab features, explained through a simplified data schema.

and students both access the same post-lab feedback page, but only educators can edit the data.

Educators are able to create post-lab questions, which are very similar to the previously discussed 'step questions', and are associated with a particular step. Labdog collates all of a student's responses to questions, written observations, help requests, and trace data for completing a step into a single 'lab report'. When students access this lab report they can respond to the post-lab questions. Educators can also access a student's lab report to look at how they have performed, and if they have completed the post-lab questions.

3.3.2 Technological Overview

Labdog was built using Ruby on Rails[3] (Rails) version 4, an open-source framework for web applications which uses Ruby, the open-source programming language[4]. Rails follows a model-view-controller (MVC) architecture. Between January 2015 - January 2017 Labdog was hosted on Red Hat's Openshift service[5] running RedHat's Linux Distribution. From January 2017 onwards Labdog was migrated to Digital Ocean server

---

[3] `http://rubyonrails.org/`
[4] `https://www.ruby-lang.org`
[5] `https://www.openshift.com/`

running Linux Ubuntu 16.04[6], served from an Ngnix[7] and Unicorn[8] servers. Both instances stored data using the open-source MySQL database[9], an ACID relational database.

The front end was developed with Twitter's Bootstrap[10] framework (version 3), though custom styles were used. The front end also relied heavily on the use of Javascript and the JQuery[11] framework. Chart rendering was done through a mixture of D3.js[12] and ChartJS[13].

The development of Labdog represents approximately 1,000–1,500 hours of effort. I have made a conscious effort to focus this thesis on the educational rationale and implications of Labdog, and therefore I am simply unable to detail the length, complexity, and nature of the development process.

Two paradigms of software development informed the process above all others: agile (Martin, 2003) and lean development (Ries, 2011). These are both modern practices embraced by entrepreneurs. At the heart of agile development is the idea of developing a minimal viable product (MVP) - a small proof-of-concept piece of software which allows developers to test assumptions before further committing resources to development. Lean software development promotes this test-driven process of making decisions and dedicating resources. This is related to the lean principle of pivoting: using the findings from these experiences and assertions to change the direction and intent of development so solve the most pressing or difficult problems, even if they did not arise in the initial software conceptualisation. Agile development is about rapidly and continuously evaluating the effectiveness of these MVPs, and deciding if features should be added, removed, enhanced, etc. During this testing, if a more fruitful or beneficial approach or functionality reveals itself, then it should be capitalised on.

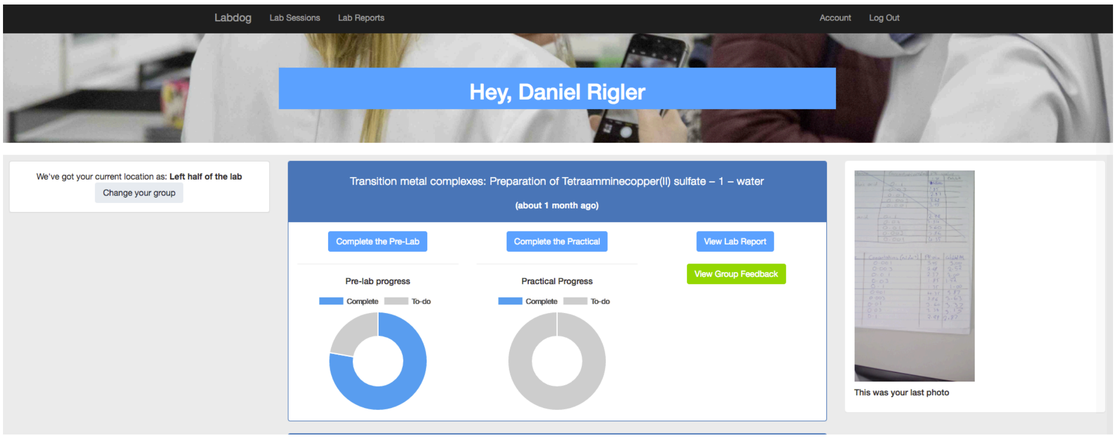
3.4 Labdog's Features

In this section I wish to describe Labdog from a more applied perspective. As with the technical discussion, I divide my discussion by pre-, during-, and post-laboratory

---

[6] `https://www.digitalocean.com`
[7] `http://ngnix.org`
[8] `https://bogomips.org/unicorn/`
[9] `https://www.mysql.com/`
[10] `http://getbootstrap.com/`
[11] `https://jquery.com/`
[12] `https://d3js.org`
[13] `http://chartjs.org`

activities. The student homepage, which is presented immediately after they log in, is shown in Figure 3.4. For each lab session made by the teacher, the student is presented with an individual box, which contains a simple open learner model (Bull and Kay, 2010) on their progress towards completing the pre- and during-lab activities. Open learner models (OLM) are student-facing interfaces into information on their engagement, and can help students monitor their progress towards goals, and enhance reflection about reaching these goals - both are SRL behaviours (Law et al., 2017).



Hey, Daniel Rigler

We've got your current location as: **Left half of the lab**

Transition metal complexes: Preparation of Tetraamminecopper(II) sulfate – 1 – water

(about 1 month ago)

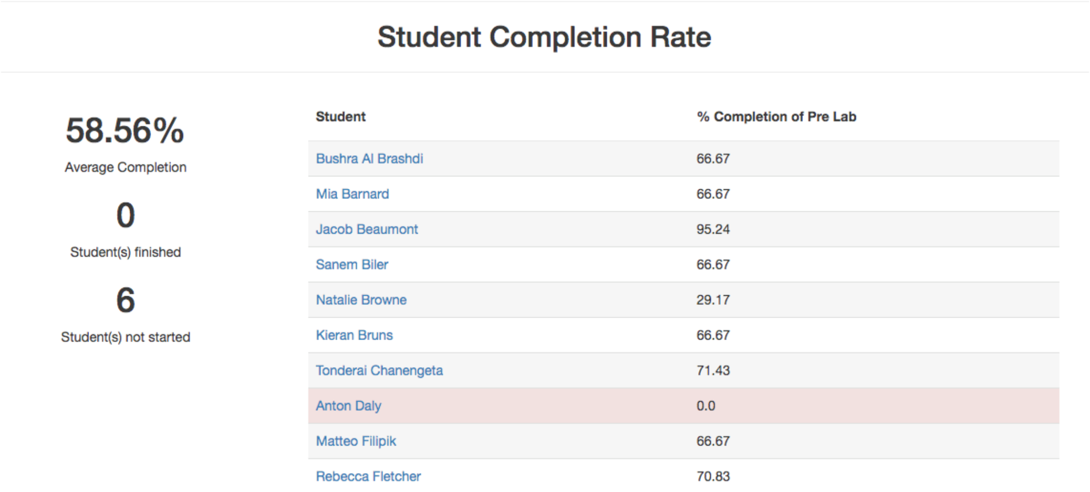
This was your last photo

Figure 3.4: The student home page of Labdog

3.4.1 Pre-Labs

Pre-lab activities are a set of tasks for students to complete before entering a laboratory. Pre-labs have been common-place in the university's undergraduate teaching lab since 2008, and are useful for reducing the cognitive demand on students during the laboratory, and can therefore increase the amount of material which can be covered in a lab session (Carnduff and Reid, 2003; Jolley et al., 2016). Pre-labs also introduce the scientific concepts relevant to a laboratory, and facilitate group- or individual feedback based on students' responses. I have discussed the need to highlight underlying concepts (Galloway et al., 2015), and the value of timely formative feedback (Beatty and Gerace, 2009; Walker et al., 2016). This focus on concepts and knowledge may come at the expense of practical competency and skills, so there is a need to balance the two (Grant and Jenkins, 2011).

Teachers can create four types of activities in a pre-lab: questions, externally-hosted resources, previewing the steps, and providing annotations. These are combined into an equal-weighting index for completion, i.e. previewing all of the steps in the practical is weighted the same as answering all of the questions. These percentages are generated for each student, and can be seen by the teacher in a simple table, as in Figure 3.5. Each student's name is a hyperlink to an individual student's pre-lab report, which provides further detail on each of the four pre-lab activities.



Student Completion Rate

58.56%
Average Completion

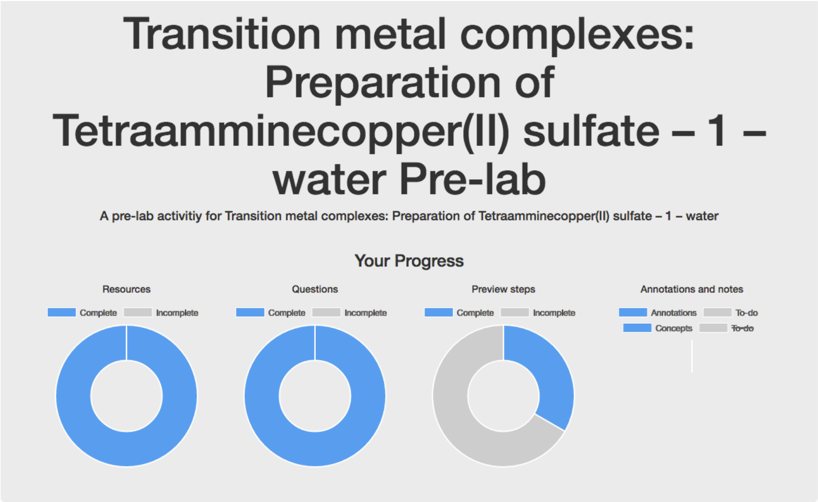
0
Student(s) finished

6
Student(s) not started

Student	% Completion of Pre Lab
Bushra Al Brashdi	66.67
Mia Barnard	66.67
Jacob Beaumont	95.24
Sanem Biler	66.67
Natalie Browne	29.17
Kieran Bruns	66.67
Tonderai Chanengeta	71.43
Anton Daly	0.0
Matteo Filipik	66.67
Rebecca Fletcher	70.83

Figure 3.5: The teacher-facing interface of each student's progress towards completing a pre-lab activity.

Similar to the home page, the student view of the pre-lab presents an OLM of their completion, divided by the four pre-lab activities, as seen in Figure 3.6. The page was designed to immediately present the current state of the students' progress, so they could quickly identify what they need to do and progress through the pre-lab.



Transition metal complexes: Preparation of Tetraamminecopper(II) sulfate – 1 – water Pre-lab

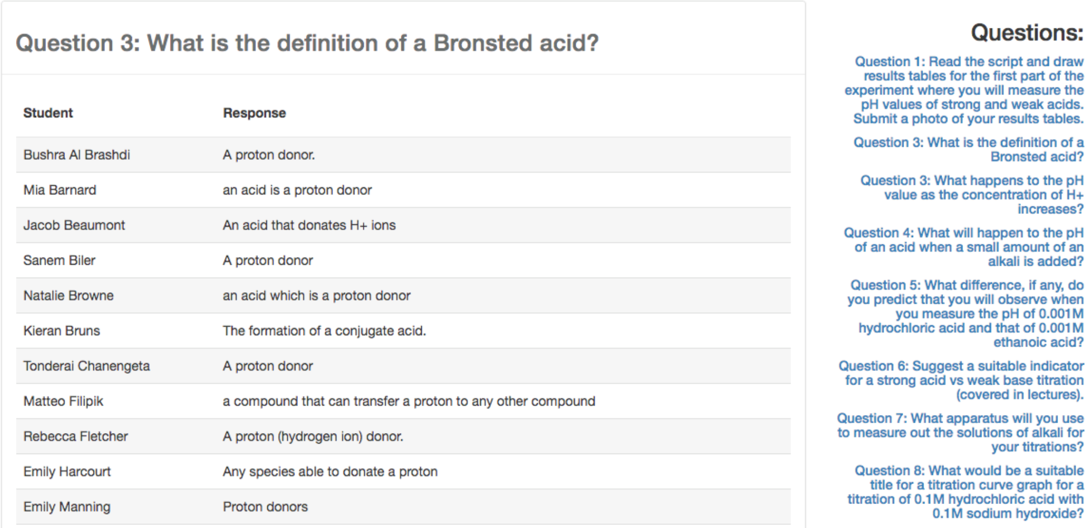
A pre-lab activitiy for Transition metal complexes: Preparation of Tetraamminecopper(II) sulfate – 1 – water

Your Progress

Figure 3.6: The student-facing open-learner model of their progression towards completing a particular pre-lab.

3.4.1.1 Pre-lab questions

Teachers are able to create three kinds of questions: open, closed, and file. Open- and closed/multiple-choice questions are self-explanatory. File questions allow students to submit any file, e.g. a picture of a results table. Teachers can access the responses to all pre-lab questions grouped by question, as seen in Figure 3.7, or by individual student by going to that student's pre-lab report. This was designed to enable quick and targeted formative assessment (Beatty et al., 2006).



Question 3: What is the definition of a Bronsted acid?

Student	Response
Bushra Al Brashdi	A proton donor.
Mia Barnard	an acid is a proton donor
Jacob Beaumont	An acid that donates H+ ions
Sanem Biler	A proton donor
Natalie Browne	an acid which is a proton donor
Kieran Bruns	The formation of a conjugate acid.
Tonderai Chanengeta	A proton donor
Matteo Filipik	a compound that can transfer a proton to any other compound
Rebecca Fletcher	A proton (hydrogen ion) donor.
Emily Harcourt	Any species able to donate a proton
Emily Manning	Proton donors

Questions:

Question 1: Read the script and draw results tables for the first part of the experiment where you will measure the pH values of strong and weak acids. Submit a photo of your results tables.

Question 3: What is the definition of a Bronsted acid?

Question 3: What happens to the pH value as the concentration of H+ increases?

Question 4: What will happen to the pH of an acid when a small amount of an alkali is added?

Question 5: What difference, if any, do you predict that you will observe when you measure the pH of 0.001M hydrochloric acid and that of 0.001M ethanoic acid?

Question 6: Suggest a suitable indicator for a strong acid vs weak base titration (covered in lectures).

Question 7: What apparatus will you use to measure out the solutions of alkali for your titrations?

Question 8: What would be a suitable title for a titration curve graph for a titration of 0.1M hydrochloric acid with 0.1M sodium hydroxide?

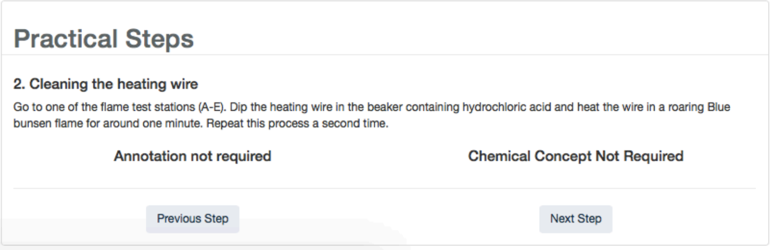
Figure 3.7: Teacher-facing interface of all students' responses to individual pre-lab questions

3.4.1.2 Online resources

Labdog allows teachers to provide links to externally-hosted resources via a hyperlink. This was most commonly used for online videos but can include documents hosted on LMSs, news articles, and any other web-based resources

3.4.1.3 Previewing practical steps

The student-facing pre-lab page presents the relevant practical to the student, one step at a time, as seen in Figure 3.8. This interface tracks when and if students have previewed each of the steps of a practical, providing a measure of pre-lab engagement, and supported by the idea that pre-labs should reduce cognitive load in the laboratory.



Practical Steps

2. Cleaning the heating wire

Go to one of the flame test stations (A-E). Dip the heating wire in the beaker containing hydrochloric acid and heat the wire in a roaring Blue bunsen flame for around one minute. Repeat this process a second time.

Annotation not required

Chemical Concept Not Required

Figure 3.8: An opportunity for students to preview the steps involved in a related practical.

3.4.1.4 Annotating Steps

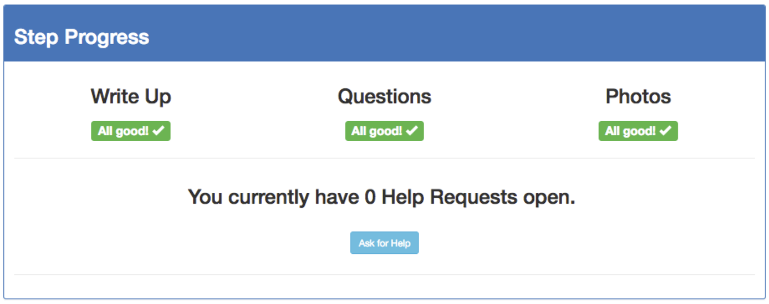
Teachers are able to identify individual steps within a practical which they would like students to annotate, or list the chemical concepts related to that particular step. Students use the same interface as above (Figure 3.8), and text-boxes would appear where input is required from the student.

3.4.2 During the lab

Labdog's most well developed and used features take place in the laboratory. Existing literature suggests that students struggle to relate observations or activities in the laboratory to the theory they learned in the classroom (Abrahams and Millar, 2008; Abrahams et al., 2013). This finding also emerged in the undergraduate interviews (Section 3.2.2).

Labdog was informed heavily by technology-enhanced formative assessment (Beatty and Gerace, 2009), where questions direct students' attention to some related chemical concept or idea, and their response would evidence a (mis)understanding. Questions could be used for simplistic recall or more cognitively advanced application or evaluation (Krathwohl, 2002) - a freedom which is discussed in greater detail in the evaluation of Labdog. Labdog also contains a number of digital lab notebook-like functions, allowing students to submit rich-text observations for a specific step (Van Dyke and Smith-Carpenter, 2017).

As was discussed in Section 3.3.1.2, and shown in Figure 3.2, teachers can control the requirements for each practical on a step-by-step basis. E.g. one step may require a photo, the next three, and the next nothing. To simplify the student experience, the student-facing interface quickly shows students what they have to do, and what they have done, on each particular step - as seen in Figure 3.9.



Step Progress

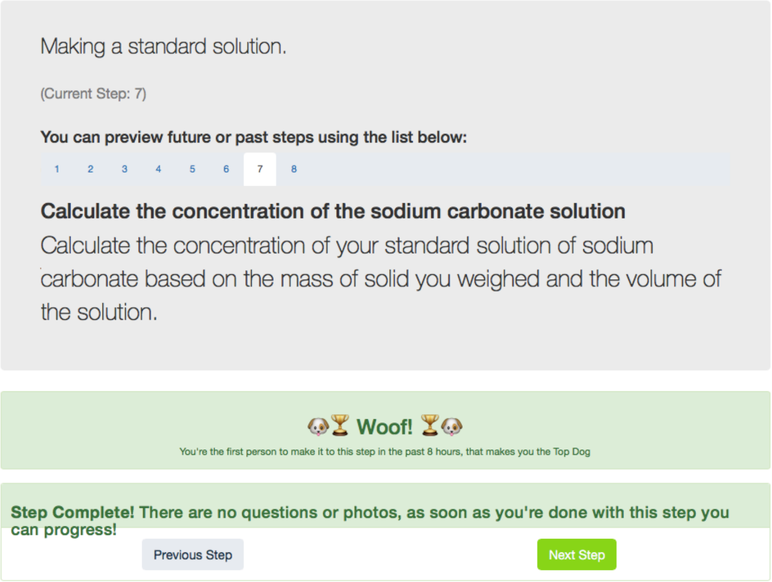
Write Up
All good! ✓

Questions
All good! ✓

Photos
All good! ✓

You currently have 0 Help Requests open.

Figure 3.9: A student's view of their current progress in a particular step in a practical - informing them of any activities they are required to do before progressing.



Making a standard solution.

(Current Step: 7)

You can preview future or past steps using the list below:

Calculate the concentration of the sodium carbonate solution

Calculate the concentration of your standard solution of sodium carbonate based on the mass of solid you weighed and the volume of the solution.

🐶🏆 Woof! 🏆🐶

You're the first person to make it to this step in the past 8 hours, that makes you the Top Dog

Step Complete! There are no questions or photos, as soon as you're done with this step you can progress!

Figure 3.10: The students' view of a single step in a practical.

3.4.2.1 Digital lab script

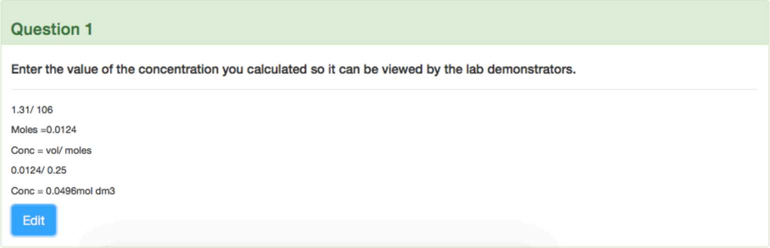
Labdog presents students with the practical activity one step at a time, limiting the information they can see at one time. The student-facing interface, Figure 3.10, clearly shows the name of the practical, the student's current step, the related instructions, and any warning. Each of these are created, and can be edited, by a teacher account.

It is conceivable that educators in schools will share resources, or obtain resources from external bodies (e.g. CLEAPSS[14]), online resources (e.g. the Times Education Supplement online), or qualifying bodies. For this reason, Labdog includes an integrated sharing service, which allows users to make copies of other's practicals.

---

[14] `http://www.cleapss.org.uk/`

Figure 3.11: The rich-text box for students to submit write ups associated with specific steps in a practical - acting as a simple electronic lab notebook.


Question 1

Enter the value of the concentration you calculated so it can be viewed by the lab demonstrators.

1.31/ 106
Moles =0.0124
Conc = vol/ moles
0.0124/ 0.25
Conc = 0.0496mol dm3

Figure 3.12: A student's answer to a question, and the opportunity to edit their response.

3.4.2.2 Record Observations

Labdog lets students record their observations or actions in relation to a specific step. As seen in Figure 3.11, this includes a rich text box, allowing students to include sub- and super-script, as well as tables.

3.4.2.3 Answer questions

Teachers associate questions in a practical to a particular step. Whenever a student is on that step Labdog presents all questions simultaneously, and students are free to answer in any order. Figure 3.12 shows a previously-given response to a question, and the ability of a student to edit this previously-given response.

3.4.2.4 Submit photos

Teachers can flag certain steps in a practical where students need to submit a photo. When students arrive at these steps, they are able to submit a photo file within Labdog.

3.4.2.5 Ask for help

Students are able to submit a simple text request for help, which is sent through to the teacher-view for a particular lab session.

3.4.3 Post-labs

SFY have an existing post-laboratory coursework: the skills portfolio, a reflection-based piece of written coursework submitted via the VLE after every laboratory session. The portfolio combines written reflection and pictures of students demonstrating competency of completing certain techniques, such as correctly weighing materials. Due to this existing assessment framework and workflow, there was little pressing need to develop many post-laboratory features. Post-labs are largely about re-presenting students with their own, and others', data. This was introduced to provide an opportunity for formative- and self-assessment.

3.4.3.1 Lab reports

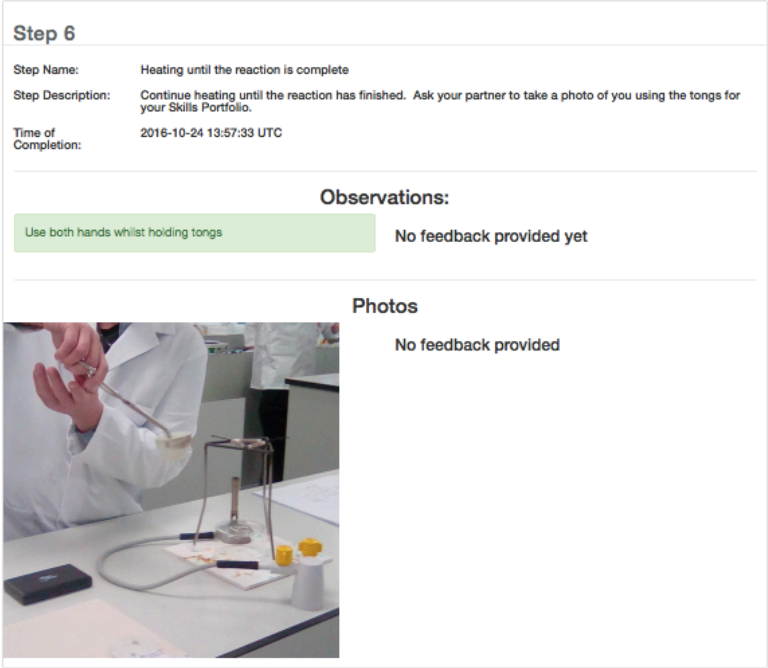
Labdog generates a lab report for each student's completion of lab session. The report contains all of the student's progression, recorded observations, answers to questions, and photos submitted. A single step from a lab report can be seen in Figure 3.13.

3.4.3.2 Review Individual Responses

After a laboratory, students have the opportunity to review all of the answers submitted to a question. This includes their own responses, as well as those submitted by the rest of the group. Figure 3.14 shows the student-facing interface.

3.4.3.3 Group Feedback

Labdog additionally collected all students' answers to individual questions into a single place, and allowed educators to provide overall group feedback. This is in contrast to the individual feedback provided on lab-reports. The student-facing group feedback page can be seen in Figure 3.15.



Step 6

Step Name: Heating until the reaction is complete

Step Description: Continue heating until the reaction has finished. Ask your partner to take a photo of you using the tongs for your Skills Portfolio.

Time of Completion: 2016-10-24 13:57:33 UTC

Observations:

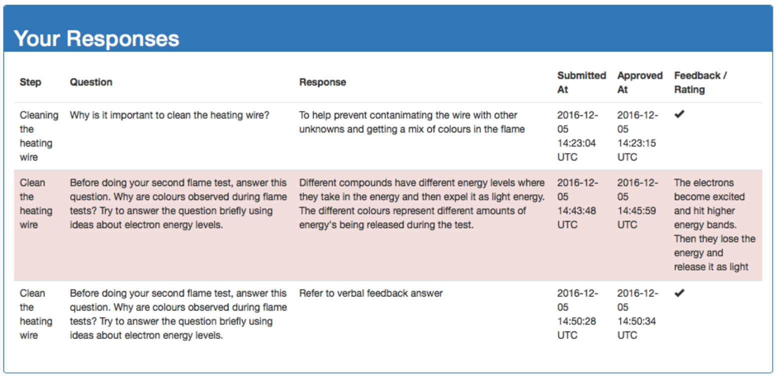
Use both hands whilst holding tongs

No feedback provided yet

Photos

No feedback provided

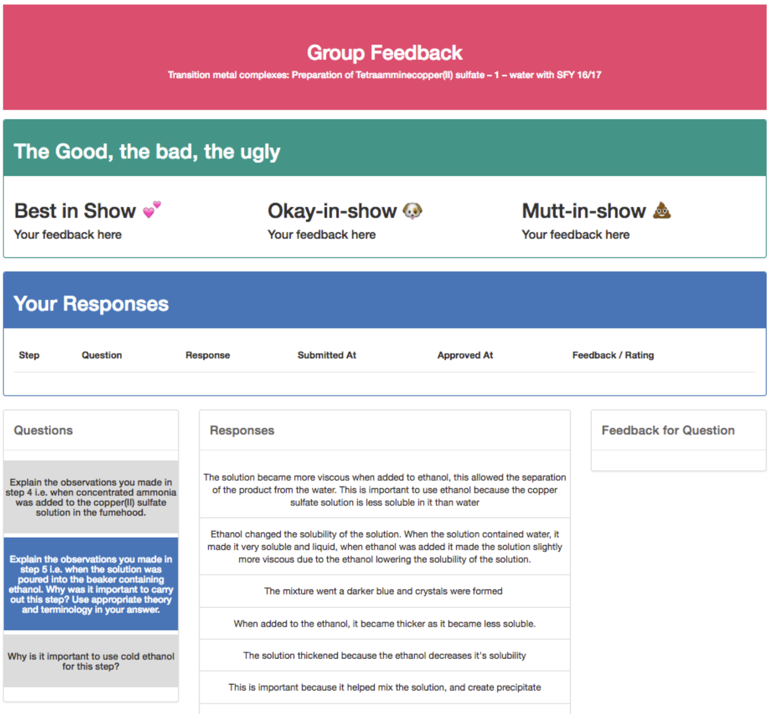
Figure 3.13: A student's view of a single step in their lab report



Your Responses

Step	Question	Response	Submitted At	Approved At	Feedback / Rating
Cleaning the heating wire	Why is it important to clean the heating wire?	To help prevent contanimating the wire with other unknowns and getting a mix of colours in the flame	2016-12-05 14:23:04 UTC	2016-12-05 14:23:15 UTC	✓
Clean the heating wire	Before doing your second flame test, answer this question. Why are colours observed during flame tests? Try to answer the question briefly using ideas about electron energy levels.	Different compounds have different energy levels where they take in the energy and then expel it as light energy. The different colours represent different amounts of energy's being released during the test.	2016-12-05 14:43:48 UTC	2016-12-05 14:45:59 UTC	The electrons become excited and hit higher energy bands. Then they lose the energy and release it as light
Clean the heating wire	Before doing your second flame test, answer this question. Why are colours observed during flame tests? Try to answer the question briefly using ideas about electron energy levels.	Refer to verbal feedback answer	2016-12-05 14:50:28 UTC	2016-12-05 14:50:34 UTC	✓

Figure 3.14: A student's view of all the responses they have submitted during a practical. Those responses highlighted in red have been rejected.



Group Feedback

Transition metal complexes: Preparation of Tetraamminecopper(II) sulfate – 1 – water with SFY 16/17

The Good, the bad, the ugly

Best in Show 💕
Your feedback here

Okay-in-show 🐶
Your feedback here

Mutt-in-show 💩
Your feedback here

Your Responses

Step	Question	Response	Submitted At	Approved At	Feedback / Rating

Questions

Explain the observations you made in step 4 i.e. when concentrated ammonia was added to the copper(II) sulfate solution in the fumehood.

Explain the observations you made in step 5 i.e. when the solution was poured into the beaker containing ethanol. Why was it important to carry out this step? Use appropriate theory and terminology in your answer.

Why is it important to use cold ethanol for this step?

Responses

The solution became more viscous when added to ethanol, this allowed the separation of the product from the water. This is important to use ethanol because the copper sulfate solution is less soluble in it than water

Ethanol changed the solubility of the solution. When the solution contained water, it made it very soluble and liquid, when ethanol was added it made the solution slightly more viscous due to the ethanol lowering the solubility of the solution.

The mixture went a darker blue and crystals were formed

When added to the ethanol, it became thicker as it became less soluble.

The solution thickened because the ethanol decreases it's solubility

This is important because it helped mix the solution, and create precipitate

Feedback for Question

Figure 3.15: The student-facing group feedback page in Labdog

Chapter 4

Piloting Labdog (2015-17)

Having introduced Labdog, in this chapter I detail three pilot studies which were conducted between 2015-2017. These earlier pilots were incredibly formative in the ongoing evaluation and design process necessitated by DBR.

4.1 Science Foundation Year (2015/16)

Labdog was first used by students in October 2015, in four fortnightly practical sessions with the science foundation year (SFY) at the UoS. At this time Labdog existed as a very early-stage piece of software (1.5) - it simply presented students with the practical activity, and the questions. Student submission of photos were not yet fully deployed, and there was no teacher-facing interface - I had not even developed lab reports for students and educators to see all information in retrospect. This short-term pilot was intended to provide rapid feedback to focus future developments and improvements of Labdog.

There were 51 students in the 2015/16 SFY cohort, and attendance to lab sessions was mandatory. Labs were held in a general teaching laboratory at the UoS, where students worked in self-selected pairs, and were asked to complete the activity in Labdog in these same pairs using only one device between them.

4.1.1 Student Usage of Labdog

All of the practical activities undertaken were part of the existing SFY syllabus. Each activity was therefore adapted from a pre-existing lab script, which included both

No.	Practical	Steps	Started	Pen.	Anal.	Pen. (%)	Anal. (%)
1	Preparation of copper (II) sulphate	18 (+5)	33	14	10	42.4%	30.3%
2	Finding the formula of hydrated copper (II) sulphate	11 (+7)	38	13	1	34.2%	2.6%
3	Measuring enthalpy changes	6 (+1)	22	16	2	72.7%	9.1%
4a	Flame Tests	6	10	5	-	50.0% -	
4b	Silver Nitrate	7 (+3)	15	4	1	26.7%	6.7%

Table 4.1: SFY student completion of Labdog practicals in 2015/16, each practical is listed with the number of steps required in lab and in analysis - presented in parentheses. The table shows the number of students (N=51) who started each practical activity, and then the the number and percentage these students who reached the penultimate step of the practical and analysis sections of the practical.

practical and analytical steps. Practical steps were designed to be completed during a lab session, analytical steps are those completed afterwards, e.g. plotting results. Table 4.1 lists the name, timing, and number of steps for each practical.

Table 4.1 also shows the number of students who started each activity, reached the penultimate step of the non-analytical steps in Labdog, and completed the analytical steps in Labdog. In total there were 51 valid student accounts associated with the SFY in this year, though recall that students were only required to complete the practical in pairs.

There is a consistently low rate of completion for the analytical steps, likely because students were required to complete such activities in their lab-books, with little to no obvious benefit from using Labdog. Here, Labdog is separated from both teacher and student workflow, which demonstrated the importance of the relationship between the software and the actions undertaken by both teachers and students.

In the first three sessions, approximately 50% of students logged into Labdog and completed at least a single step - the expected level given students were working in pairs. However student uptake drops heavily in sessions 3 and 4, where only 20-30% of the student population logged in. There is, however, a spike to 70% in week three (measuring enthalpy), though it is unclear why.

The overall decline is symptomatic of the problems which emerged over the four sessions: frequent technical and logistical errors prevented the use of Labdog. This led to the perception of Labdog as unreliable by both students and teachers. Furthermore, there were issues with students bringing and charging their own devices, especailly as labs would start at 2pm, when many students' phones would already be low on battery.

Additionally there was a lack of obvious or immediate benefits from using Labdog. The SFY lab is staffed by DR and two postgraduate teaching assistants (PGTAs). During this trial, Labdog had no real-time features, and so could not provide useful information to educators in a timely way. This clearly demonstrated the distinction between collecting data within a database, and presenting it in a meaningful and logistically easy way to users.

4.1.2 Student Opinion of Labdog

Students were surveyed for their perception of Labdog after activities 1 and 2. Using the UoS' online survey tool[1] the surveys were e-mailed to students shortly after the lab sessions where 13 and 24 students responded to each survey respectively. These surveys came before significant reductions in student uptake (session 3), so the responses cannot necessarily help explain the reduction in student use at that point.

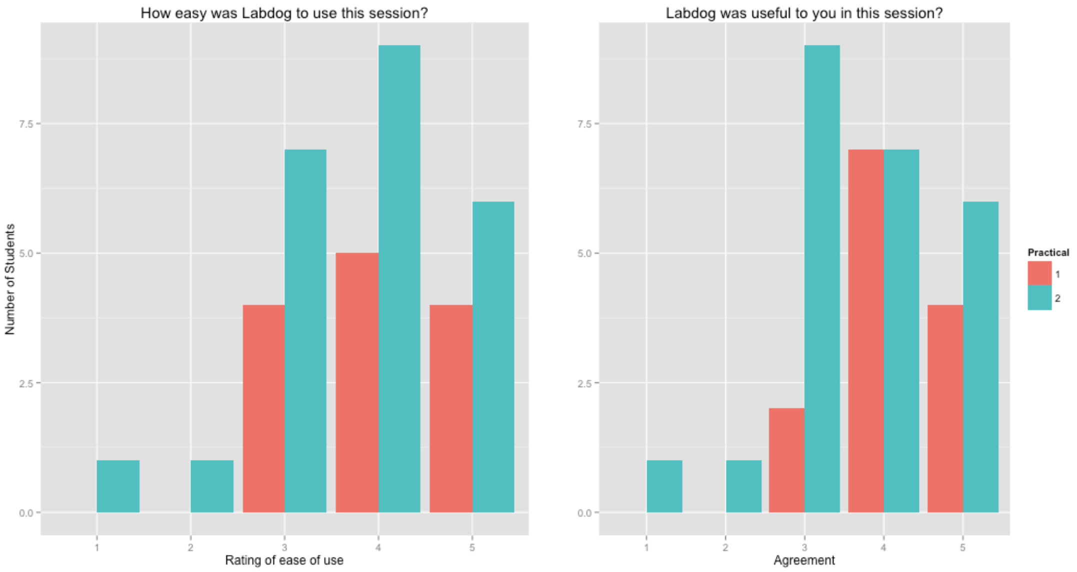


Figure 4.1: SFY students' (2015/16) responses to two evaluation questions asked after the first and second sessions of using Labdog. The left-hand graph shows students' rating of how easy Labdog was to use; whereas the second graph visualises their subjective ratings of usefulness.

---

[1] https://isurvey.soton.ac.uk

Students were asked to rank, on a five-point likert scale, both the ease-of-use and overall usefulness of Labdog. Figure 4.1 visualises students' responses to these questions. Labdog can be seen as both easy to use, and useful, though not unanimously so.

There is a noticeable difference in student-perceived usefulness between sessions 1 and 2. Specifically there is a decreasing student-reported usefulness, where ease-of-use remains relatively consistent between the two week. To provide qualitative insight into this trend, I asked students to provide text-based comments or feedback in an open question. In the first week, a total of 13 responses were left, of which five were entirely positive, and the rest provided at least some critical element. Several themes of critique emerged:

- Improvements to the instructional design, e.g. wording or ordering of the activity in Labdog (6 students);
- Improved or clearer layout (4 students);
- The difficulty of using Labdog alongside paper-based instruction or lab books (4 students).

The re-design or re-structure of content is relatively simple, changing the layout or order of information to best help the student. The last issue, using Labdog alongside paper-based mediums, is more deeply related to the nature of Labdog within the laboratory environment. This theme re-emerged during the second survey, in which I asked students to explain if, and how, Labdog hindered them. Eight of 13 responses commented on having to use using multiple media during a lab session. Addressing this problem involves changing the activities given to students in a laboratory session, something which was not advisable with an early-stage product, with little evidence that it improved student learning.

4.1.3 Conclusions

After session four, given the severely reduced student engagement, myself and DR made the decision to end the pilot and focus on improving Labdog. A number of themes for improvement had emerged, primarily the need to improve the underlying stability and maintain students' good-will towards the software.

Labdog's use during the 2015/16 year was admittedly problematic. Recurring technical and instructional problems hindered the both the teacher and student experience. I do not believe the evaluations capture this feeling, largely because my efforts during the period were spent attempting to fix multiple technical problems with Labdog. These were compounded by the logistical problems of intermittent wi-fi, and students bringing insufficient or non-charged devices.

In retrospect I believe I expected a great deal from students. They were required to bring their own digital device to use software which crashed or broke frequently, and gave them no reward or incentive for their compliance, and was not summatively required. Furthermore I required educators to support me in this, when the software presented them with little-to-no benefit in their activities.

4.2 Twilight Outreach Events (2016)

Labdog was next used in January 2016, after several months of technical improvements (Figure 1.5). During this time I developed lab reports for individual students, as I started considering how data could be presented to educators and students in a way which is both familiar, and useful - to evidencing both understanding and completion of the practical. I worked on several breaking technical issues, specifically better error handing - previously an error for a single student would ‘block’ the server, causing the server to be inaccessible to all students. Understandably, this became an extreme limitation in the student and educator experience, so I treated it as a priority.

Labdog was used at seven days of outreach events, organised within the UoS. During these sessions local AS-level students complete a natural product extraction from the household spice nutmeg in the undergraduate teaching lab.

The events presented a large audience of new users who could provide useful feedback on Labdog. The one-off, and unassessed nature of the events meant it was a relatively low-stakes environment in which to do so.

I transferred the pre-existing lab-script into Labdog. Student pairs, which were self-selected, were provided with a laptop computer which I had already logged into Labdog - reducing the technical workload on students. Five minutes of the pre-planned thirty-minute introductory presentation was dedicated to introducing Labdog to students, and I was available throughout each session to troubleshoot. Students were also given a paper copy of the lab script, on which the location of questions in Labdog was denoted with a paw-print symbol, which served as a reminder for students to use Labdog.

Date	N. Started	N. Penultimate	N. Last	% Pen.	% Last
13/01	18	9	9	50.0	50.0
14/01	16	6	6	37.5	37.5
15/01	17	13	11	76.5	64.7
18/01	16	10	8	62.5	50.0
19/01	15	9	8	60.0	53.3
20/01	19	12	11	63.2	57.9
21/01	20	13	9	65.0	45.0
Overall	**121**	**72**	**62**	**59.5**	**51.2**

Table 4.2: The number and percentage of students who started the natural product extraction on each date, and who reached both the penultimate and last steps during the 2016 twilight events.

After extracting the natural product, students must characterise their product in three separate ways. Although these activities were in Labdog, they have been excluded from analysis due to low participation - I suspect due to high logistical requirements of moving a Laptop between three separate work stations.

4.2.1 Student usage

In total, 121 student accounts logged into Labdog over the seven sessions, each representing a pair of students. No attendance list was taken, and it is therefore not possible to comment on student engagement by demographic. Anecdotally, I consistently ensured that many or all pairs of students at least logged on to Labdog during a particular session. This was occasionally hindered by technical problems with Wi-Fi.

Throughout the events, 72 accounts (59.5%) reached the penultimate step of the practical, and 62 (51.2%) finished. Student completion, divided by day is given in Table 4.2. The data represents a higher and more consistent completion rate compared to the 2015/16 SFY data (Table 4.1).

I examined the data of individual steps in the practical to see where/when student retention fell most. Of students who did not reach the penultimate step (n=49), there were two steps where abandonment spiked: step 7 (n=9, 18.4%) where students use rotary evaporation to remove a solvent, and step 12 (n=16, 32.7%), the final purification step, where students perform a vacuum filtration using cold acetone.

Students took longer to complete these steps. The time spent on each step among the whole cohort is shown in Figure 4.2. The figure clearly shows that students typically spent longer on steps 5, 7, 9, and 12 than they did on other steps. However, only steps 7

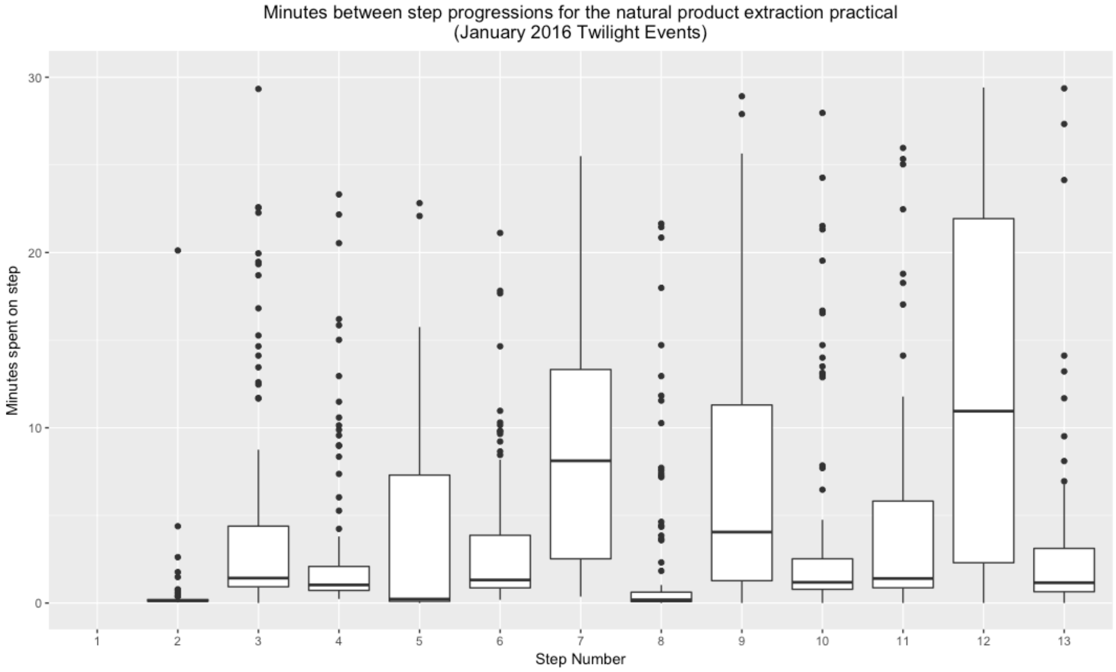


Figure 4.2: Boxplot showing the time (minutes) between student progression from one step to the next in the 2016 January twilight events.

and 12 saw notably student attrition, with 0 and 3 students dropping out on steps 5 and 9 respectively. It is important to also consider the variation in time spent on individual steps, where outliers provide valuable information. While the upper quartiles in time spent completing steps 5 and 9 are relatively high, the mean time spent during these steps is relatively comparable. Steps where more students were more likely to spend longer before progression saw the highest drop-out rates.

Without experiential information from students it is only possible to speculate about why these steps caused students to stop using Labdog. From personal experience I suspect that abandonment increased at step 7 because of the technical complexity of completing a rotary evaporation. Students are instructed to seek the help of a demonstrator when using the equipment, which is both time-consuming and can be cognitively demanding. Similarly, I repeatedly saw students return to their workstation after performing the vacuum filtration, wherein they felt they have fallen behind, or as though they need to rapidly progress with the rest of the practical. To bring Labdog up to date with their actions would have placed another barrier to progression. Alternatively, students may have simply forgotten to return to Labdog after leaving their workbench for an extended period. Students likely felt as though they have built momentum in their actions, and may wish to carry this forward without being interrupted by questions or other activities not associated with completing the practical.

It is somewhat difficult to design against this problem. One could implement a series of smaller steps, to reduce the average time taken to complete each one, but this would likely interrupt students when they are trying to complete an extended and complex task. As students used laptop computers, they could not easily be transported during these period. Contrastingly, using fewer but longer steps may make students feel like they don't want to return to using Labdog when they've not used it in so long.

A more realistic or logistic solution is to have demonstrating staff remind students to continue using Labdog in moments where abandonment may increase. Once students have completed a more demanding or extensive task, especially if they had to leave their work space, demonstrating staff should encourage and remind students to complete the required steps in Labdog before progressing. This highlights the importance of integration between technology and realistic practice.

4.2.2 Student evaluation

Student pairs were asked to complete a questionnaire at the end of each session. This survey was anonymous and completed online. In total, 65 student pairs responded, of which 56 reported using Labdog throughout the session. Of these 56 respondents, 37 identified themselves as females, 16 males, and 3 did not give gender. Finally, 37 students reported being predicted as performing at an A*-A grade, 16 at a B-C, and 3 did not give grades.

Closed-answer questions

Students were given six closed-answer questions, five of which used five-point likert scale, and one used a binary yes-no response. Within the Likert scale a one represented strong disagreement and five strong agreement.

Overall, students were positive about Labdog. They reported it as easy to use (Figure 4.3), and helpful (Figure 4.4). 17 respondents (30%) rated the difficulty of questions as $\geq 4/5$, suggesting that the majority of students did not find them too difficult.

On a more critical note, 17 respondents (strongly) agreed that ‘Labdog slowed me down too much’, and only 22 responses rated enjoying using Labdog at $\geq 3/5$. Of the 54 students that responded to the question, only a marginal majority (57.5%) of students reported that they would want to see Labdog be used in their college for practicals.

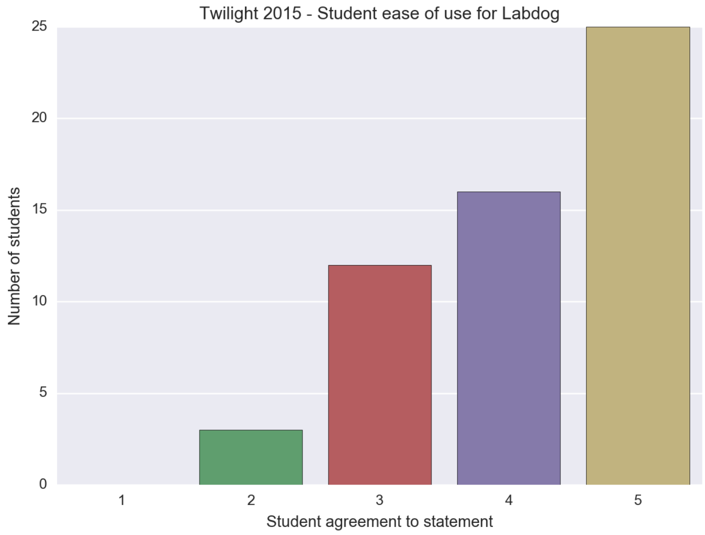


Figure 4.3: Responses from local A-level chemistry students who attended Twilight outreach events in 2015/16. Figure shows number of students who agreed to the statement 'Labdog was easy to use', along a five-point likert scale.

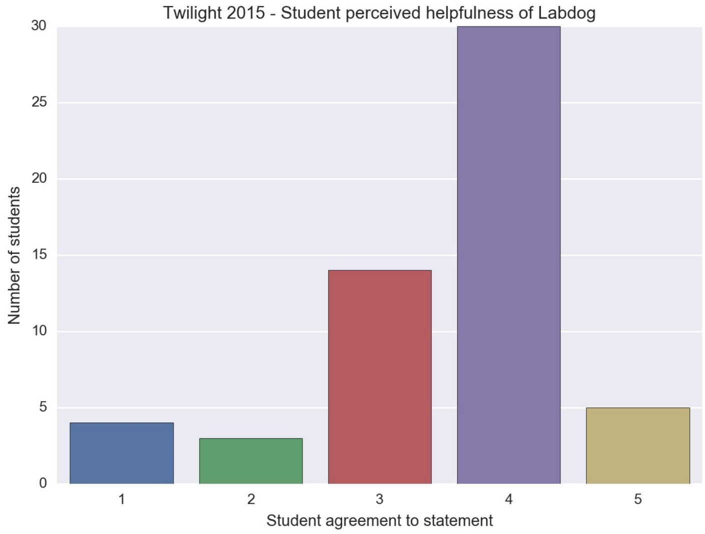


Figure 4.4: Responses from local A-level chemistry students who attended Twilight outreach events in 2015/16. Figure shows students' rating of the helpfulness of Labdog along a five-point likert scale.

Open-answer questions

Two open-answer questions were included in the survey. 51 students provided a comment to the first: 'How do you think Labdog helped you during the session?', with an average of 19.4 words per response (SD=13.1). 23 responses were positive, 3 were negative, and 16 contained both positive and negative elements.

Many (20) students identified how Labdog helped develop their understanding. These comments related to both procedural and conceptual understanding. Comments like "*It was helpful because it made us think about why we were doing certain things*" and " *The questions I was made to do on labdog helped me understand what I was doing a lot better*" are exemplar responses in this category.

As previously found in the SFY pilot, several students felt that Labdog slowed their progress, or added too much work (14). For example one student commented that Labdog was "*long-winded, giving less time for practical work*", and that they would rather be left to just get on "*with the steps of the practical rather than paused to answer the questions* ". Several students noted that they had already been given the instruction on paper, and that Labdog was just a re-iteration of this. However, in counter to this, 19 students made comments that Labdog enhanced the design of their instruction. One student stating that it "*gave us concise instructions*", another that there were "*clear and easy to follow instructions*".

A follow-up question asked students for any other comments, which received a smaller number of responses (17). These were spread across positive (8), negative (6), and a mixture of both (3).

Four students further iterated how the questions in Labdog were beneficial to their learning. Several students (7) commented on the functioning or design of the software itself, suggesting the layout was clear and usable, though perhaps a bit cluttered. Four students identified that the design/execution of the practical could be improved, e.g. with clearer questions, instructions, or feedback on answers.

Contrary to the negative trend seen with SFY students, only a very small number (2) of students reported being distracted by the presence of multiple media. Where SFY students are assessed on their actions in the laboratory (through post-lab skills portfolio) and are required to keep a lab book, the A-level students had no such formal requirements - they were able to simply complete the practical on Labdog. It is also worth noting the greater ease of use associated with using a laptop computer compared to a mobile device with a smaller screen.

4.2.3 Conclusions

The results from this evaluation suggest that students may not necessarily enjoy using Labdog, but that they are often able to identify how it benefits their understanding.

Although instructional design should at least acknowledge student enjoyment of the activity, it should prioritise conceptual and procedural competencies.

Outreach events are fundamentally different to the SFY, and are more about encouraging a student through a practical activity, at times promoting enjoyment over understanding or rigour. Additionally, the populations in SFY and outreach events are different, specifically the twilight evaluation was mostly filled out by high achieving females.

The results from these early evaluations shows how Labdog can be an imposition on students during the lab, though less so than previously seen in the SFY. The added workload, and separation of Labdog from the activity itself may lead to students not using Labdog, which manifests itself in certain steps of a practical more than others. Deciding how activities in Labdog are structured, i.e. what makes up a single step, is likely important in addressing this issue, and so too are the actions of demonstrating or teaching staff.

4.3 Twilight Outreach Events (2017)

Identical twilight events were carried out in January 2017. In October 2016 we started to use Labdog with the SFY, but one final pilot was planned for 2017 in order to provide a discrete, fast, and focused evaluation of the latest iteration of Labdog. Minor changes were made to the practical activity: it was divided into smaller steps, resulting in 16 steps, compared to 14 in 2016.

By this time, I had solved a lot of the deeper technical problems with Labdog, namely better error handling and prevention. Essentially this came by reducing the number of unexpected behaviours that the interface or server of Labdog could not interpret, handle, or work around. At this time, I was able to focus further on the student experience of using Labdog, as well as the demands of educators. Perhaps the single greatest addition during this time was the creation of pre-lab activities, which were available from the offset of the SFY cohort. I also introduced a number of small-scale improvements, e.g. the ability of students to change their passwords or usernames, the ability for educators to approve/reject responses, and for students to re-submit their responses. The freedom and time to add so many significant improvements came from an increased confidence in the potential of Labdog, and having solved a number of deeper, and critical, technical problems.

Date	Started	Penultimate	Last	Pen. (%)	Last (%)
16th	15	10	8	66.7	53.3
17th	23	12	11	52.2	47.8
19th	20	9	8	45.0	40.0
20th	17	6	6	35.3	35.3
23rd	16	9	8	56.3	50.0
24th	12	4	4	33.3	33.3
25th	20	11	11	55.0	55.0
26th	14	9	8	64.3	57.1
Overall	**137**	**70**	**64**	**51.1**	**46.7**

Table 4.3: The number and percentage of students who started and completed (to both the penultimate and final) steps in the January 2017 Twilight events.

Labdog was used over eight days in 2017, but the data from one of those days was removed due to my absence, resulting in no advertising or explanation of Labdog. As with the previous years, 2-4 sixth-form cohorts attended, each institution bringing 5-15 students. Students worked in self-selected pairs, using an Amazon Kindle Fire tablet, not laptops. Students were given an updated five-minute introduction to Labdog at the beginning of the session, and were asked to complete a ten question evaluation survey at the end of it. The survey asked three demographic questions, four closed-answer likert-scale questions, and three open-answer questions.

4.3.1 Student Usage

Student uptake and usage of Labdog can be seen in Table 4.3. Compared to 2016, there is a small decrease in the percentage of students reaching both the penultimate (−8.4%) and final (−4.5%) steps in the practical. This suggests an overall lower uptake from students, but the proportionally smaller decrease in of students reaching the final step suggests those who were using Labdog were more likely to finish the practical. This reduction may simply be the result of natural variation, or perhaps an ongoing issue with wi-fi in certain parts of the teaching lab, or from the change from laptop to tablets.

Similar to the previous year, there were a number of points during the practical where student abandonment of Labdog peaked. This time we see a very high (45; 30.4%) number of students abandon Labdog between steps 10-12 (14, 12, and 19 student user accounts on each step respectively). In these steps, students are asked to dissolve their product in hot acetone, cool their acetone, and then perform a vacuum filtration. A high drop was seen in student-use of Labdog at this point in the practical in 2016. As before,

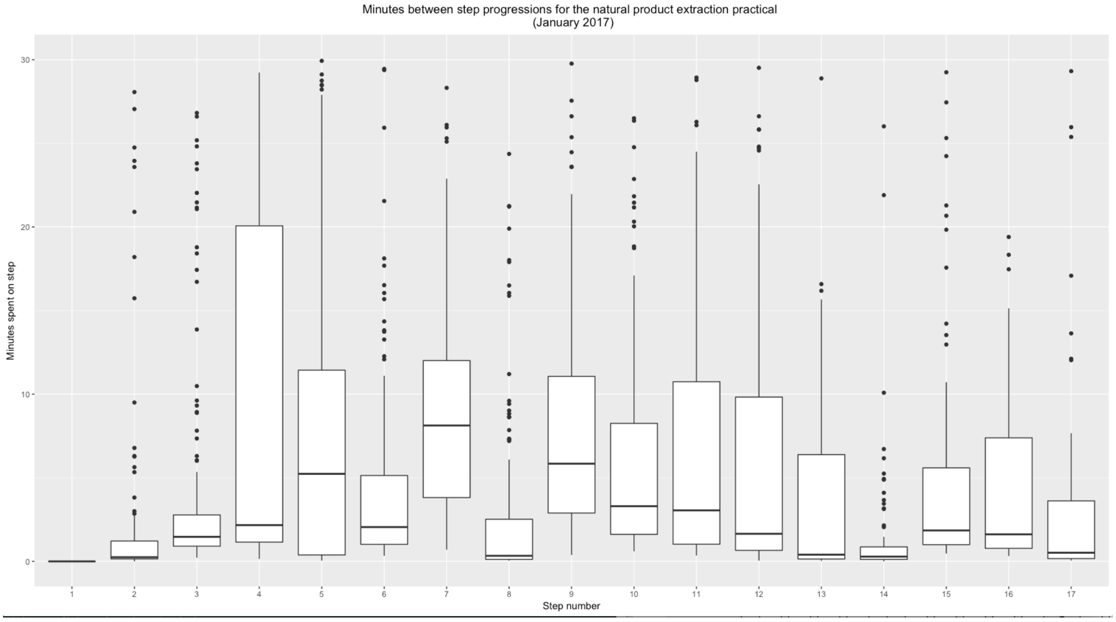


Figure 4.5: Boxplot showing the time (minutes) between student progression from one step to the next in the 2017 January twilight events.

this coincided with time spent away from the workstation, causing students to build momentum in their actions which they did not wish to stop by returning to Labdog.

Unlike 2016, however, this year saw a considerably lower (7; 4.7%) reduction in the number of accounts which stopped using Labdog during the rotary evaporation step. As can be seen by Figure 4.5, step 7 still takes students a notably longer time to complete than other steps. Conversely, steps 10-12 did not take drastically longer to complete. It is worth noting that the difference between the faster and slower quartiles is unevenly spread: the slower half of the student population were further from the average speed than the fastest.

This weakens the argument that steps which take longer are more likely to lead to a reduction in student retention. It is perhaps possible to explain this increased retention on Labdog during rotary evaporation by a focus from demonstrating staff (including myself) on actively encouraging students to use Labdog past this previously troublesome step. Conversely, steps 10-12 present a finer division of three discrete actions, all of which take place away from the workstation and in quick succession - i.e. the students are heating and cooling the same materials in the same vessel, so it is perhaps likely that many students followed their paper-based script instructions at these steps simply because it was easier. This returns back to the idea of momentum being built by students which they then resent having to break in order to answer questions, before carrying on with the procedure.

4.3.2 Student evaluation

A total of 99 students responded to the survey, of which 88 reported they had used Labdog throughout the session. I performed an exploratory analysis on these 88 responses, of which the majority (n=80) were AS students, with only four A2 students, and five 'other'.

The students' opinion of Labdog was overwhelmingly positive. On a 1-5 scale, and 82 rated the ease-of-use as $\geq 4/5$, and likewise 62 students rated it's usefulness as $\geq 4/5$, and 58 reported feeling that using Labdog helped them learn more than if they had not had used it. In this last question, 23 students responded as 3/5, suggesting that Labdog neither benefited nor hindered them drastically. This is supported further by the fact that only seven students rated Labdog as 2/5 and no students rated it as 1/5.

Students were also asked to rate the difficulty of the questions they were asked in Labdog, and responses are shown in Figure 4.6. Data suggest that the questions were not too difficult, with many (36) responses rating it $\leq 2/5$, though most (40) students rated the difficulty at 3/5, suggesting that students felt the questions were appropriately challenging.

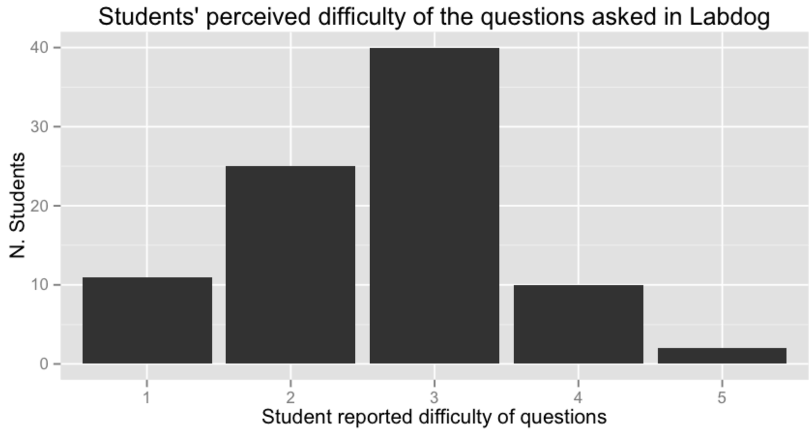


Figure 4.6: The response of 88 AS-level chemistry students about the difficulty of the questions they were asked in Labdog during a natural product extraction outreach event.

I used an open-answer question to ask students if they would like to use Labdog in their college. The majority (82%) of respondents were positive, with 11 saying they would maybe want to use it, and seven saying that they would not. Three of these seven responses stated that they preferred the current paper-base system for practicals, which was also noted by two students who responded 'maybe'. Students believed that paper “help you quickly check if you need to do something immediately after a step” or “quickly write down observations and information”.

Students were asked to detail if they think Labdog helped or hindered their learning during the session. Again, the majority of comments were positive: 63 positive, 2 negative, and 9 contained a mix of both. Several themes emerged, the two most common of which were that Labdog presented the practical instructions in a clear and concise way (26); and also that Labdog helped students think about the chemistry behind their actions (22). A smaller number of students (12) reported directly that Labdog improved their understanding or helped them learn.

A group of students commented on the interface or content of Labdog. For example, one student “liked [the] simple instructions”, and another student stated that “the instructions were clear”. Student comments about the easy navigational features of Labdog present a confusing narrative. Although the students had a copy of the printed lab script, they were encouraged to stay within Labdog, which makes it difficult to say if Labdog was easier to use than the equivalent paper-based lab script. The written instructions in Labdog were identical to those in the paper script, though it is possible that students appreciate being presented a focused view of the practical, i.e. one step at a time. For example, one student reported that Labdog “helped to break up practical into steps”, and another stated that “it’s easier to see what you’re doing while you’re doing it without having to leaf through pages”.

Other students focused on the cognitive benefits of Labdog - namely the questions, and enhanced engagement with the relevant scientific concepts. As one pair of students responded: “It made us really think about what we were doing instead of just adhering to a method and doing things like a robot” and another group of students found “it made us think about the physics of the processes”. Simply put, about a quarter of responses reported that Labdog achieved its desired goal: “It made me question what I was doing”.

4.4 Consequences to Following Research

Though diverse, I do not believe the data discussed in this section represented the richness and impact that these formative experiences had. Someone attending the SFY in 2015/16 would almost fail to recognise the use of Labdog in the 2017 twilights. Improvements in interface and technical stability meant that Labdog worked as intended in these later sessions.

In the 2015/16 academic year there was a seeming separation of Labdog from a teacher and student perspective. One of the biggest breakthroughs came when I engineered Labdog data to become available in real time. Students' progression in a practical, their responses to questions, or their requests for help - are all real-time dependent. If educators were able to see this information at the right time, it could impact teaching 'on the ground' as opposed to in retrospect. By making such things possible, I was able to better integrate Labdog into the educators' workflow.

Additionally, over this time I started paying more attention to instructional design, i.e. how practicals were put into Labdog. Designing steps and questions which allowed students to build momentum where appropriate, and questioning where necessary proved to be a difficult and on-going balance.

That students reported that Labdog made them question what they were doing is incredibly positive. Previous literature suggests that this is both valuable and difficult to achieve. This finding motivated continued work to improve Labdog, despite critical evaluation and experience, and the pilots proved essential in motivating, rationalising, and recording progress against the initial vision of Labdog as a tool to help all students develop an understanding of their actions in the laboratory. This lead Labdog to become a stable and viable piece of technology, enabling it to be used throughout the 2016/17 academic year in a year-long study.

Chapter 5

Working with the Science Foundation Year (2016/17)

After implementing from the formative pilots, Labdog was used in a year-long trial with SFY students in 2016/17. This offered the chance to use Labdog over a longer time-frame. The chapter opens with the structure and content of the SFY, and continues by evaluating students' use of Labdog: their pre- and during-lab engagement. Lastly, I look at the student opinion of the use of Labdog, to examine its impact on learning.

As previously discussed, Labdog was the most advanced, stable, and capable educational technology by the time the 2016/17 academic year began. The final major change came from the addition of real-time features, which I delivered between the first and second laboratory sessions - this provided the first feature added with the during-laboratory time frame in mind. This was crucial in developing Labdog into both a pedagogical and instructional tool, therefore providing value to educators and demonstrators. This resulted in a positive feedback mechanism where greater educator support lead to increased motivation, and focused feedback to further improve Labdog from an interface perspective. I therefore made countless small changes to the software over the academic year, most of which improved the experience of using Labdog.

5.1 The 2016/17 Science Foundation Year

With the exception of one coursework assessed practical, Labdog was used for every laboratory session by the 2016/17 SFY cohort. SFY students attend a chemistry laboratory every two weeks (attending biology labs in the intermittent weeks). Labs were held on a Monday and Tuesday, with students randomly assigned to a group for either day, between 14:00 - 17:00 in a general purpose science teaching laboratory at the University of Southampton's National Oceanography Centre (NOC). Students worked in self-selected pairs during the session.

5.1.1 Student Demographics

The 2016/17 SFY cohort consisted of 46 students (female n=26, 56.5%), of which one was re-sitting externally. Student age ranged from 18-40, with a mean age of 21.15 and a standard deviation of 4.89. The majority (95.65%) of students lived in Great Briton, and 31 (67.39%) were of British nationality.

5.1.2 Structure of the Academic Year

The Labdog pre-labs were released to students on Thursday or Friday of the preceding week. At the beginning of the lab session demonstrators would check students' completion using the Labdog interface. Students who had completed the pre-lab were permitted to start working immediately. Otherwise they were expected to complete the pre-lab activities before continuing.

The university provided 25 7-inch Kindle Fire tablets, enough for each student to use one during a session. The tablets connected to the wi-fi and had a camera, the only essential features for Labdog. Anecdotally there appeared to a be a student prefernece for using their own devices, largely their phones.

A summary of each chemistry lab undertaken in the year is given in Table 5.1. Students used Labdog for each of these sessions, however during week 6 (synthesis and analysis of trimyristin) the students migrated from one server to a newer, higher capacity server. There were a number of logistical and technical problems involved in doing this, and a complete dataset for the students could not be retrieved.

5.2 Student use of Labdog

It is essential to know if, and how much, Labdog was used by students. Such information provides essential context when discussing the impact of Labdog, and is essential to answering the research questions. In the following pages I detail student use of Labdog

Session	Date	Practical A	Practical B	Data?
1	10/10/16	Preparation of copper sulphate		Y
2	24/10/16	Oxidation of Magnesium		Y
3	07/11/16	Preparing a standard Solution	Titrations	Y
4	21/11/16	Coffee-cup calorimetry		Y
5	05/12/16	Identification of unknowns: Flame Tests	Testing for Halides	Y
6	30/01/17	Extraction and analysis of Trimyristin		N
7	27/02/17	Preparation of Esters	Testing for carbonyl compounds	Y
8	13/03/17	Strong and Weak Acids	Titration Curves	Y
9	24/04/17	Preparation of Tetraamminecopper (II)		Y

Table 5.1: A list of practicals completed by the SFY cohort in 2016/17, and if full data is available for analysis.

in the 2016/17 academic year. Labdog evolved continually over the course of the year, and as a result some data is only available for semester two. As such, the discussion and evaluation of Labdog's usage has been split by semester.

5.2.1 Semester one

Students attended five practical sessions in semester one (October-December 2016), completing seven practicals in total. Over these five sessions, 42 unique students logged on to Labdog. Descriptive statistics of student usage of Labdog are given in Table 5.2. Students have been grouped by three categories:

1. **Started** - started the practical on Labdog;
2. **Penultimate** - reached the penultimate step in Labdog;
3. **Complete** - reached the end of the Labdog practical.

An average of 90% of students started using Labdog, with an average of 78% reaching the penultimate step over the five weeks. This suggests that Labdog was used consistently throughout the practical sessions in the first semester.

It is worth noting the high (18.77%) standard deviation in students reaching the penultimate step over the five weeks. This suggests an inconsistent use of Labdog

Practical	Steps	Started		Complete		Penultimate	
		N	%	N	%	N	%
1	17	39	92.86	34	80.95	35	83.33
2	8	41	97.62	38	90.48	38	90.48
3a	8	38	90.48	37	88.10	37	88.10
3b	16	38	90.48	14	33.33	14	33.33
4	10	41	97.62	19	45.24	38	90.48
5a	8	34	80.95	32	76.19	33	78.57
5b	9	36	85.71	32	76.19	35	83.33
Mean		**38.14**	**90.82**	**29.43**	**70.07**	**32.86**	**78.23**
Std		**2.36**	**5.61**	**8.55**	**20.36**	**7.88**	**18.77**

Table 5.2: SFY student completion of all the semester 1 (October-December) practicals in 2016.

over the period, which can likely be attributed to practicals 3b and 4: performing a titration and coffee-cup calorimetry. Removing these two practicals increases the average completion percentage to 82.35%, and reduces standard deviation to 6.65.

In order to better understand when and why students stopped using Labdog I examined these practical activities. Both of these practicals contained iterative tasks: In the calorimetry practical students were required to record a temperature every thirty seconds; and performing a titration involved drop-wise dilution and constant observation. As such, during these practicals students had neither ample time, or cognitive resources, to complete both the practical activity, while using Labdog.

To support this idea, the average time taken between steps in these practicals was greater in this practicals than in others. In a well-designed practical one would expect to see consistent, and relatively low, time between steps. Figure 5.1 visualises the time, in minutes, between progression from one step to another for each practical activity in semester one. The mid-point within each box represents the median, with the upper and lower limits representing the 75th and 25th percentile respectively. Practicals 3b and 4 have both a higher median and variance in time spent between steps. Statistically, the distribution in time (in seconds) between step progressions for practicals 3b and 4 was statistically significant and different from that of all other semester one practicals (ANOVA[1] $P < 0.001$). Unless mentioned, this and all other statistical analysis was completed using R.[2]

---

[1] Both populations were normally distributed, but an uneven sample size prevented an independent t-test.

[2] The statistical programming language. `https://www.r-project.org/`

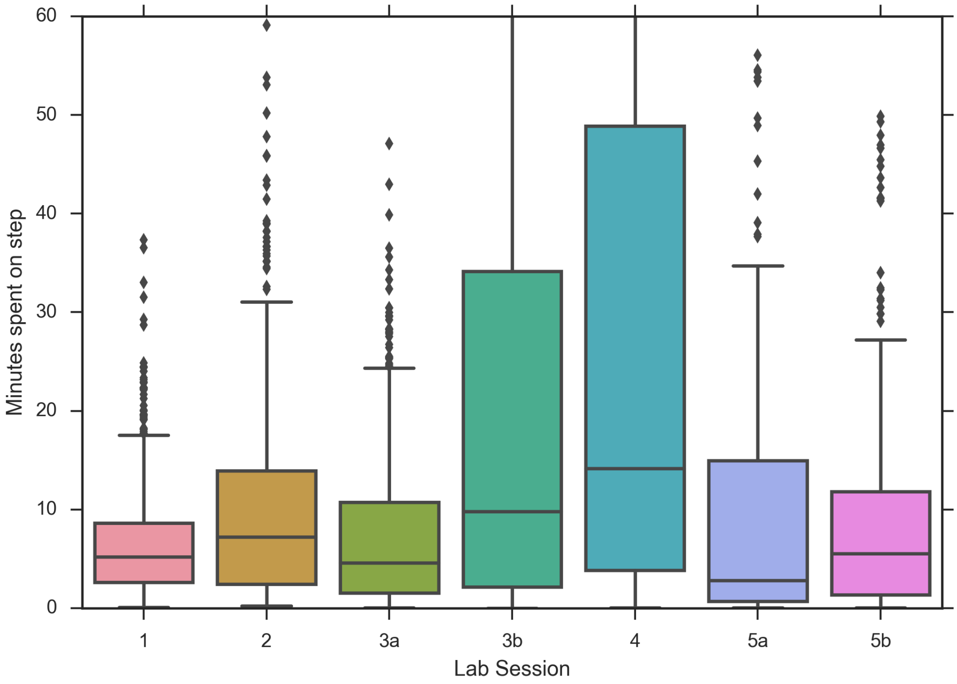


Figure 5.1: A box plot of the number of minutes spent on each step in each practical in semester 1 with SFY students in 2016/17.

Week	Subject	Q	R	S	A?	Completion	Started
7	Esters	5	1	7	N	58.38	33
8	Acids and Bases	8	4	7	N	58.56	36
9	Transition Metals	6	1	9	N	59.39	35

Table 5.3: Details of the pre-lab activities required of SFY students in 2016/17. The number of quesions (Q), resources (R), Steps (S), and if annotations or concepts are required (A) are noted, alongside the mean percentage completion, and the number of students (of N=42) who started the practical activity.

5.2.2 Semester 2

Pre-labs

In semester two, I developed an equal-weighted index to record student completion of pre-lab activities, results for which are shown for practicals 7-9 in Table 5.3. The index equally weights students' completion of each of the pre-lab activities: answering questions, previewing steps, using resources, and giving annotations/concepts. The index only accounts for elements used by the educator, for example if an educator sets questions and requires students to preview the practical, these would account for up to 50% each.

Three pre-labs were used in semester two, creating 126 unique student pre-lab sessions. Of these, 26 (20.6%) remained at 0% completion - of which 15 (57%) came from 5 student accounts which did not complete any pre-lab sessions over semester two. The other 11 unused pre-lab sessions were divided between 9 other accounts, demonstrating that the majority of students were logging into Labdog to complete the pre-lab activities.

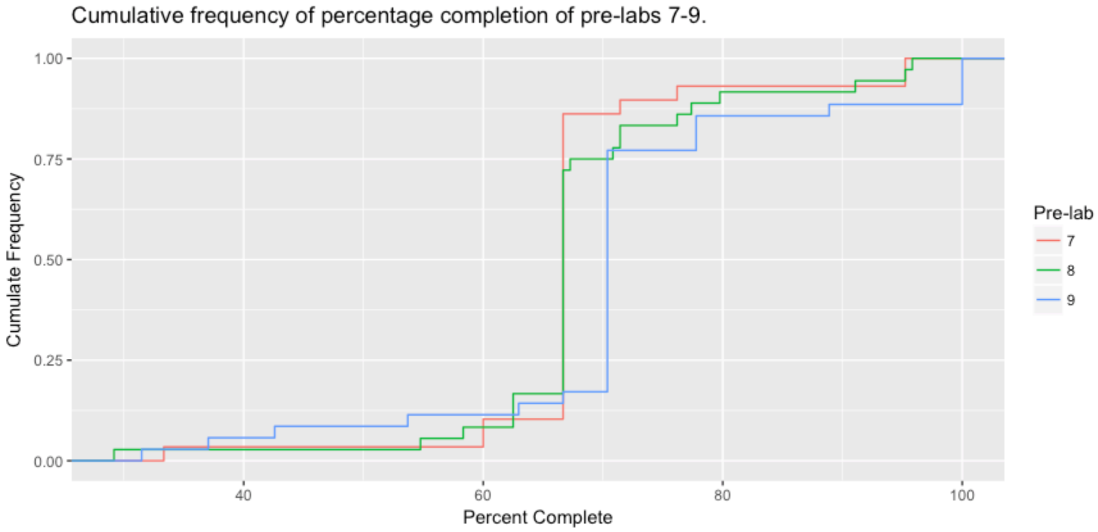


Figure 5.2: Cumulative frequency distribution of non-0% pre-lab completion rates of students in pre-labs 7-9, for SFY students in 2016/17.

The distribution of the 100 non-0% pre-lab sessions are shown, divided by pre-lab, in Figure 5.2. After the initial number of 0% completion, only 12% students completed between $0-60\%$. There is a universal, drastic rise at 60-80%, which accounts for 87% of students. The single largest increase ($\sim 40\%$) falls between $65-70\%$ completion rate, and represents two-third completion of the pre-lab. Further investigation revealed that only the minority of students, $7.4\%, 9.1\%$ and 27.6% for each week, previewed the complete practical in the pre-lab. Comparably $> 97.6\%$ of students completed both the question- and resource-based activities across all pre-labs.

I believe this low level of completion is due largely to the fact that students are given access to, and expected to read, the lab-script in the VLE before attending the session. Although this document does not contain any of the questions that students are expected to answer, it details the practical procedure. This data does not suggest, then, that students were not previewing the practical, but rather simply not doing so in Labdog. It is therefore unlikely that students would see the need to duplicate the work in Labdog, for its own sake.

The other aspects of Labdog's pre-labs can be considered well utilised. Students accessed the resources given to them, and answered the questions posed. Although there were

Practical	Steps	Started		Complete		Penultimate	
		N	%	N	%	N	%
7a	11	29	80.56	28	77.78	28	77.78
7b	7	30	83.33	28	77.78	30	83.33
7c	9	29	80.56	26	72.22	27	75.00
8a	5	34	94.44	34	94.44	34	94.44
8b	7	34	94.44	24	66.67	28	77.78
9	9	31	86.11	24	66.67	25	69.44
$\bar{x}$		**31.17**	**86.57**	**27.33**	**75.93**	**28.67**	**79.63**
σ		**2.11**	**5.87**	**3.40**	**9.44**	**2.81**	**7.80**

Table 5.4: SFY student completion of 6 of 7 practicals in semester 2 (February - March) 2017.

a small number of students who did not complete any of the pre-lab activities, this consides with the number of students we saw drop-out of their studies.

In-labs

In Semester two (February - April 2017) SFY students completed eight practicals over five lab sessions. Data from session six is unavailable, and session 10 is an assessed practical activity. To prevent student confusion and distraction in an assessed environment, Labdog was not used in session 10. The remaining data is presented in Table 5.4. The percentages have been adjusted in semester two to account for the 36 unique students who logged on over the course of the semester - as mentioned previously, these are students who abandoned or suspended their studies.

Second semester data shows comparable student use of Labdog to semester 1 (Table 5.2). There are small increases in the overall percentage of students completing(+5.86%) and reaching the penultimate step (+1.4%) of practicals in Labdog. This may be related to the removal of students who abandoned their studies, leaving students who were more likely to use Labdog. There is negligible change in the standard deviation of percentage of students starting the practical, suggesting a more consistent use of Labdog across the semester.

There is a more drastic change in the standard deviation for both the percentage of students finishing a practical ($S_1 = 18.77, S_2 = 9.44; \Delta = 10.92$), and reaching the penultimate step ($S_1 = 20.36, S_2 = 7.8; \Delta = 10.97$). In relative terms, we are seeing the attrition rates approximately half - suggesting a more consistent use of Labdog.

Evidence of repetitive or tedious actions disrupting the use of Labdog recurred in practical 8b, where students perform a titration. The percentage of students reaching the penultimate step falls suddenly from an average of 80% in all other practicals, to 66%. Figure 5.3 visualises the time (minutes) spent between steps in each practical. Although the median value is comparable to the other practicals, the upper quartile demonstrates that the slowest students were notably slower, compared to other practicals in the semester. Statistically, there is a significant (one-way ANOVA, $P = 0.045$) difference in the time spent on steps (in seconds) for this practical, compared to all others in the semester. This provides more evidence that Labdog is ill-suited to iterative or repetitive activities, likely detracting from students' cognitive resources, in turn having consequences to the related learning.

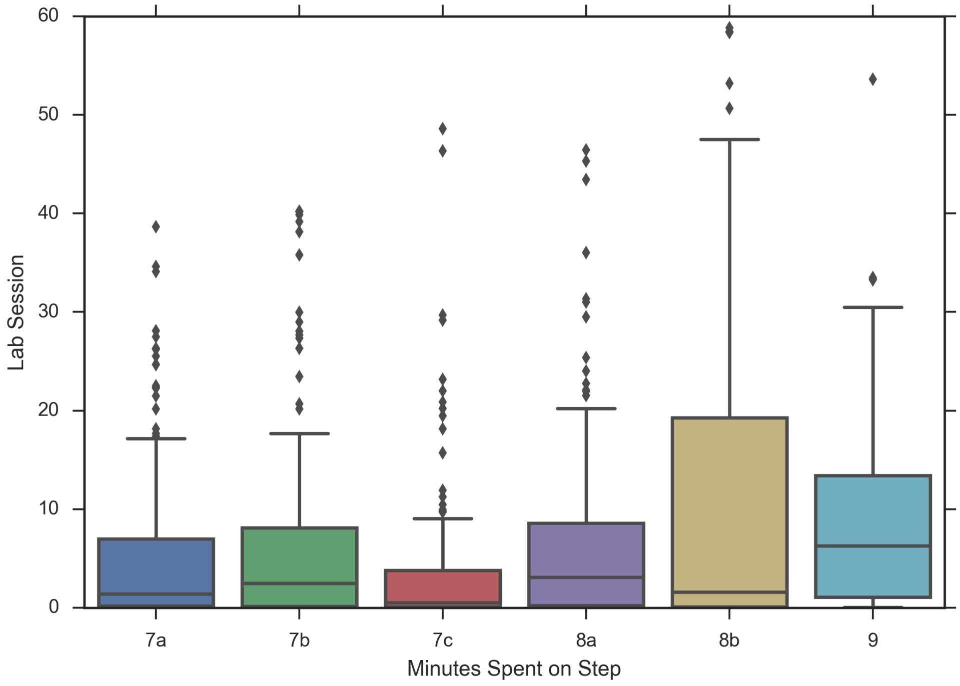


Figure 5.3: Box plot of the number of minutes spent on each step in each practical in semester 2 with SFY students in 2016/17

5.2.3 Student Usage: Conclusions

The second research sub-questions requires us to examine if, and to what extent, Labdog was used by SFY students. To answer this, I constructed an analytic engine into Labdog to record if and when students progress from one step in a practical to the next. Across both semesters, typically between 70-80% of students who started using Labdog would reach the end of the activity in Labdog.

Valuable insight could be gained by examining the practicals where student engagement dropped. Specifically, three practicals over the two semesters saw a decreased usage - each similar in that they required iterative actions or measurements from the students, e.g. the performing of a titration. During these practicals, students were unable to use Labdog at the same time as completing the practical itself, and therefore usage decreased.

Despite these fringe cases, the data presents a good argument that Labdog was well, and consistently, adopted by students over the academic semesters. This addresses a potential fault in the research, wherein Labdog would not be used by students. This provides assurance that data used to answer other research questions has a degree of validity.

5.3 Student Perception of Labdog

The third research sub-question asks about the student-perceived impact of adopting Labdog into the laboratory environment. To investigate this I used both survey and interview methodologies, the benefits of which I have discussed and rationalised is Section 3.1.1 (page 73). In the following section I outline a custom survey conducted in December 2016 – January 2017, and another custom survey conducted at the end of semester two.

5.3.1 Semester 1 Surveys

To quickly assess the quality of Labdog early on, I designed a small survey which was given to students after sessions 1 ($N = 34$) and 4 ($N = 36$). These surveys were intended to provide rapid and broad insights into the technical success of Labdog in the early weeks. I believe technical success in these earliest sessions is essential to developing good rapport with the cohort.

The surveys were hosted online, and were distributed to students via e-mail. Students had to identify themselves using their student ID, and then respond to ten statement items along a five-point likert scale from ‘Strongly Agree’ to ‘Strongly Disagree’. Results to these questions are presented in Table 5.5. The survey also contained four items which students respond to using an unlabelled five-point rating scale, where 5 represents the strongest level of agreement, which are given in Table 5.6.

Question	**Week**	**Response**				
		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
During the laboratory session, Labdog encouraged me to think more about what I was doing and why, more often than I think I would have done if I was not using Labdog	1	11	18	3	4	0
	4	6	20	6	1	1
During the pre-lab activities I would often stop to make sure I understood the related chemistry concepts and ideas.	1	6	17	11	2	0
	4	2	25	5	2	0
During the pre-lab activities I would often stop to make sure I understood what I would be doing during the lab	1	10	21	3	1	1
	4	5	23	5	1	0
During this week's practical session I feel I was actively involved and engaged in what I was doing, not just following a set of instructions.	1	14	19	1	1	1
	4	12	17	4	1	0
I arrived at the practical session with a good knowledge of the steps involved in the practical I was about to conduct.	1	10	20	5	1	0
	4	10	15	3	6	0
My understanding of the chemistry related to the practical improved more in pre- and post-laboratory activities than in the lab itself.	1	5	13	12	6	0
	4	3	9	14	7	1
My understanding of the chemistry related to this practical has improved a lot over the course of the pre-, during-, and post-laboratory activities required of me.	1	12	17	6	1	0
	4	9	21	3	1	0
Overall, Labdog improved how much I learned from this week's practical.	1	11	16	6	3	0
	4	9	16	7	1	1
Using Labdog helped improve my understanding of the chemical concepts relevant to this practical.	1	5	15	12	4	0
	4	1	22	8	2	1
Using Labdog helped improve my understanding of the practical and analytical techniques used over the course of this practical.	1	7	19	8	2	0
	4	3	20	9	1	1

Table 5.5: SFY student responses to likert-scale questions in a semester 1 survey in 2016 after sessions 1 and 4

Question	**Week**	**Response**				
		1	2	3	4	5
How easy was Labdog to use this week?	1	1	3	12	16	4
	4	0	3	7	7	17
How useful was Labdog to you this week?	1	0	3	9	17	7
	4	1	3	9	10	11
I found Labdog more beneficial to my understanding this week compared to last week	4	2	4	8	12	8
I found using labdog easy this week compared to last week	4	2	1	6	11	14

Table 5.6: SFY student responses to numerical answer questions in a semester 1 survey in 2016 after sessions 1 and 4.

Responses present a favourable image of the role of practical work in students' education. It appears that no single phase in the laboratory process, i.e. pre-, during-, and post-laboratory, were universally more valued by students. Students reported that the pre-lab activities helped them both prepare for the practical task and develop an understanding the relevant chemical concepts (81.75% (Strongly) Agree[3]), and the practical procedure (76% S/A).

Students reported that Labdog improved how much they learned, after both sessions one (75% S/A) and four (69% S/A). When separated into conceptual and practical/analytical learning, students still credit Labdog, though less strongly with conceptual (59.72% S/A) than practical/analytical (68.06% S/A). It is possible that this difference is a result of the student-perceived purpose of the laboratory, where they feel practical/analytical gains are more important, easy, or likely. Nevertheless, students reported feeling 'actively engaged' with the practical procedure, and that it was not just a 'set of instructions' (85.5%S/A), though the wording of this statement means this cannot be attributed completely Labdog.

A Mann-Whitney U test was conducted to identity questions with the biggest difference in distribution between surveys. Though no statistically significant differences were found, five questions returned a close to significant ($P < 0.1$) result. Two of these questions asked students if the practical and pre-lab activities were 'worthwhile', but on reflection this wording is too subjective, and without follow-up qualitative insight, such quantitative results cannot meaningfully contribute to the discussion on Labdog. The three other items which remained were:

---

[3] The average of students responded either 'Strongly Agree' or 'Agree', abbreviated to S/A

1. Using Labdog helped improve my understanding of the chemical concepts relevant to this practical. (P=0.089; Figure 5.4)
2. During the laboratory session, Labdog encouraged me to think more about what I was doing and why, more often than I think I would have done if I was not using Labdog (P=0.081; Figure 5.5)
3. How useful was Labdog to you this week? (P=0.65; Figure 5.6)

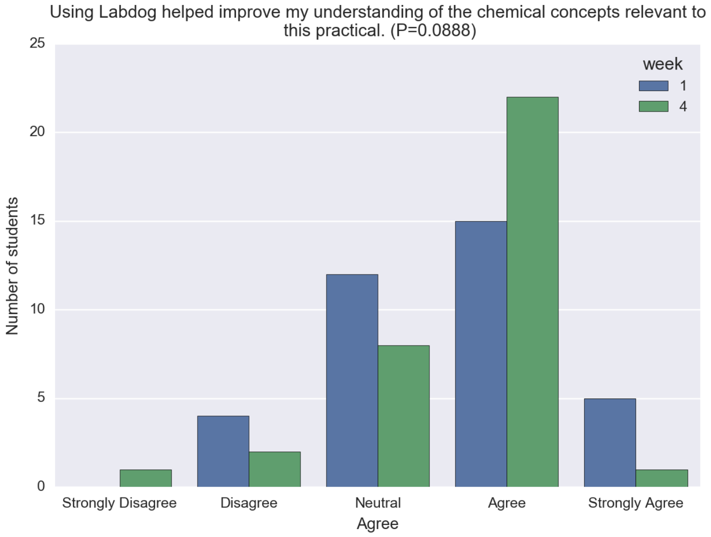


Figure 5.4: SFY student responses to a question on the impact of Labdog on their understanding after sessions 1 and 4 (2016/17)

The first question, Figure 5.4, concerns the perceived development of chemical understanding from using Labdog. Across both surveys most students report that they feel that Labdog helped them develop their understanding. Over time, there is a reduction in the number of non-positive responses., suggesting a higher proportion of students reported Labdog as beneficial to developing understanding over time.

Similarly, Figure 5.5 visualises how more students reported that Labdog caused them to stop and think about their actions and the underlying science, more so than they believed they would not have done without Labdog. There is a clear distribution towards (strongly) agreeing with this statement over time. However, there seems to be a less favourable change between sessions 1-4, with a shift towards more ‘neutral’ and away from ‘strongly agree’, suggesting that Labdog did not provide the same strength of benefit in the later session. As before, this could be due to the design of the activity in Labdog, or perhaps the integration of Labdog into the lab sessions. This therefore potentially presents a continuation of the themes identified previously: that simultaneous

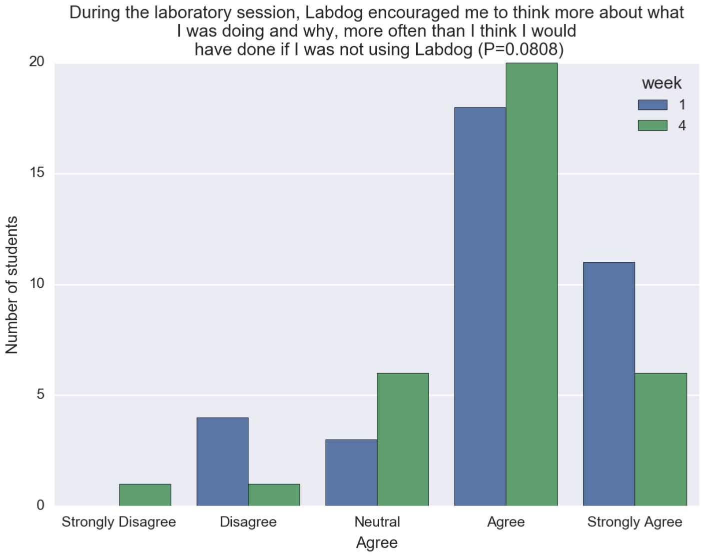


Figure 5.5: SFY student responses to a question on the impact of Labdog on their understanding after sessions 1 and 4 (2016/17)

use of software and certain iterative practical procedures detract from the benefits of Labdog to perceived student learning.

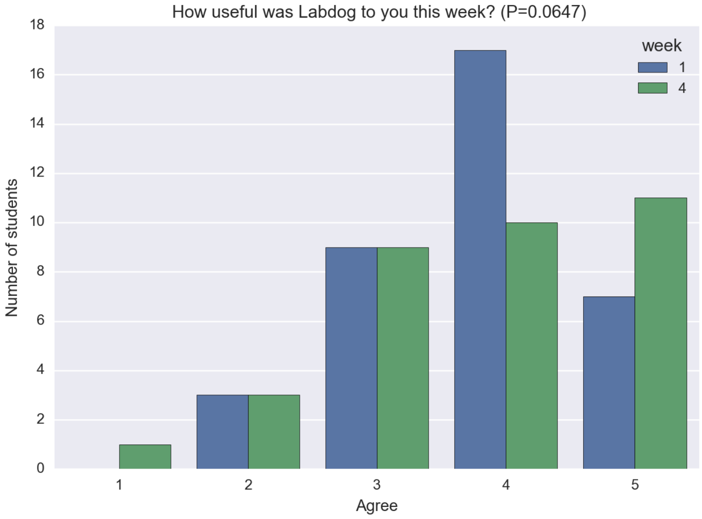


Figure 5.6: SFY student responses to a question on the impact of Labdog on their understanding after sessions 1 and 4 (2016/17)

Visualisation of students' perceived usefulness of Labdog is presented in Figure 5.6. There is an overall positive trend across both surveys, with few students providing responses at levels 1 or 2. Student responses in the second survey shift more towards 'strongly agree' than 'agree', suggesting that even if there were issues with instructional design or technological integration, Labdog was seen as useful by students, though only quantitative data can explain this more fully.

5.3.2 Interviews

To provide such qualitative insight, I designed a semi-structured interview protocol. This extends the undergraduate interviews I conducted (section 3.2), which revealed that the laboratory is seen as inconsistently useful and beneficial. Given the fundamental differences in experience and ability between SFY and UG cohorts, I purposefully limit the scope of comparing the groups directly, however. Instead, I focused the analysis largely on the role of Labdog in promoting learning, characterised through SRL, in the laboratory environment.

5.3.2.1 Methodology

SRL is a multifaceted idea of learning with no agreed definition, though prominent literature (Borkowski et al., 1990; Zimmerman and Campillo, 2003; Schunk, 2008; Bjork et al., 2013) guided me to identify the following components as important to capture:

1. **Affective** (2 questions), which concerned student enjoyment of the laboratory;
2. **Attitudinal** (4 questions), which asked students for their opinion of the teaching laboratory in general;
3. **Cognitive and Metacognitive** (5 questions), which investigated the cognitive and metacognitive processes associated with the laboratory;
4. **Motivations** (2 questions), which examined internal and external motivating factors affecting students' actions in the laboratory;
5. **Use of Labdog** (5 questions), which asked students specifically about Labdog, and their interactions with it

All interviews were conducted by Steve Barnes (SB), a post-graduate researcher and teaching assistant who demonstrates with SFY labs and workshops. I did not conduct the interviews myself, as I believed that students would give untruthful responses about Labdog. SFY students are aware I am the lead developer of Labdog, and am heavily invested in its use. SB has been part of the SFY teaching staff throughout the development of Labdog, and has built a both formal and informal rapport with the

Student	**Gender**	**Repeating?**
Dan	M	N
Elise	F	N
Emily	F	N
Ivy	F	N
Jacqueline	F	N
Matteo	M	N
Megan	F	Y

Table 5.7: List of SFY students interviewed in December 2016

Purpose of the teaching lab	Students
Put theory into practice	5
Visualise theory	3
Complement Lectures	2
Future or Prof. Skills	2
Highlight chemical processes	2
Improve conceptual understanding	2
Improve procedural understanding	2

Table 5.8: SFY students' perceived benefits of the teaching laboratory to their chemical education.

students, giving him both the knowledge and interpersonal relationship to make him preferable to a completely independent interviewer.

Before the interviews, I briefed SB on the interview procedure, explaining each question and the depth of responses I was looking for, giving him a chance to openly discuss the questions. Interviews were advertised to all SFY students by a cohort-wide e-mail, and mentioning it in lectures. SB organised a time and a place to meet each student who volunteered, and I transcribed and analysed all of the interviews from an audio recording. I used the same methodology as discussed with the UG interviews (81).

A total of seven students, representing 16% of the SFY cohort, were interviewed. The students' names, genders, and if they are repeating the SFY can be seen in Table 5.7. The interviews typically lasted between 15-25 minutes.

5.3.2.2 Student-perceived purpose of the laboratory

SFY students were asked to describe the purpose of the teaching laboratory within their own personal chemical education. A summary of the student-stated purposes is shown in Table 5.8. Given what emerged from the UG interviews, the SFY interviews contribute very little to this discussion: students found the laboratory largely as a way to integrate theory into practice, with students distinguishing the development of conceptual and procedural

Similar to UG students, the most common justification of the laboratory was as a chance to visualise and better understand the related chemical theory. Five students mentioned that the laboratory was a chance to link theory and practice: two who mentioned just observing the theory, and the three who mentioned both observing and implementing their theory into practice - a theme which emerged form UG interviews. The quote below is a typical response which illustrates this:

> "It makes you... visualise and think more about what you've learned, because sometimes when you learn it, it's all theory and it's in your head... so the labs help you to apply that and you understand it a lot more " (Emily)

Unlike UGs, SFY students gave fewer and narrower purposes of the teaching lab. This may be attributed to both the fewer number, and shorter duration, of interviews. Additionally, SFY students were much less likely to mention the benefit of laboratory work to their employability or professional competencies. Although this was the least frequently mentioned item by the UGs (12/19 students - 63%, Table 3.6, page 89), it was mentioned by only two of seven SFY students. SFY students are both earlier in their academic career, and are not yet committed to study chemistry at a HE level, which likely explain this.

5.3.2.3 The student-perceived impact of Labdog on learning

Where these interviews provide novel and valuable insight is in the discussion of the impact of Labdog. Students were asked about their overall enjoyment and perception of Labdog. Most students (n=6) presented balanced arguments about Labdog, with one students being entirely positive. This suggests that Labdog was a balanced, non-detrimental, factor to the student experience.

Students reported that Labdog was useful to them: "I would say it's very useful, and I think the intention behind it is sound, and that it is an enhancement to the practical laboratory experience" (Jacqueline). This appears to be a sentiment shared outside of those interviewed:

> "I feel like with anything there are still improvements that can be made, but in terms of how things are right now, it seems to be helping me and

How Labdog Improved Amount Learned	Students
Feedback from teacher	4
Answering questions	4
Helps engage with practical	1
Centralised record of activities	1
Pre-labs	1
Step-by-step guide in Labdog	1

Table 5.9: The reasons for Labdog improving the quantitaty of material learned during practical sessions, given by five SFY students.

How Labdog Improved Quality Learned	Students
Answering questions	4
Engage with the practical	3
Access formative feedback	2
Highlights scientific concepts	1
Provides constant activities	1

Table 5.10: SFY student perception of how Labdog increases the quality of the material learned.

> everyone else that I have spoken to interns of understanding the chemistry."
> (Megan)

Specifically, students appreciate the enhanced conceptual focus, and simplification of workflow from Labdog. However, students often reported drawbacks from technical problems. To understand the benefits to learning we need to examine answers given in the 'cognitive and metacognitive' phase of the interview, where students were asked to comment on the amount, quality, content, and mechanisms behind their learning in practical work.

When asked to comment on the amount of material learned, five students stated that it helped them learn more material, one stated that it had a mixed impact, and another that it had no impact. The reasons for increasing the amount of material learned can be seen in Table 5.9, showing the dominance of question-based learning and formative feedback from demonstrators.

In a similarly encouraging finding, five students reported that Labdog improved the quality of their learning, while two responded that it had a mixed effect. Broad categories of student answers to these question are given in Table 5.10. Again, question-based learning and formative feedback emerge as dominant factors. Three students also report on engaging more with the practical.

When discussing how using Labdog changed learning from practical work, no new factors emerged. This suggests a saturation of ideas about the mechanisms of student learning from and with Labdog (Fusch and Ness, 2015). As for evidence regarding the impact of questions in Labdog, one need only to look at the ubiquity of students' positive comments on the nature of questions in Labdog. Below are a series of comments, one from each of the students interviewed, in which they talk about the benefits of being asked question in Labdog:

> "I think the idea of Labdog is that you answer it whilst you're doing it, which is much more likely to be beneficial, because you can then understand while you are doing, and you're more likely to understand what you're going to do next or what you have done previously" (Elise)
>
> "Rather than its just seeming like one big subject that I don't understand, everything is broken down simply. We've got Labdog with us always and there are always people on hand to help us with the questions even if they seem too big" (Megan)
>
> "To actually do it and understand it at the same time, and Labdog really does do that, it really does emphasise on 'you've got your results here, what do they show you?' 'what does this colour mean? What does the precipitate mean?'. It just helps to make you understand it." (Dan)
>
> "I think Labdog has more of an effect in the chemical concept side of things, this questioning of what have you seen and why have you seen it. Asking those questions, which, I don't know if it was just written down in the schedule, but it just seems much more interactive in the Labdog format." (Jacqueline)
>
> "again it is the question aspects, You actually have to think about it you can't just fly through it... you can get extra help if you need it" (Matteo)
>
> "you answer those questions and you have the lecturer coming up to you and correcting you, and I think that is really good... And I feel like that is really memorable as well, I think it is really helpful... I don't think I would've considered that if it wasn't on Labdog" (Ivy)

> "I think it has made me sit there when I'm doing things and think 'why am I doing that?' and it makes me think about why I am doing that, and I get me to think about why we're doing certain practical things, or safety things" (Emily)

Students clearly see educational benefits from using Labdog. Students' previous attribution to question-based learning and formative assessment were hugely influential considerations throughout future reflection and development (Section 3.4) of Labdog. I wish to discuss both of these elements in further detail, to help better understand how students felt Labdog benefited their learning.

Real-time and compulsory questions. Having to answer questions in the lab, while completing the practical, was repeatedly identified by students. For example, Elise stated that when asked questions during the flame tests practical, Labdog meant that "you're not just looking at pretty light, you're wondering why it's coming out that colour, something that you probably wouldn't consider if you didn't do the practical [with Labdog]". Several other students also identified this:

> "It is not just about doing the practical in list, it's about understanding what you're doing as well. So I think it really helps with that." (Ivy)

> "I think it definitely has positively affected how you think about what you're doing in the practicals, because for me I often just do things and then just look at it, and you'll follow the method and then just carry on because you've done it, with the results and the graph, and you'll go 'okay, cool' " (Elise)

> "yes I think it has made me sit there when I'm doing things and think 'why am I doing that?' and it makes me think about why I am doing that, and I get me to think about why we're doing certain practical things safety things" (Emily)

> "in Labdog at the time you're learning about it, you're understanding it at the time that you're doing it. [yeah] You can go away knowing what oyuve done there." (Dan)

Students reported that their conceptual understanding was improved by being asked questions about the chemistry as it arose during the practical. Megan, who was repeating the SFY, admitted that she previously ignored concepts she did not understand: “last year I used to go in, bash [the practical] out, get out as quickly as I could because I didn’t have a clue what was going on. So when it came to the end of the lab, you know drawing enthalpy diagrams or doing calculations, most of the time I will close my book can be like ‘I’ll do that at home’ but I wouldn’t do it at home because I didn’t understand it”. But being forced to engage with questions, specifically helped Megan, especially when they are designed to address minute details of particular concepts:

> “Whereas this sort of ties it together and breaks it down, like I said, if I didn’t understand one concept in chemistry and that is the concept for the practical, That will be broken down into like 25 different questions which really, they simply spell it out, I know sometimes it seems like the questions are silly but you need them to be broken down as simply as possible, just so that when you look back in it on the Sunday or the Saturday, or whenever you do it you can see the steps as if you’re teaching it to someone else which to me anyway is how I revise, just seeing it step-by-step and you can see the way the knowledge is building on previous knowledge.” (Megan)

Similarly, Ivy implied that being required to answer the questions meant that she didn’t need to actively pursue help, and that teaching staff:

> “actually come up to you and then you realise that you made a mistake, They correct it for you, And I feel like that is really memorable as well”

The ability of educators to help students overcome problems during a class was one of the motivating factors which caused the emergence of flipped teaching (Bergmann and Sams, 2012). This may specifically benefit lower ability students who may not recognise their misunderstanding and actively seek help (Kruger and Dunning, 1999; Zell and Krizan, 2014).

Questions were not always seen as positive by students. It is important to consider what questions are being asked, and when they are given to students. A similar series of concerns has arisen in the discussion of CRS, one which was addressed in physics by the presence of the forces concept inventory (Hestenes et al., 1992), which provided physics educators with a series of pre-made, well-focused questions which assessed identifiable concepts, e.g. Dufresne and Gerace (2004).

Unfortunately, no such resources exist for this context. The development and palcement of questions was performed *ad hoc* by the teaching staff, and one which developed continuously through discussion with staff and reflection on experience (Wilson and Read, 2017).

Students' responses provide some insight into what makes a good, or bad, question. Two students stated that the time and effort required engage with Labdog as detrimental to both the practical and conceptual aspects of the lab. Elise referred to Labdog as "tedious at times... because you're trying to fill out Labdogs and steps, and sometimes it's just a waste of the quality of your time in the labs". Similarly, Matteo stated that sometimes answering questions would cause him to "lose a bit of the flow of the experiment, but in the long run it is better, I imagine, because I am actually thinking about what is going on". Lastly, Ivy stated her frustration that "you can't move on using Labdog unless you answer that question, but it kind of slows you down and you think 'I just want to get on with the practical now' ".

These students all refer to the experience of answering the questions, rather than the content or focus of the questions themselves. Specifically, students berated questions which were long-winded, or prevented them from feeling like they were making progress. Psychologically this is similar to the concept of 'flow' (Csikszentmihalyi and Csikszentmihalyi, 1992), where sustained immersion in a problem or action improves outcomes. This resembles the experiences discussed in the Twilight pilots (Chapter 4), where students were most likely to abandon Labdog when in phases of extended novel or complex procedures, e.g. using a rotary evaporator. This raises the pragmatic concern of the remit and definition of a step, and the wording or focus of a question - which differs from the theoretically-focused development of concept inventories (Krause et al., 2004). This exists within the idiosyncrasies of the laboratory environment, which the adoption of DBR can effectively acknowledge and incorporate.

Teacher-led formative feedback. The immediate verbal and written feedback from demonstrating staff in the laboratory appears to be beneficial to students' learning during practical work:

> "I can sit in a laboratory situation without Labdog, and write a load of rubbish, and the teacher might not ever know if I've written it right or wrong, and Labdog just enables the teacher to actually understand that I do have an understanding of it." (Dan)

> "you answer those questions and you have the lecturer coming up to you and correcting you, and I think that is really good. And Whereas in biology

> workshops you're just doing the practical and you're not really, If you are having issues it is hard to ask for help." (Ivy)

> "I think Labdog's capacity to facilitate this timely feedback, is like one of its greatest positives. And that has enhanced the quality of the lab experience." (Jacqueline)

Such a finding is largely unsurprising as formative assessment has been repeatedly proven as a beneficial educational tool (Nicol and Macfarlane-Dick, 2004; Sadler, 2010), which can be enhanced through technology (Beatty and Gerace, 2009). Labdog can be used to record students' misunderstandings, which is especially important in helping those with novice-level understanding (Kirschner et al., 2006), and direct teacher instruction to improve them.

This validates the integration of CRS principles into Labdog (Section 2.2.1). Questions which cause students to demonstrate misunderstandings provide material for discussions and corrections with students (Crouch and Mazur, 2001). The real-time nature of Labdog means such conversations can happen when the content is still relevant and students are familiar with the incorrect thought processes or rationale which lead them to submit an incorrect response. This heralds to the original purpose or justification of flipped teaching (Bergmann and Sams, 2012).

5.3.3 Semester 2 Surveys

Though not a unanimously positive experience, by 2016 Labdog presented itself as a stable technology. Following from this, I designed a more survey to follow up on the issues raised in semester one.

5.3.3.1 Methodology

I designed a quantitative survey which was delivered at the end of semester two. The survey contained 19 item: 17 closed-answer statements and two open-answer questions. All statements were preceded by the qualifier "Over the course of this entire academic year...", and students stated their agreement to the statements using a five-point likert scale from 'Strongly Disagree' to 'Strongly Agree'. For brevity, in the results I record these as the degree of agreement 1-5, where 1 is Strongly Disagree and 5 is Strongly Agree. The items fell into five areas, identified and discussed in the previous sections:

1. **Application** - Applying, and recognising previously-learned chemical concepts during laboratory work;
2. **Engagement** - Students' reported interaction with chemical concepts;
3. **Learning** - Other questions about the learning process;
4. **Technical** - The occurrence and subsequent interference of any technical or infrastructural problems from Labdog as a piece of software;
5. **Understanding** - The development of understanding of underlying chemical principles.

The two open-answer questions were designed to collect qualitative information on the student-perceived usefulness and benefit of Labdog to their learning:

1. Please explain if and why Labdog was/not useful to you this year.
2. Would you consider using Labdog to be worthwhile? Explain your answer.

5.3.3.2 Results

A paper version of the survey was given to SFY students during the penultimate week of the second semester, and their responses were digitised. 22 students ($\sim$ 55% of the cohort) responded to the survey. Below I discuss each thematic area in turn, and then the open-answer question.

A Cronbach's alpha test was conducted for each thematic area of closed-answer questions to determine their interrelatedness, to provide a more targeted and specific discussion (Tavakol and Dennick, 2011).

Application of Chemistry

Students repeatedly reported that the laboratory helped them apply concepts to problems. I created three questions, shown in Table 5.11, to examine these aspects. A Cronbach's alpha of responses of 0.893 suggests the items were highly inter-related.

It appears that Labdog helped students feel as if they were applying chemistry to their actions. Most notably, 12 (57%) students responded 'agree' or 'strongly agree' (S/A) that Labdog helped them recognise when they were 'putting [their] knowledge of chemical

Question	Level of agreement				
	1	2	3	4	5
Labdog helped me see the chemistry in what I was doing	0	6	5	9	2
Labdog helped me better recognise when I was putting my knowledge of chemical theory into practice	0	7	2	9	3
Labdog helped make me feel as if I was applying my chemical knowledge to real world scenarios	0	9	8	3	2

Table 5.11: SFY student responses to questions about Labdog in relation to the application of chemistry. All statements preceded by the qualifier "Over the course of this entire academic year...".

theory into practice'. A similar number (11, 52%) responded S/A that Labdog helped them 'see' the chemistry in what they were doing.

No students strongly disagreed with either of these statements, though a number of students disagreed or remained neutral. This neutrality suggests that approximately half of the respondents do not feel Labdog had any impact on them recognising the chemistry related to what they were doing. Additionally, about a quarter (27%) of students disagreed that Labdog helped them better see the chemistry in what they were doing.

Only a minority (5, 23%) of students reported that using Labdog helped them feel as if they were applying their chemical knowledge in a real-world scenario. Disagreement and neutrality to this statement are comparable, 9 and 8 respectively (77% total). It seems possible that this issue relates more to the instructional design of Labdog as a whole, which does not have a focus on problem- or project-based learning. It is plausible that students are aware that they are completing introductory laboratory activities designed to familiarise them with equipment and concepts. Students are likely aware that this does not resemble an industrial or research environment.

I designed Labdog to integrate with the existing instructional design of the SFY. In turn, the laboratories are constrained by the financial and equipment resources, time, and the need to effectively map on to existing Further Education-level qualifications. Many of these activities resemble cookbook chemistry, where students are given a set of instructions to follow. Although certain practicals presented students with decisions, e.g. in the choice of chemicals during a titration, students are likely aware that the lab does not resemble a 'real-world' setting. Labdog was not designed specifically to help

Question	Level of agreement				
	1	2	3	4	5
Using Labdog made me feel more engaged with the practical work	2	4	3	8	3
Using Labdog encouraged me to ask more questions about my actions in other laboratory work (e.g. Biology)	0	12	2	4	4
Labdog encouraged me to think more about what I was doing, and why, during laboratory sessions than I think I would have done if I was not using Labdog	0	4	2	11	5
Using Labdog helped me better visualise the underlying chemistry during the practical session	0	5	3	12	2
Using Labdog made me think about some aspects of the chemistry that I would not have otherwise thought about	0	3	2	12	5

Table 5.12: SFY student responses to questions about Labdog in relation to the engagement with chemistry. All statements preceded by the qualifier “Over the course of this entire academic year...”.

educators highlight the application of chemical principles into real-life settings, so called context-based education (Parchmann et al., 2006; Christensson and Sjöström, 2014).

The limit of this work’s scope prevented it from redesigning the syllabus. As such, I believe that the finding that Labdog does not help students see their actions in terms of ‘real world scenarios’ is important in guiding future work, but does not reflect on the educational shortcomings of Labdog.

Engagement

Previously, SFY and UG students reported that the laboratory helped them feel as though they were engaging with the relevant scientific principles. I designed five questions to collect information on this, shown in Table 5.12. An alpha score of 0.826 was returned for these items, suggesting high inter-relatedness.

In general, these responses are favourable to presenting Labdog as a pedagogically-focused educational technology. Half of the students responded S/A to the statement that Labdog helped them feel more engaged with the practical work, with six (27%) students strongly/disagreeing (S/D). To expand on this, the majority (16) of students agreed that Labdog encouraged them to ‘think more about what I was doing, and why, during laboratory sessions’, compared with only four students who responded S/D. Similarly, the majority of students (17) reported that Labdog made them think about some area of chemistry that they would not have otherwise thought about. This may be attributed

Question	Level of agreement				
	1	2	3	4	5
Overall, Labdog has had a positive impact on my learning	1	2	7	10	2
Labdog has not made me a better chemist	1	9	5	5	1
The data available in Labdog (e.g. photos and observations) has made completing my skills-portfolio easier	0	8	6	6	2

Table 5.13: SFY student responses to questions about Labdog in relation to the learning of chemistry. All statements preceded by the qualifier "Over the course of this entire academic year...".

to the majority (14) of students who report that Labdog encouraged them to actively visualise the related chemical concepts.

These results are incredibly encouraging, showcasing Labdog as a beneficial educational technology. They suggest that Labdog has been successful in making students more aware of the underlying chemistry to their actions. Though by way of caveats, it does not appear that the use of Labdog encouraged students to approach their biology laboratories any differently. This suggests that Labdog perhaps offers only context-specific benefits, where the design and structure of the laboratory sessions are crucial.

Learning

Several items about the broad learning process were either too broad to fit into the other categories or did not thematically fit. I designed four questions of this type, which are presented in Table 5.13. A Cronbach's alpha of 0.141 was returned, suggesting that these items did not measure the same underlying factor, likely due to this breadth.

It appears that only a minority (3) of students would disagree with the statement that Labdog had an overall negative impact on their learning. However, of the remaining students, 12 agree and seven remain neutral. These results suggest that while Labdog some positive impact on learning, there is a notable portion of respondents (31.8%) who feel it made no significant difference either way.

A similar pattern emerged when students were asked if they think Labdog makes them a better chemist: with 10 agreeing that it does, six disagreeing, and five neutral responses. This positive trend is encouraging, with only a quarter of students disagreeing that Labdog makes them better chemists. While it is important to acknowledge that this is

Question	Level of agreement				
	1	2	3	4	5
The technical problems I had with Labdog severely reduced its impact on my learning	2	8	6	4	2
Labdog became more technically stable	0	3	4	11	4
I had frequent technical problems with Labdog	1	5	6	8	2

Table 5.14: SFY student responses to questions about the technical aspects of Labdog. All statements preceded by the qualifier “Over the course of this entire academic year...”.

not a unanimous trend, this provides further support for the notion that Labdog was a successful ediucational technology.

The last item examined the use of Labdog as an e-portfolio, however no clear benefit emerged from responses. An equal number of students, eight, both agreed and disagreed to this statement, with six neutral students. Such a mixed result does not support the notion that Labdog integrated well with existing coursework or assessment structures. This is likely a result of the separation between the students’ assessed skills portfolio and the use of Labdog. Although Labdog may have complemented such activities, the two remain fundamentally separate, with students using word-processing software and the VLE to complete these assignments.

Technical

At numerous points in the research, students raised concerns about the technical instability of Labdog. Given that I made significant efforts to improve the stability of Labdog in preparation of 2016/17, it is essential to then evaluate the impact of these changes. These questions, and their response counts, are shown in Table 5.14. Responses to these questions returned a Cronbach’s alpha of -0.36, suggesting these items do not measure a unifying factor. The alpha value is negative is a result of negative correlations between items.

As expected, the majority (15) of students agreed that Labdog became more technically stable over the course of the academic year, with only 3 disagreeing. Despite this, half (11) of respondents S/A that they experienced ‘frequent technical problems’ with Labdog over the entire academic year. The fact that this is not a universally shared opinion suggests that these likely come in the form of one-off events, e.g. students being logged out of Labdog after periods of inactivity, or an inability to upload a specific photo.

Question	Level of agreement 1	2	3	4	5
Labdog helped improve my understanding of the chemical concepts related to the practical work at hand	0	3	4	12	3
The questions asked in Labdog during the laboratory were very helpful for developing my understanding of the related chemistry	0	1	6	11	4
The feedback from teachers / demonstrators in the lab, based on my answers to the questions in Labdog, helped improve my understanding much more than just answering the questions alone would have done	0	2	1	9	10

Table 5.15: SFY student responses to questions about how Labdog helped develop their understanding.
All statements preceded by the qualifier "Over the course of this entire academic year...".

Six (27%) students agreed that these problems 'severely reduced' the impact of Labdog on their learning, while around half (10) disagreed that this was the case. Therefore while it is important to take note of these issues, to guide future technical improvement to the software, it heartening to see that such reports come from only a minority of students.

Understanding

There is an implicit need to examine how, and if, students felt Labdog impacted their understanding of the chemistry related to their practical work. The following section completes the image of Labdog which has been constructed through the previous sections, shedding more light on the impact of Labdog as an educational technology. I designed three questions to probe this aspect, given in Table 5.15. A Cronbach's alpha of 0.754 suggests a moderately-strong interrelation between these items.

Labdog helped students 'improve my understanding of the chemical concepts related to the practical work' for the majority (15 S/A, 68%) of responses, with only a minority responding neutral (4) and negative (3). Though not unanimously beneficial to students understanding, one could easily argue that Labdog has had a more positive impact on student-perceived learning.

Unsurprisingly, the majority (15, 68%) of students agreed that the in-lab questions in Labdog were 'very helpful to developing my understanding'. Both formative feedback and student-educator interaction were attributed by the vast majority (19) of students. These responses provide further strong evidence that Labdog is most empowered to promote learning through formative feedback.

Open-answer questions

Student-perceived usefulness

18 students responded to the statement ‘please explain if and why Labdog was/not useful to you this year’. Responses were classified as either positive or negative, and the two categories were non-exclusive - i.e. a student can report Labdog to be of mixed use. Responses were split almost evenly: five students responded that it was useful, six that it was not, and seven provided a mixed opinion.

The most prominent benefit from students concerned the asking of questions. Six students identified the questions as useful. For example, one student stated that “the questions did cause me to stop and think about what was going on with the chemistry of the practical”, and another that “they made us think about the chemistry”. The subsequent feedback from educators, based on students’ responses, was mentioned by two students, one of whom stated that this made Labdog “worth it for this alone”. This agrees with data from surveys and interviews, as well as wider psychological (Karpicke and Grimaldi, 2012), pedagogical (Clark, 2012), and educational technology (Beatty and Gerace, 2009) literature. Labdog has succeeded in achieving the benefits of question-based learning and the subsequent teacher-led formative assessment, which are prolific and praised in the literature, and which guided the creation of Labdog.

Though not removed altogether, Labdog has likely make cookbook chemistry less prominent. As one student stated: “it tested my understanding of the key chemical ideas... without Labdog I would have just followed the script to complete the practical”. This idea emerges in another student’s response, that “I was forced to think about/reflect on topics as opposed to having a superficial understanding”. The prominence of cookbook chemistry in the literature, and my previous experience, was fundamental to the conception of Labdog.

Existing literature attempts to achieve the same thing through a more significant redesign of laboratory curriculum/structure, or the integration of some more external technology. Some educators have used evidence-based portfolios (McClean et al., 2016), electronic lab notebooks (Van Dyke and Smith-Carpenter, 2017), or digital badges (Seery et al., 2017) as additions to their existing structure. Others have taken more radical re-designs to include problem-based learning (Polman, 2000) or inquiry-based practicals (Hunnicutt et al., 2014; Zimbadi et al., 2015). The work with Labdog represents minimal change to the structure of the practicals themselves, and although it has required a shift in educator practices to rely more on the information provided through a web-interface, both educators and students are largely performing the same actions in the laboratory: performing a pre-planned practical, and helping correct students' misconceptions or errors respectively. The evaluation of related studies, like those cited above, have often relied heavily on the educator narrative of logistical and technical success, with little attention to the impact on student learning. Contrastingly, this research demonstrates a middle-ground between the two: step-by-step practical activities, typically associated with cookbook chemistry, delivered alongside the need to evidence actions and understanding, subsequently eliciting evidence of (mis)understanding. This process appears to have benefited both student and teacher.

Five students stated that Labdog helped them develop their conceptual understanding. One student identified that Labdog helped them "to visualise and understand chemistry practically as much as theoretically". Three students stated that using Labdog allowed them to relate and connect the chemical ideas involved in practical work to the theoretical concepts covered in lectures. This is a crucial point, and one which relates to the chemical triplet, one of the fundamental concepts in chemical education (Talanquer, 2011; Taber, 2013). Having students visualise, utilise, and communicate submicroscopic processes represents a significant problem and area of interest in chemical education - one which has often been overlooked (Flood et al., 2014). This demonstrates the potentially unique pedagogical benefits of Labdog, and an area for future work, i.e. examining when and how the most relevant learning takes place.

A number of students were also openly critical about the impact of Labdog. Such comments are essential in the iterative improvement process associated with DBR. Students most commonly criticised Labdog for distracting them from the practical activity at hand, or for being technically unstable. One third of students felt that Labdog distracted them by adding too much to their workload. Exemplar statements are that "it took too much time and attention from carrying out the practical", or that "it took my attention away from the practical I was completing as most of my time was spent writing in Labdog the steps".

Similarly, two students felt that Labdog produced undue stress: "it only added stress and pressure during the labs as it always made us fall behind", and "Labdog distracted me from focusing on the practical to the point where I made mistakes and finished the practical late". One other student felt they were unduly duplicating work between Labdog and their post-laboratory coursework.

Most technical comments were non-specific, citing "too many technical problems", "technical flaws", or "it wasn't working correctly". Several students provided more specific detail, e.g. one student mentioned "constant issues with uploading pictures", while another two critiqued the graphical interface, saying it wasn't "mobile friendly" or "hard to navigate". Only one student reported technical problems were resolved over time.

Four students critiqued the design of questions, e.g. "on occasion there were too many questions", or a request for easier questions because "harder questions = confusion = lack of confidence". Other students commented on the nature of the steps themselves, e.g. for having "unnecessary steps, many of which were later removed". Similarly, one student stated that they simply preferred to use paper over digital technologies.

Was using Labdog worthwhile?

In total, 21 responses were given to the question 'would you consider using Labdog to be worthwhile? Explain your answer'. The majority (14) of students reported that using Labdog was worthwhile, with five students disagreeing, and two providing unclear statements.

Unsurprisingly, Labdog was most commonly seen as useful because of the questions. Five students stated that Labdog was worthwhile because of the benefits to their understanding: "not only do you conduct an experiment but you get to know principles behind [it] and concepts". Another five students identified that Labdog helped them feel more consciously engaged in the underlying chemistry, e.g. it "made me think more into what was happening during the practical".

Educator feedback from the questions was mentioned by another three students: "that is when it is most useful as you can discuss things which otherwise may not crop up. That is when Labdog was most successful". It was not so much the questions themselves, but the subsequent feedback and interactions with educators. One student identified the pre-lab activities as making Labdog worthwhile to them, because it allowed them to arrive to labs with “a foundation of understanding of what I had to do for that experiment”.

Five students did not agree that Labdog was worthwhile, presenting a more varied series of justifications. Two students cited the technical problems with Labdog as sufficient to make Labdog not worthwhile, “too many system crashes, taking up too much time” and “lots of technical problems that deleted my answers I had already entered”.

Two students seem to be against the notion of using digital technologies. One of these students stated that Labdog would be worthwhile “probably for other people...I find it easier to just have the paper script in from of me”. The other stated that “I think there are other ways (more productive) to stimulate learning in the labs (minus technology)”. These answers do not highlight any specific attribute of Labdog, but rather suggest against using a LaRS approach in general.

Two other students cited the issues of time demands or distractions. One student refers to their previous answer on the usefulness of Labdog, in which they cite that Labdog distracted them from the practical work at hand “to the point where I made mistakes”. The other student stating that Labdog “took too much time and attention away from carrying out the practical”.

5.3.4 Discussion: Constructing a narrative of learning

Student responses provided valuable insights into the use of Labdog. Interestingly, few new findings were discovered, which did not emerge from previous qualitative data. This suggests that there is no urgent need to collect further data on the same subject (Fusch and Ness, 2015).

Labdog appears to have successfully promoted student focus on the underlying scientific concepts (Table 5.12). Thus, Labdog presents a viable medium to address cookbook chemistry (Domin, 1999b), as one student stated: “[Labdog] has provided an insight of chemistry. It has taken chemistry procedures out of something like cooking recipe, to what it actually needs to be like”.

The student-perceived benefits are attributed largely to formative questions and feedback (Table 5.15). However, students raised valid concerns around technical issues (Table 5.14) and the competing demand of Labdog on their focus in the lab. Although I addressed many technical problems, it is clear that some remain, and that students do not easily forget negative experiences.

Beyond the technical scope of Labdog, this research has raised concerns around the design of a LaRS-enhance laboratory. It is important to balance the pedagogical benefits with psychological concerns of students' cognitive resources (Paas and Ayres, 2014), and logistic constraints of time.

It is no new finding that that the complexities of the laboratory environment can hamper student learning. However, it is part of a new body of work, e.g. Galloway and Bretz (2016), examining such problems from the student perspective. Other novel research has attempted to integrate technology into the lab, to alleviate logistical constraints (Weibel, 2016; Van Dyke and Smith-Carpenter, 2017), but Labdog is novel in its explicit focus on pedagogical benefits. Students repeatedly reported Labdog as beneficial to their learning, namely though formative questions and feedback. Such findings have recurred throughout multiple formats of investigation, suggesting validity and reliability. However, before discussing their implications further, I wish to examine the learning process in more specific detail, in an attempt to move away solely from student-reported learning outcomes.

Chapter 6

Collecting Evidence of Learning in Labdog (2016/17)

Having demonstrated that Labdog was used for the 2016/17 academic year, and detailing the student-perceived impact on practice and learning, I wish to examine its impact on learning using a series of other data. Firstly, I used a pre-published quantitative instrument, the meaningful learning in the laboratory instrument (MLLI) to compare student-reported learning between SFY and UG students.

Secondly, I use a corpus of SFY responses to in-lab Labdog questions, and the associated educator feedback. I use this data to identify evidence of students' (mis)understanding during the lab, and better demonstrate how Labdog was used during the laboratory sessions. DBR encourages ongoing evaluation of the research context (Collins, 1992; Wang and Hannafin, 2005), and this represents an examination of Labdog outside of user-reported information.

Lastly, I examine a series of open-answer responses given by two of the three members of SFY laboratory teaching staff. This provides insight into the educator-perceived impact and role of Labdog, supporting all previous narratives.

6.1 Measuring Meaningful Learning in the Lab

I have previously (e.g Chapter 4 and Section 5.3) used custom evaluative instruments to quantify the experience of using Labdog. The complexity of learning in a real-world context (Barab and Squire, 2004) means the development of a valid and reliable instruments is an area of study in its own right, e.g. Krause et al. (2004). Although the use of *ad hoc.* tools has provided me with rapid and adequate insight, more rigorous instruments provide comparable, consistent, and previously-validated data.

I therefore decided to adopt a previously validated survey instrument: the Meaningful Learning in the Laboratory Instrument (MLLI). The MLLI was constructed to help instructors "design instructional materials to bridge the gap between instructor goals and student expectations" (Galloway and Bretz, 2015, p.1149), specifically in the laboratory environment. The original MLLI consisted of 31 items, phrased as statements, which were split between positive (14) and negative (16) wordings, to which students would agree using a slider-bar between 0-100. There was one indicator item, 16 cognitive, eight affective, and six cognitive/affective. These three components make up meaningful learning, in the authors' definition (Novak, 1998; Bretz, 2001), itself influenced by the constructivist paradigm (Novak, 1998).

To validate and evaluate the instrument, Galloway and Bretz (2015) worked with 614 American chemistry students (436 general, 178 organic chemistry) who completed the MLLI at the beginning and end of a semester. The authors conducted factor analysis and exploratory factor analysis (EFA; through Eigenvalues and pattern matrices). EFA was conducted on both pre- and post-test surveys to identify, and then remove, the items which may not belong to a particular factor. Furthermore the authors interviewed 13 students, in which the student would describe their interpretation of an item and justify their response.

Cronbach's alpha test was also applied, and returned a high (> 0.7) α for affective and cognitive items, but a lower score (< 0.7) for cognitive/affective. As Cronbach's α is a measure of "how correlated student responses are on a set of items" (p.1153) this suggests that students generally separate cognitive and affective factors, and therefore where the two are mixed (affective/cognitive) they hold less consistent ideas.

6.1.1 Methodology

The MLLI therefore offers a robust tool to collect student-reported learning along the constructivist paradigm Novak (1998). With the benefit of fair comparison, I decided to collected data from both SFY and first year UG cohorts in 2016/17. Data was collected at the end of semester one (December 2016) and two (April 2018).

Due to the greater number of UG students, the survey was distributed to UGs as an online pre-lab activity. A number of UG students completed the survey more than once, likely because of technical issues, where this happened I removed the attempt which took the shorter time, assuming that data they spent more time on considering would be more honest. SFY students completed the MLLI on paper during a compulsory workshop session at the same week it was given to UGs, and I digitised their responses. Both UG and SFY students gave their student IDs, which allowed cross-semester comparison. Though distribution in paper versus online medium may have some minor effect on students' responses, this difference can be largely discounted (Le Corff et al., 2017).

It is important to note a number of differences between this and the original authors' methodology:

1. I used the MLLI to examine only past, and not expected, learning in the laboratory. I believed the inherent differences in the entry requirements and prior experience between SFY and UG cohorts would give the cohorts drastically different expectations. Specifically, the diversity in SFY requirements would contrast the largely A-level history of the UG cohort.

2. I used 1-7 ordinal responses, and not the 0-100 integer sliders used by the original authors. I did so to keep a consistent data-type between all surveys in the research project, as it would be inconsistent to measure agreement with a statement as continual in one survey but ordinal in another. I used a likert-scale which contained two 5% intervals at either end ('0-5%' and '95-100%'), two 15% intervals ('5-20%' and '80-95%'), and three 20% intervals. I adapted the indicator item to account for this change.

3. A technical error with the first SFY questionnaire meant SFY students only responded to the first 28 (of 31) items MLLI. For equal comparison this was kept the same in the second semester.

6.1.2 Results

After removing duplicates and responses from students who failed the identifier question, 246 total responses were included in the analysis. Counts by cohort and semester are shown in Table 6.1. Responses to all questions were not statistically normally distributed ($P < 0.001$ Shapiro-Wilks).

Semester	Year		
	SFY	UG	
1	24	94	**118**
2	22	106	**128**
	46	**200**	**246**

Table 6.1: Number of responses made to the MLLI in 2016/17.

Comparing within cohorts

I firstly wished to examine the differences in the responses made by individual students across both instances of the survey. However, of the 32 unique SFY students who responded to either MLLI, only 13 ($\sim 28\%$) students completed both. Due to this small sample size, relatively and absolutely, I do not believe that a paired-sample analysis would provide valid or reliable data.

I therefore treated the SFY data as non-independent, and conducted a Fisher's test, on all response from both semesters. The only significant difference was found in the item "I worried about getting good data" ($P = 0.007$), no other tests returned a $P \leq 0.3$. This item has little pedagogical value, presenting a limited scope for discussion. As such, I continue to to discussing the difference between SFY and UG cohorts, where more fruitful data was present.

Comparing between cohorts

A Pearson's chi-squared measured differences in distribution of responses to individual questions between SFY and UG cohorts. This generated P-values for semester 1 and 2 data, which are presented in Table 6.2. Statistically significant values have been denoted with an asterisk (*). Items with a $P \leq 0.1$ are denoted with a dagger (†). I bring these less statistically different items into discussion because they suggest a noteworthy degree of difference between cohorts.

I also identified a number of items in the MLLI which were relevant to the focus of Labdog. These items are highlighted in blue in Table 6.2, and are discussed in the following section.

Question	P-value		
	Sem. 1	Sem. 2	% change
Learned chemistry that would be useful in my life	0.576	0.587	1.91
Worried about finishing on time	0.076†	0.245	222.37
Made decisions about what data to collect	0.592	0.684	15.54
Felt unsure about the purpose of the procedures	0.164	0.824	402.44
Experienced moments of insight	0.423	0.853	101.65
Was confused about how the instruments work	0.412	0.567	37.62
Learned critical thinking skills	0.772	0.327	-57.64
Was excited to do chemistry	0.702	0.12	-82.91
Was nervous about making mistakes	0.054†	0.545	909.26
Considered if my data makes sense	0.495	0.484	-2.22
Thought about what the molecules are doing	0.531	0.099†	-81.36
Felt organised	0.769	0.958	124.58
Developed confidence in the laboratory	0.32	0.745	24.58
Worried about getting good data	0	0.106	
Thought the procedures to be simple to do	0.711	0.569	-19.97
Was confused about the underlying concepts	0.056†	0.026*	-53.37
Got stuck but kept going	0.004*	0.724	18,000
Was nervous when handling chemicals	0.277	0.161	-41.88
Thought about chemistry I already knew	0.399	0.931	133.33
Worried about the quality of my data	0.54	0.349	-35.37
Was frustrated	0.42	0.08†	-80.95
Interpreted my data beyond only doing calculations	0.865	0.815	-5.78
Focused on procedures, not concepts	0.024*	0.302	1158.33
Used my observations to understand the behaviour of atoms and molecules	0.521	0.266	-48.97
Made mistakes and tried again	0.502	0.0447*	-91.1
Was intrigued by the instruments	0.011*	0.24	2081.82
Felt intimidated	0.293	0.402	37.20

Table 6.2: P-values (Pearson's chi) in difference in response distributions to MLLI items between UG and SFY cohorts, in semester one and two, and the percentage change between semesters.

6.1.2.1 Statistically notable differences

Responses to the following items were significantly differently distributed between cohorts. Distributions to each item are shown in Table 6.3.

1. (During this semester I ...) worried about getting good data (Semester 1, $P < 0.001$)
2. ..."got stuck" but kept going (Semester 1 $P = 0.004$)
3. ...focused on procedures, not concepts (Semester 1, $P = 0.024$)

4. ...was intrigued by the instruments (Semester 1, $P = 0.011$)
5. ...was confused about the underlying concepts (Semester 2, $P = 0.026$)
6. ...made mistakes and tried again (semester 2) (Semester 2, $P = 0.045$)

Getting good data. In semester 1, UG students reportedly worried more about getting 'good data'. This is likely due to the assessment of UG labs, which requires students to submit a thorough data analysis with their lab-report. Conversely, SFY students complete a skills-based portfolio in which they must demonstrate a competency and understanding of some practical skill(s), e.g. handling a pipette or using a fume cupboard. This latter approach is similar to the digital badging previously discussed (Hensiek et al., 2017), and does not focus so heavily on the accuracy or correctness of data. In semester two, the trend remains in UG students, but a larger proportion of SFY students begin to report being worried about the data.

'Got stuck' but kept going. UGs reported this feeling more often than SFY students. The wording of this item is unclear: it could be interpreted as asking how frequently students got stuck, or how often they felt they overcame adversity when it arose.

Differences in both content, and student-instructor ratio could contribute to this difference. UG labs feature more complex procedures and therefore students may be more likely to feel stuck. SFY labs contain between 20-30 students, and 3-4 members of demonstrating staff. Additionally, SFY demonstrating staff are the teaching staff for lectures and workshops, building consistency and rapport with students, as opposed to the common rotation of PGTAs in the UG lab.

Made mistakes. Where SFY students were more likely to agree with the statement, UG students submitted more wide-spread responses. This may suggest that SFY students are more persistent. It is possible that this Labdog caused students to recognise their mistakes more than UG students, through formative feedback. Additionally, it is possible that a greater demonstrator-student ratio, and demonstrator consistency in the SFY meant that mistakes were more often identified.

Question	Sem	Year	N	0-5%	5-20%	20-40%	40-60%	60-80%	80-95%	95-100%
Worried about getting good data	1	SFY	24	4.17	20.83	20.83	12.50	4.17	37.50	0.00
		UG	94	0.00	3.19	7.45	13.83	24.47	32.98	18.09
“Got stuck” but kept going	1	SFY	24	20.83	16.67	20.83	20.83	4.17	12.50	4.17
		UG	94	1.06	15.96	15.96	19.15	23.40	18.09	6.38
Was intrigued by the instruments	1	SFY	24	4.17	0.00	29.17	20.83	20.83	20.83	4.17
		UG	94	1.06	9.57	6.38	11.70	36.17	20.21	14.89
Made mistakes and tried again	2	SFY	22	0.00	0.00	0.00	22.73	36.36	31.82	9.09
		UG	104	0.96	4.81	14.42	15.38	26.92	25.96	11.54

Table 6.3: Percentage frequency distribution of responses to each statistically significant question in the MLLI, between UG and SFY students.

Intrigued by the instruments. UGs reported being intrigued by the instruments they use. This is largely outside the control or influence of Labdog, but is rather a function of facilities of the UG laboratory, which represent a wider, and more sophisticated selection than the SFY have access to.

6.1.2.2 Labdog-relevant items

Certain MLLI questions were particularly relevant to Labdog's pedagogical intentions. These questions are highlighted blue in Table 6.2. Of these items, two had a statistically significant difference between SFY and UG cohorts. These were: i) students' focus on procedures not concepts, shown in Figure 6.1, and ii) reported confusion on the underlying concepts, visualised in Figure 6.2.

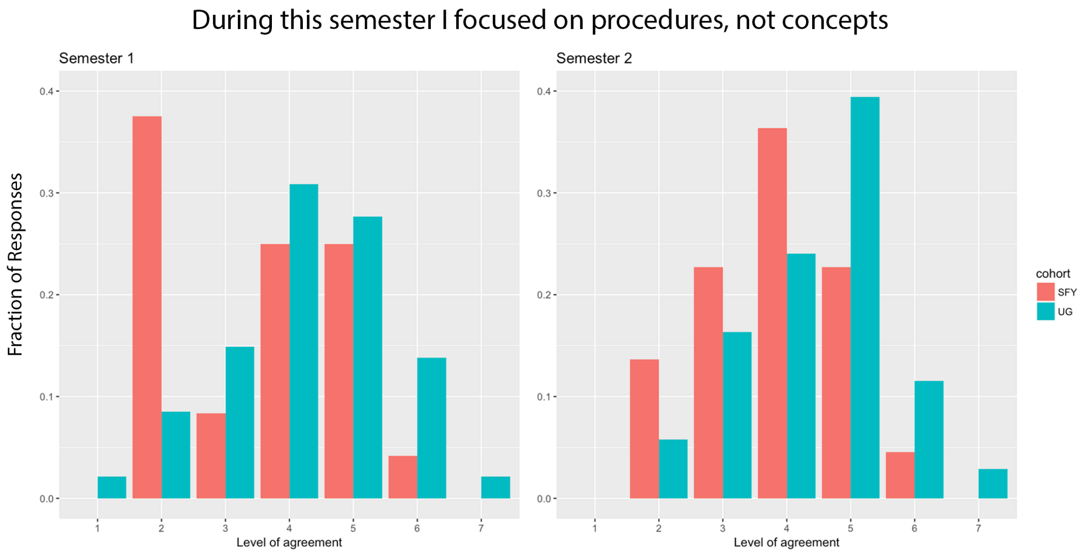


Figure 6.1: Student responses to the MLLI item "during this semester I focused on procedures not concepts" from semester 1 and 2, split by cohort.

Focus on procedures or concepts. Labdog can be used to draw students' focus to concepts over procedures. Although the teaching laboratory can be used to improve both practical competence and scientific understanding (Taber, 2015), a combination of poor conceptual understanding and unclear instruction often sees students often focus on process over concepts (Abrahams and Millar, 2008; Galloway and Bretz, 2016). Given that SFY students used Labdog, we would hope to see a greater focus on concepts, compared to UG students.

The data, visualised in Figure 6.1, supports this notion across both semesters, though there is only a statistically significant difference in semester 1. In semester 1, 74.4% of UG agreed $\geq 40\%$ to focusing on procedures not concepts, compared to 53% of SFY students, in semester 2 this difference was still present but less stark: 77.9% of UG compared to 63.9% of SFY.

This idea that SFY students thought more about the submicroscopic element of the laboratory (Johnstone, 1982) is supported in two other questions, where students reported they: "thought about what the molecules are doing", and the other asking if they "used [their] observations to understand the behaviour of atoms and molecules". Using the percentage of students who agreed $\geq 40\%$ as a measure of agreement, SFY students reported thinking more about what the molecules were doing ($S_1 = 55.9\%; S_2 = 86.3\%$), compared to an initially higher but slightly decreasing percentage of UG students ($S_1 = 79.8\%; S_2 = 77\%$).

There is a similar upward trend in students interpreting observations at the molecular level. SFY students showed a greater gain between semesters ($S_1 = 77.47; S_2 = 86.5\%$), compared to lower and consistent UG levels ($S_1 = 76.7; S_2 = 78.74\%$). It is reasonable to attribute at least a portion of these differences to students' use of Labdog. This claim is strengthened by the messages which emerged from SFY students during interviews (Section 5.3.2).

It is important to acknowledge differences in instructional factors which could also contribute to these differences. Specifically, I think it important to acknowledge that UGs were required to handle more complex concepts, procedures, equipment, and data. Despite this, UGs reported feeling their procedures were more simple. By contrast, more SFY students found their procedures more simple in semester one, but less simple in semester two.

Related to this, SFY students reported developing more and more consistent confident in the lab. Despite finding the procedures less simple, there is relatively little change in UG students' reported focus on the molecular aspects. Contrastingly, SFY students report a simultaneous increase in perceived difficulty and focus on practical procedures, as well as developing their confidence. Some of this difference can likely be explained by the use of limited cognitive resources by using equipment and completing procedures (Paas et al., 2010). However, it appears that SFY students were able to develop confidence and adopt a molecular focus while handling reportedly similar difficulties in the lab. This suggests that Labdog may help students better utilise cognitive resources for conceptual gain. Alternatively, it could be that Labdog has some impact on student's perceived confidence and mastery of practical procedures - a previously unconsidered idea.

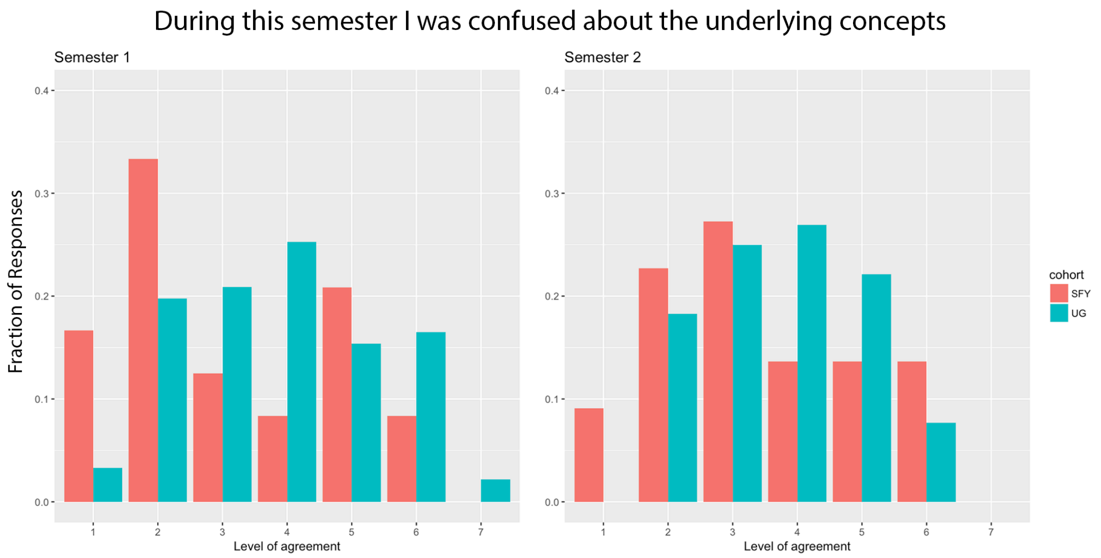


Figure 6.2: Student responses to the MLLI item "during this semester I was confused about the underlying concepts", from semester 1 and 2, split by cohort.

Student-reported confusion. The MLLI asks students to agree with the statement "[I was] confused about the underlying concepts". This gives some indication of both students' knowledge, and their ability to handle and process new or previously-covered chemical concepts which arise during a laboratory session.

One could reasonably propose that students who use Labdog could become either more or less confused than if they had not. Students with poor understanding often have high confidence in their knowledge and capabilities (Kruger and Dunning, 1999). The formative feedback associated with the use of Labdog could therefore lower the confidence of previously confident, but low ability, students. Contrastingly, competent students could have their ability recognised and validated, or reduce confusion over the concepts (Black et al., 2003) - lowering confusion

Across both semesters (Figure 6.2) SFY students are less likely to report feeling confused about the concepts. UG students move from a relatively homogeneous to a more unimodal distribution, with agreement peaking at 40-60% in semester 2. SFY responses, in comparison, move from a bimodal distribution in semester 1, with peaks at 5-20% and 60-80%; to a more unimodal distribution in semester 2, peaking at 20-40%.

Again, it becomes essential to consider the wider context of UG versus SFY material. UG students are handling more complex concepts and equipment, and therefore one may expect to see greater reported confusion. However, good instructional design would have the difficulty calibrated with students' ability, which in turn is related to better learning practices (Engeser and Rheinberg, 2008; Undorf and Erdfelder, 2013). Anecdotally and historically, SFY students have struggled with chemistry, both the material and the laboratory. Additionally, only a minority (perhaps 10%) of SFY students continue to study chemistry at HE. Therefore a direct comparison with chemistry UGs needs to be considered carefully, in terms of the ability, self-efficacy, and motivations of students in chemistry.

Despite these differences, the results show that SFY students report being less confused about the chemical concepts. Further research is needed to attribute this to Labdog in specific, however it plausible that Labdog is leading students to be less confused about the chemical concepts they are handling in the laboratory. The study of confidence (Kruger and Dunning, 1999) or self-efficacy (Bembenutty, 2009) are idiosyncratic and complex areas to study, however there is clear value in understanding how, when, and if Labdog prevents or reduces student confusion.

6.1.3 Concluding remarks

The MLLI is a pre-existing, validated instrument which quantifies student-report learning in the laboratory environment. The MLLI was completed by both SFY and UG cohorts at the end of each semester in 2016/17. The analysis examined the difference between UG and SFY cohorts, as low repeating sample size (n=13) prohibited paired-sample comparison within the SFY.

Within the survey, UGs reported a greater interest in the equipment and the procedures themselves. Whereas SFY students demonstrated a greater focus on the conceptual component of the laboratory. Notably, SFY students report comparatively less confusion over scientific concepts in the lab, and a greater focus on 'what the molecules are doing. SFY students reported developing a higher and more consistent confidence in the laboratory. Whereas UG students reported a less high, but more drastically increasing confidence between semesters.

This can be paired with student-perceived complexity of the procedures, where SFY thought the procedures became less simple between the two semesters, but UGs reported finding them more simple. This raises interesting questions about the impact of

Labdog in increasingly complex environments and problems. Specifically, how does it affect students' competency, confidence, and self-efficacy. Such factors affect learning (Prat-Sala and Redford, 2010) and metacognition (Flavell, 1979; Cook et al., 2013) and how students form, and act upon, opinions of their confidence are multifaceted (Dinsmore and Parkinson, 2013). Furthermore, examining how these relate to students' ability and confidence (Kruger and Dunning, 1999) could identify which students Labdog benefits the most.

Attributing the above differences to Labdog is difficult, given contrasts in course content, laboratory equipment, assessment, and staffing between SFY and UG laboratories. However, the data strongly supports the themes which emerged in the previous chapter: that Labdog helps students focus on the conceptual component of the teaching laboratory. What's more - it may reduce confusion and increase confidence.

6.2 Responses to open answer questions

Students repeatedly reported that formative questions and feedback are what made Labdog most beneficial. In this section I examine how these questions were used by educators and students, commenting on the extent of learning evidenced.

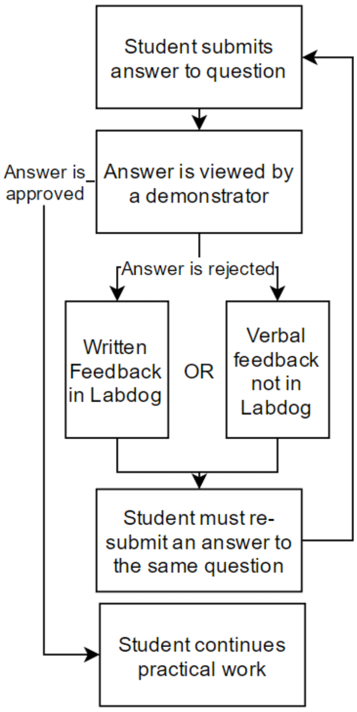


Figure 6.3: Labdog workflow for submitting a response to a question, and providing feedback.

The process of answering a question during a lab session in Labdog is represented graphically in Figure 6.3. Students submit an answer to a question; responses are then presented to educators through a real-time interface; educators then approved or rejected the response. If rejected, the demonstrator provides either written feedback in Labdog, or in-person verbal feedback which they would then note in Labdog. If a response is rejected, students are required to re-submit a response to the same question, and cannot progress through the practical activity until they have done so.

6.2.1 This over-use of questioning in semester one

During seven practical activities over five sessions during semester one, 36 questions were posed to students - averaging 5.1 questions per practical. In total, 624 responses were submitted, averaging 17.3 per question. In total, 59 (9.4%) of the responses were resubmission to prior feedback. As a crude measure of students' engagement with the questions, I calculated the average number of words in a response to each, which is visualised and sorted in Figure 6.4.

During the semester I focused my efforts on technical support and improvements - primarily the real-time features necessary for educators to take advantage of students' responses. I therefore conducted the bulk of semester one's analysis in retrospect, which prevented it form impacting practice.

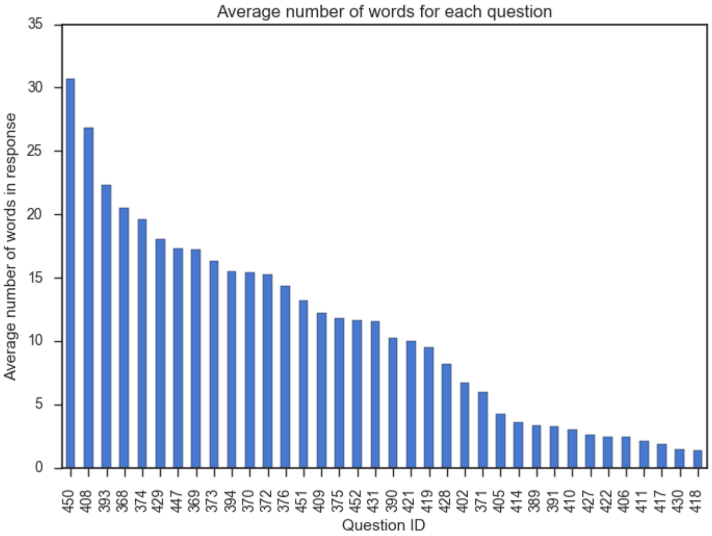


Figure 6.4: Average number of words in responses to each question posed in semester one (N=36), sorted in descending order.

It quickly became clear that questions were over used in semester one. No pre-existing research has been done using a LaRS technology, and so no guidelines exist to inform what makes a ‘good’ question. In their eagerness to take advantage of the benefits of formative feedback, educators posed too many questions to students. Asking students too many questions, obviously put too much demand on the students, as was noted during pilot (Chapter 4) and SFY evaluations (Section 5.3).

The incompatibility between LaRS-usage and the practical work were highlighted in practical activities 3a and 5 (Table 5.2, page 132). These experiences inspired a reflective online post[1].

It is clear that the length of a responses varied widely between questions. I was largely uninvolved in the generation of questions - the subject knowledge and teaching experience of the teaching staff far outweighed my own knowledge. The questions which emerged did so naturally from practising and experienced educators, and a thematic analysis of them lead me to identify six distinct types of questions (given below). The number of questions in each of these categories in semester one is given in Table 6.4.

- **Observation**: Explain an observation using the related chemical concept(s);
- **Method**: Explain or justify a methodological choice;
- **Definition**: Definition a related technical or scientific term;
- **Reaction:** Give mechanistic or empirical details of a chemical reaction;
- **Calculation**: Report a numeric result from a chemical reaction;
- **Other**: Anything which did not fit into the above.

There was a significant (one-way ANOVA, $P < 0.001$) difference in the length of responses within each question type, visualised in Figure 6.5. Students gave longer response to questions about an observation or method, but provide shorter responses for calculation, definition, or reaction questions.

Calculation questions were the most common. These would often stop students mid-practical and have them workout some value or equation before continuing. Though these are important chemical and mathematical concepts, I would argue that providing a definition or calculation are not providing “purposeful checks on [students’] understanding” (Galloway et al., 2015, p.153), instead promoting recipe-following or simple recall (Krathwohl, 2002). This does not represent moving away from the historically shallow use of the laboratory (Abraham, 2011).

---

[1] `http://edtechandchem.ghost.io/quick-pivots-in-labdog/`

Type	Count	Example
Calculation	12	What was the mass of oxygen gained during the formation of magnesium oxide?
Method	8	If you didn’t rinse the beaker, what effect would this have on the concentration of your standard solution?
Observation	7	Why does some solid remain in the flask?
Definition	2	What does ’in excess’ mean?
Reaction	2	The reaction of copper (ii) oxide and sulphuric acid gives a salt and one other product name the two products of this reaction
Other	5	When you have completed the evaluation survey [Link to survey]. Answer this question, remember that we can double-check your response

Table 6.4: The number of questions posed to SFY students in Labdog during Semester one of 2016/17, divided by category and with examples.

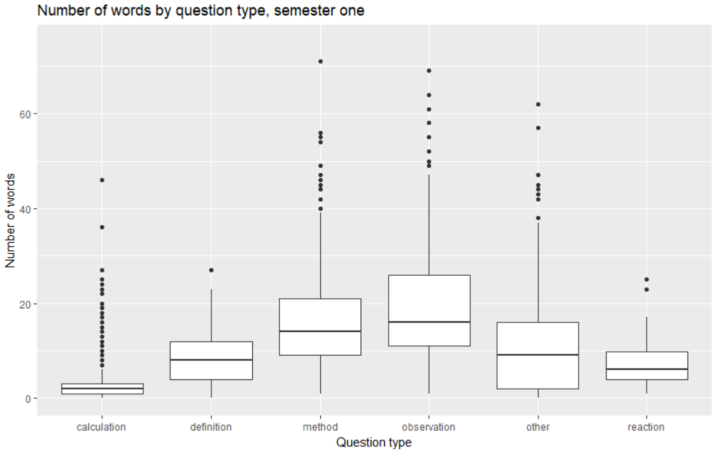


Figure 6.5: Boxplot distribution of the number of words in a response to each question type given in Labdog during semester one.

Through ongoing reflection and discussion, the teaching team began to refine the instructional design principles for questions. By the end of semester one, we believed that every question asked during a practical should be only be answerable during the laboratory itself (Wilson and Read, 2017). This involved asking fewer questions overall, specifically less calculation, definition, and reaction questions. Such questions could still be posed, but in pre- or post-lab environment.

6.2.2 Examining students’ responses in semester two

In total 12 questions were asked to students across six practical activities in semester two, shown in Table 6.5. A number of questions could not be clearly divided between the ‘observation’ and ‘reaction’ category, I therefore created a hybrid (observation/reaction) category. Such questions are characterised by asking students about some observation, and then explicitly prompting some relevant theory.

An average of 2 questions were asked per practical, compared to 5.1 in semester one. A total of 360 responses were made, which were significantly longer, in terms of number of words, (one-way ANOVA $P < 0.001$) in semester two ($\bar{x} = 18.12$), compared to semester one ($\bar{x} = 12.21$). 26 (7.2%) responses were re-submissions to written feedback in Labdog. Responses to the definition question were much longer, because it asks students to list the identity of five compounds - necessitating more words.

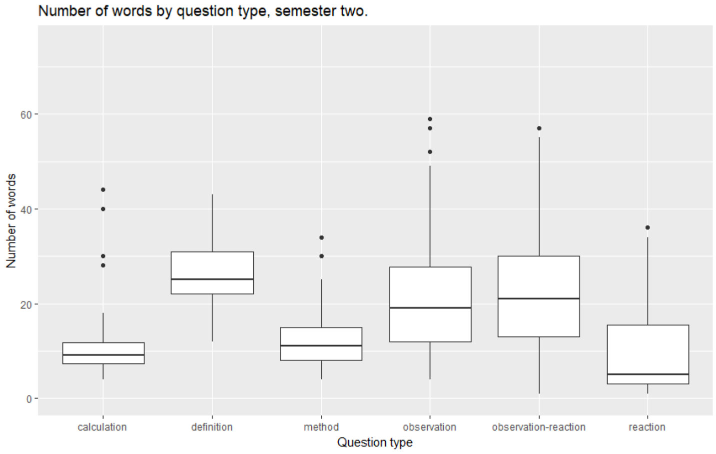


Figure 6.6: Boxplot distribution of the number of words in a response to each question type given in Labdog during semester two.

There was a significant (one-way ANOVA, $P < 0.001$) difference in the length of responses given to each type of question in semester two, visualised in Figure 6.6.

6.2.3 Thematic analysis of observation/reaction responses

The increased focus of questions to linking the macro and submicroscopic suggest that students’ responses are more likely to discuss relevant scientific principles. To explore this idea, I conducted an exploratory qualitative analysis of students’ responses to the observation/reaction questions posed in semester two. I take these questions as a case-study: if students are demonstrating poor engagement with Labdog when the questions being asked are most relevant and valuable, then the student-use of Labdog must be appropriately questioned.

Type	**Count**	**Example**
Observation/Reaction	3	What do you notice about the pH values of the solutions of ethanoic acid compared with those of HCl with the same concentration? Explain this observation using appropriate theory covered in last week's lectures.
Observation	3	Explain the observations you made in step 4 i.e. when concentrated ammonia was added to the copper(II) sulfate solution in the fumehood.
Reaction	3	The formation of an ester from a carboxylic acid and an ester is an equilibrium reaction. In these reactions, you used an excess of the alcohol. With reference to appropriate theory relating to equilibria, explain the effect of using excess alcohol on the yield of the ester.
Calculation	1	After testing the unknowns, identify which is a ketone and which is an aldehyde and give your answer below.
Method	1	What is the purpose of the sulfuric acid? (hint: it is regenerated at the end of the reaction)
Definition	1	What's the difference between the equivalence point and the endpoint in a titration? Mark both on your graph and explain the difference below.

Table 6.5: Number of questions of each type asked in Labdog during semester 2 to SFY students.

This section does not provide a comparison of students' responses between semester one and two, given the poor, often overwhelming, design of questions in semester one. Future cohorts with more consistent design and experience could allow such comparison.

6.2.3.1 Methodology

All responses to all (three) observation/reaction questions were downloaded from Labdog. The questions were:

- **Question 1 & 2:** The chemical test performed here involves a redox reaction. Which species is gaining electrons during the reaction, and what is formed by this process? What is the observable change you see as a result of this process as the outcome of a positive test? (Practicals 7b: Silver mirror test; and 7c: Fehlings Test)

- **Question 3:** What do you notice about the pH values of the solutions of ethanoic acid compared with those of HCl with the same concentration? Explain this observation using appropriate theory covered in last week's lectures (Practical 8a: Strong and weak acids)

I adopted a thematic analysis approach to students' responses (Braun and Clarke, 2006). This allowed an open-ended exploration of the data. This low-level interpretation of data does not necessitate the development of more intricate theories or detail, as seen in Grounded Theory (Vaismoradi et al., 2013). However, the process itself remains similar: systematically and iteratively (re)reading and (re)coding the information to generate understanding (Vaismoradi et al., 2016).

6.2.3.2 Results: Question 1.

In practical 7b students need to distinguish between an aldehyde and a ketone. In this practical, students add Tollen's reagent, which contains diamminesilver(I), to their solution. Tollen's reagent is an oxidising agent, and when added to the students' aldehyde, the Ag^+ ions in $AgNO_3$ are reduced to produce metallic Ag. This species is not soluble in the solution and so precipitates out, forming a thin layer of silver which coats the test-tube and creates a 'silver mirror' effect which the students observe as a visual cue that the reaction has taken place.

Twenty eight responses were made to this question: one was a re-submission, and two were misplaced as answering question 2 but included in this dataset. Nearly all (26) responses correctly identified that the silver ions were being reduced, and many (19) defined this as the gain of electrons. Far fewer (6) students stated that this related to the oxidation, i.e. gaining of electrons, of the aldehyde.

While many (19) students recorded that a precipitate formed, a smaller subset (14) gave a causal link between the reduction of Ag^+ ions and the formation of the silver mirror. An example of a student not making a causal attribution is: "Ag^+ is gaining electrons, is reduced. Ag precipitates out of the solution". Though this student has made two correct statements, it is unclear if the student understands the connection or chemical pathway. Likewise, the language in some responses suggests that students hold misconceptions, for example one student explained the precipitate as: "Ag^+ gains an electron and becomes Ag. As a result, silver became visible and silver mirror formed around the test tub". Here, the use of the word 'visible', as opposed to precipitated, not-dissolved, or solid, suggets perhaps a misconception or misunderstanding.

Response	Format	Feedback
Silver ions is being reduced forming a silver mirror	Written	What is the charge on the silver ions? What are they reduced to?
Tollens' reagent is the oxidising agent so is being reduced and gaining an electron. The formation of silver mirror is formed in positive test	Written	What specific species in the Tollen's reagent is reduced? Use your observation the help you.
The aldehyde is oxidised (oxidation is the loss of electrons) and forms a silver mirror unlike the ketone which can not be oxidised	Written	What species is reduced? Use you observation to help you.
The silver again because from silver plus getting silver and we have a product of silver mirror	Written	What about electrons?
Silver nitrate is being reduced. It then forms a precipitate which should be silver...	Verbal	The silver + ions are gaining electrons to form Ag atoms. The silver atoms then form around the test tube.
The aldehyde was oxidised and the silver was reduced. The ketone can't be further oxidised but the aldehyde can.	Written	This is nearly correct but you need to be more specific when you say "silver". What state is the silver in?

Table 6.6: Student responses and feedback to a question asking students to observe and explain the 'silver mirror test', when distinguishing ketones and aldehydes

Six responses (21.4%) were rejected by demonstrators as in need of improvement.[2] The feedback, and their associated answers, are shown in in Table 6.6. All of the educator-authored feedback is posed as questions for students to consider, in agreement with the idea of Socratic dialogue. The feedback consistently asks students to specifically state the species, or formal charge, on both the reactant (Ag^+) and product (Ag) ions. When the whole dataset of students' responses were examined, however, only 16 responses (57.1%) detail the species of the reactant (13, 46.4%), product (12, 42.9%), or both (9, 32.1%) Ag species. Depending on the stringency of the educators' requirements, i.e. do students need to mention both reactant *and* product species, between 37.5-66.6% of inappropriate student answers were not picked up through Labdog. This suggests a need for better communication between teaching staff, and a clearly communicated set of expectations or guidelines for answers to questions - which could form part of educator training for the use of Labdog.

---

[2] Feedback noted as given 'verbally' has the student summarise the feedback they were given, whereas 'written' feedback is authored by the educator. In most instances, took the opportunity to re-submit a corrected answer, and not summarise the feedback.

6.2.3.3 Results: Question 2.

In practical 7c students likewise attempt to distinguish between an aldehyde and ketone, this time using Fehling's solution, which contains Cu^{2+} ions. When added to an aldehyde the copper ion is reduced to Cu^{+}, which forms a brick red precipitate of copper (I) oxide.

A total of 30 responses were given to this question. Of these, 14 (30%) recorded the formation of a red precipitate, and one student reporting a silver precipitate. Similar to the previous question, the majority of students identified that copper was being reduced (24), with many (17) identifying the Cu^{2+} reactant or stating the Cu^{+} product (16). A minority of students (7) specified the aldehyde as the source of the electrons in the reduction.

No students give a chemical/physical causal relationship between oxidation and the formation of the precipitate. The fact that no students link their observation to the chemical process undermines the nature of the question. The fact that students mention both of these elements suggests that students possess the understanding, however. Therefore, the potential cause for concern is likely in the wording of the question, which asks: "which species is gaining electrons during the reaction, and what is formed by this process? What is the observable change you see as a result of this process as the outcome of a positive test?". The question asks students two separate points, almost as two separate sub-questions: to detail the reaction, and then their observation. Perhaps a relatively simple change in wording to ask students to "explain how your observation relates to the chemical reaction taking place" could elicit such understanding.

Despite this criticism, students generally demonstrated a good level of understanding. As such, nine responses (30%) were rejected by educators, shown in Table 6.7. Similar to before, educators are asking students to detail both the product and the reactant. Several responses describe the reaction qualitatively, e.g. 'Cu^{2+} gain electrons from the aldehyde'. One piece of feedback is given to a student who appears to have two reactions confused, as they report observing a silver precipitate. The last two items of feedback ask students to detail more of the chemical reaction. The root cause of most of this feedback is simply a lack of specificity: not clarifying which molecules are involved in the reaction, which does not necessarily indicate a low level of understanding. This in itself, is a criticism, as the questions are intended to identify and demonstrate such (mis)understandings.

Response	**Format**	**Feedback**
Cu^{2+} gains electrons from the aldehyde	Written	It does! What is formed?
The copper ions from the Fehling's solution is being reduced, it is the oxidising agent so is gaining electrons. The positive test will turn a brown red colour	Written	What is the charge on the copper before and after the reaction?
Cu^{2+} gain electrons from the aldehyde	Written	It does! What is formed?
CuO?	Written	Cu^{2+}
The aldehyde is oxidised to carboxylic acid by losing electrons, the copper ions from the copper sulphate solution are reduced as they gain electrons	Written	What is the charge on the Cu ions before and after the reaction ?
Cu3+ ions are being reduced to Cu^{2+} ions I.e these are gaining electrons and the aldehyde is being oxidised I.e losing electrons	Verbal	
the copper ions is gaining electrons to make copper ions with a plus 1 charge. As a result, a silver mirror is formed.	Written	Are you sure?
Copper changes colour from blue to brick red (ours went green)	Written	Be specific! They are copper ions. What is the charge on the ions?
The sulphate is the spectator ion. The species which is gaining electrons is the aldehyde and in this case the copper ion is gaining one electrons to create a plus one charge. The copper then turns brick red and does not form a precipitate because copper still has a plus one charge	Verbal	The copper 2+ ions are gaining electrons. They then gain one electron to form cu+ ions. This can then react with one oxygen to form cu2O. This forms a brick red precipitate.

Table 6.7: Student responses and feedback to a question asking students to observe and explain the formation of a red precipitate when distinguishing ketones and aldehydes

When examining the questions, the majority of students noted that Cu^{2+} was being reduced (18, 60%) to form Cu^{+} (17, 56.6%), with 16 (53.3%) students giving details of both species. This analysis suggests that educators rejected between 43.75-50% of responses which failed to meet the standards they had set for other students. Additionally, three student (10%) gave the wrong species - only one of which was pulled up by the educators.

6.2.3.4 Results: Question 3.

In this question, students are asked to explain the relative difference in the acidity of ethanoic (higher pH) and hydrochloric (lower pH) acids. Students use the pH of both compounds at the same concentration, a measure of the concentration of H^+ ions in the solution. A stronger acid is characterised by a higher concentration of H^+ ions, i.e. a lower pH. The presence of H^+ in the solution suggests a more stable conjugate base, another way of considering the strength of an acid. In this question, students are being asked to discuss the relationship between pH, acidity, H^+ concentration, and dissociation.

37 responses were given to this question, of which 34 commented on the relative pH of both ethanoic and hydrochloric acid. The phrasing of these statements varied: some students (15) stated than ethanoic acid had a higher pH, and a similar number (10) stated that HCl had a lower pH. Six students gave comparative qualitative statements, e.g. that HCl was a stronger acid than ethanoic acid, but did not mention pH values. Two students defined ethanoic acid as more basic than HCl. Only one student incorrectly identified the stronger/weaker acid.

Many (33) students made some attempt to explain the difference in the strength of the acid. Approximately half (18) made the statement that a lower pH represents a stronger acid (or vice-versa). While many other answers eluded to this, several did not make the statement explicitly. A greater number (25) explained the differences in pH values due to the tendency or likelihood of the acid to dissociate, with many (18) linking acidity and the greater presence of H^+ ions.

Nine (27%) responses were initially rejected, though only eight are shown in Table 6.8 as the last was incorrectly rejected. The feedback suggests that educators are not satisfied with the level of detail students' explanations. All seven pieces of written feedback ask students to discuss acid strength as a result of, or in relation to, the dissociation of the compound and/or the concentration of H^+ ions. This continues in verbal feedback, where students have still not linked acidity to dissociation. Three of the responses are rejected, at least partly, because students do not report having kept the concentration of their acids the same.

Response	**Format**	**Feedback**
The pH of ethanoic acid has a smaller difference in comparison to the pH of HCl which changed quite drastically. This is because it is a weaker acid compared to hydrochloric acid	Written	You could also compare the pH values of the two acids at the same concentration.
Ethan pic acid pH values are much higher as they are weaker acids than HCl regardless of concentration	Written	What does the strength of the acid mean for the concentration of hydrogen ions in solution?
The pH value increased as the value of concentration decreased. The pH of ethanoic has higher values than the HCl pH. The hydrochloric acid is more dissociatiate than ethanoic pH as it has more protons	Written	Hydrochloric acid is a strong so dissociates fully. Ethanoic acid doesn't fully dissociate. At the same concentrations, ethanoic acid has a lower concentration of protons and a higher pH value.
HCl is a strong acid compared to ethanoic acid which is weak, therefore the pH of HCl will be much more lower, acidic compared to ethanoic acid	Verbal	[None reported in Labdog]
Ethanoic acid is a weaker acid.this means that the pH of ethanoic acid will be higher than that of HCl, which is a strong acid and, will have a pH closer to 1	Verbal	This is because the ethanoic acid will not disocciate and will not create many H^+ ions. Whilst HCl is a strong acid and will discocciate completely, creating more H^+ ions and producing a lower PH.
The pH values are greater in ethanoic acid compared to those of hydrochloride acid, when at the same concentration. HCl is a stronger acid	Written	Can you explain this in terms other concentration of prints in solution?
The HCl is more acidic (lower pH values)	Written	Can you explain this with the theory of strong/weak acids.
HCl is completly dissociated but CH3COOH is just partially dissociated. So there are less H^+ in ethanoic acid so the pH is less	Written	A lower pH means there are more protons in solution, so more acidic.
The pH value of HCL and ethanoic acid decreases as concentration increases. This is because the concentration of H^+ions decreases as the solutions are diluted, making the pH less acidic as the concentration increases. HCl is a strong acid, so it highly dissociates while ethanoic acid is a weak acid and only partially dissociates	Written	Compare the pH values of the two acids at the same concentration.

Table 6.8: Student responses and feedback to a question asking students to observe and explain the difference in pH between ethanoic and hydrochloric acids.

Upon examining all students' responses to this question, 12 did not mention dissociation - suggesting that 25% of incorrect answers were not identified by educators in Labdog. This is a notably lower percentage than the previous two questions. Additionally, two further responses suggest that students were changing the concentration of the acids - where educators may therefore have missed a small number, but notable percentage (40%) of incorrect responses.

6.2.4 Concluding remarks

The type and number of questions posed to students during the laboratory changed drastically over the 2016/17 academic year. During semester one there was a greater focus on students performing calculations and defining terms. This use of questions did not agree with the underlying pedagogical principles employed in the design of Labdog, i.e. having students think about the actions of molecules (Johnstone, 1982; Galloway and Bretz, 2016). To this extent, Labdog was not necessarily addressing the shortcomings identified in prior literature (Abraham, 2011).

As a result students were asked fewer, and more conceptually focused questions in semester two. To assess the quality of students' responses, and determine if students engaged meaningfully in Labdog, I conducted an exploratory analysis of three archetypal questions given in semester two. Each of these questions asked students to link some observation with the relative scientific theory.

The evidence strongly suggests that students engaged with the questioning processes. It is important to learn from those students who demonstrated misconceptions, or who did not provide adequate detail. In the examples discussed, students often did not give detail of the ionic species involved in a reduction reaction, or make clear links between acidity, H^+ concentration, and the stability of a conjugate base. Examination of the responses also revealed that educators may have been picking up only approximately half of responses which lacked such detail. This suggests that there is a need for more careful or consistent use of feedback in Labdog from educators.

The investigation also suggests that students are engaging with chemical concepts through Labdog. This offers further, non student-reported, data that Labdog provided pedagogical benefits to students. This analysis suggests that at least equal weighting should be given to misunderstanding, and to students' desires to 'get on' with the practical work. Such students may demonstrate rote memorisation over the development of an understanding in the attempt to finish the practical (Bretz et al., 2013).

More work is needed to better identify and understand this idea, and also how it can be addressed. This investigation suggests that future improvements need to combine instructional design principles, to inform the content and placement of questions, and the actions of educators when providing formative feedback in the lab. Further future work could also investigate students' responses to question in Labdog to develop a more detailed and context-specific understanding of students' understanding.

6.3 Labdog from the educator perspective

The entirety of the data discussed in this thesis, up until this point, has examined Labdog from the student perspective. However I have repeatedly spoken about the importance of simplifying and empowering the educator workflow, in order to promote the adoption of Labdog from an educator perspective.

The limited time and resources available meant that this issue could not be integrated formally into the research questions. This should therefore be considered as a vignette of the educator experience and perspective of using Labdog. Despite brevity, I believe this to provide another, valuable, perspective into the use of Labdog.

6.3.1 Methodology

In 2016/17 there were three members of staff demonstrating in the chemistry practicals for the SFY: DR, the principle supervisor on this research; SB, a post-graduate teaching assistant (PGTA) who conducted the SFY interviews, SB is also a Ph.D. student in chemical education research; and Jen Barber (JB), a PGTA and Ph.D. student in physical chemistry. JB was not asked to participate due to extenuating personal and academic factors.

I developed a five-item, open-answer questionnaire for SFY laboratory educators. After the 2016/17 academic year had ended, I provided each demonstrator the questions in an electronic template, asking them to fill in their answers. I saw this as favourable to spoken interviews due to the intensive requirements of design and analysis of such rich qualitative data, and the wider scope of the research.

I asked demonstrators to detail the impact of using Labdog on each of the following aspects, in relation to the 2016/17 academic year:

- Students' learning in the laboratory;

- Students' overall experience of laboratory work;
- Their awareness of students' knowledge both before and during the laboratory;
- The type of activities they performed during the laboratory, as educators;
- Their overall workflow surrounding a laboratory session.

I conducted a thematic analysis (Braun and Clarke, 2006) on written responses to identify themes. However the small sample size, and inherently less detailed data, made the analysis a much briefer process than in any of the previous qualitative data methodologies described.

6.3.2 Results

In the following section I discuss each of the five points of the survey, drawing on quotes from SB and DR to illustrate points where appropriate.

Students' learning

Both DR and SB believed that Labdog was beneficial to students' learning. Their reasons echo those raised previously in this thesis: that both questions and the related formative feedback benefited students' learning. As SB states:

> "Following the introduction of Labdog and the questions that students are required to answer to progress, the students are having to think a lot more about what exactly it is they are doing and, more importantly, why they are doing it. Even if the students answer these questions incorrectly, the demonstrators are immediately able to see where students have misconceptions, and can then go to those students and discuss why their answer was incorrect. Hence, the impact on students' learning has been immense; not only are the students gaining the practical skills they need to be successful, they are also developing the understanding underpinning those techniques and the chemical processes occurring."

The benefits extend across both pre- and during-lab Labdog activities. In the context of pre-labs, the use of Labdog meant that "students were thinking more about what kind of things they would have to do in the lab... meaning that they could either look it up before the session or come to the session with questions for the demonstrators relating to the content, both indicating a positive impact on the students' learning and thought processes" (SB).

DR noted "it's unlikely that all students would have engaged with such thought processes without prompting", suggesting at the unique value of Labdog. To achieve this, DR states that it's important to "ask the right questions at the right time" - echoing the sentiments of the previous chapter. Although DR acknowledges the need for more rigorous evidence, "I feel it is likely that this leads to a significant positive impact on learning from laboratory work".

Students' overall experience

Both DR and SB mentioned the early technical problems with Labdog. They report some issues as ongoing, however, with DR stating that "some students resist engaging with Labdog, seeing it as a distraction from the procedure they are trying to undertake", also noting "technical issues and slight clunkiness in the interface, which [have] been greatly refined in advance of the current year". Additionally, SB justifies students' scepticism because Labdog "brought them more work in the lab, in the form of questions. In addition, sometimes the technology would fail, and as such the students did not want to rely on a tool that they could not necessarily trust to always work".

However, prolonged use, and the improvements in both student-perceived benefits and educator practices, eventually lead to a smoother, and more beneficial student experience: "experience of 17/18 indicates that good instructional design can lead to Labdog integrating seamlessly into the laboratory workflow for students, leading to a positive impression of its impact" (DR), and "with time, the reliability of Labdog improved greatly, and so too did students' experience of it. It was clear" (SB).

Both educators wished to mention the importance of formative feedback to improving the experience. SB reported that "over time the students felt.. the additional questions... [were] of greater benefit than not having them". Likewise, DR reported the related feedback as beneficial to students: "Perhaps the biggest single impact is in support for the provision of virtually instantaneous feedback... students seem to be incredibly grateful... in an era where feedback is becoming ever more important, this is definitely a significant positive impact of the use of Labdog."

Awareness of students' knowledge

Both educators reported that Labdog was successful in improving their awareness of students' (mis)understanding. I have commented before on this from a student perspective, and so will not labor the point. SB summarises the impact well:

> "Students' responses to questions for both the pre-lab and in-lab were accessible to all demonstrators... This proved very positive for us as demonstrators, as it allowed for us to target our approach a lot more; rather than wandering around the lab asking if students were okay, we had their responses to review and discuss with each student... Without Labdog, it would have been impossible to gauge each students' level of understanding at key points before and during the practical session, so it has been of great benefit to us"

DR highlights how the presentation of this data "in a way which makes it easy for educators to gauge where students are prior to the laboratory" is preferable over the current LMS, which "would require vastly more work to collate the data, and this simply would not be possible". This highlights the importance of adopting an applied approach such as DBR, which required me to consider both the necessary data from a technical point, and the pragmatic presentation of information of educators in the lab. Ultimately, this allowed DR to "give the highest level of personalised feedback to the greatest number of individual students that I've ever been able to deliver in a real-time teaching environment of any kind".

During-lab activities

The educators report that Labdog has helped develop a more conceptually focused laboratory environment. It went beyond "simple procedural queries, health and safety matters and good laboratory practice, by being able to identify students who are struggling with a particular concept and actually directly teaching them the theory, in a way that was never possible without such insight" (DR). Similarly, SB reported that "the introduction of Labdog allowed primarily for us to see what students understood and what they didn't, meaning that the discussion between demonstrators and students was significantly more focused that it had been previously".

DR reports this has lead to an "increased the number of high-quality questions that students pose during laboratory sessions", which he suggests evidences "more affective learning taking place in the laboratory itself". Though anecdotal, it is excellent to see this theme emerge - and provides promising direction and focus for future work.

Overall workflow

Both educators suggest that Labdog has integrated well into their practice before and in laboratory sessions, as well as suggesting that it has improved their practice. SB simply stated that "Labdog made everything very efficient... it meant that our time was being put to greater use, as those students who required our help more were receiving it."

DR reports that it has provided a "greater degree of structure for my teaching practice... Not only am I better organised... but the insights provided support me in identifying students who need assistance the most at the time they need it, and this means that my time is used much more effectively during the laboratory session itself." This interaction between using Labdog and focused student-teacher interaction was repeated by SB, who "would often do one round of the lab, and then move on to check the students' responses in Labdog. This would then inform my next movements, ensuring that I speak to those students who struggled first before speaking to those who were getting on okay".

DR reflects that this had wider implications, as students responses would allow him to "identify common themes which can then be addressed in other aspects of our teaching. It has been revolutionary in terms of its impact on my practice". This is another suggestion at the potential further, previously unrecognised, impact of Labdog: that it could inform teaching for specific concepts outside of the laboratory environment.

6.3.3 Concluding remarks

This short vignette of the educator experience of using Labdog helps complete the holistic picture of the experience of using Labdog. Though inherently anecdotal and experiential in nature, the educator-reported impact and experience of using Labdog provide further favourable evidence for its use.

The evidence discussed in this, and the previous, chapter present Labdog as a functional educational technology which brings both pragmatic and pedagogical benefits. This also represents the last piece of experimental work within the thesis. The following chapter summarises what is known, before discussing the implications for future work, and providing a literature-based discussion on the most important elements.

Chapter 7

Conclusions and implications for future work

In this final chapter I draw together the evidence and experience described previously to answer the research questions, posed in Section 1.6 (page 38). I begin with a very brief restatement of the theoretical foundations which rationalised the focus and design of the work. I then discuss the primary findings in the context of the research questions. In doing so, I identify several theoretical areas which could be used to increase the efficacy of future research in the area. I close the chapter by describing such potential future research, outlining a number of research questions and methodologies posed by this work.

7.1 Summary of the work

This research has involved the development, piloting, and evaluation of Labdog - a novel web-based educational technology for the teaching laboratory. Labdog is an online framework which educators can use to digitise traditionally paper-based practical instructions, i.e. lab scripts. The features of Labdog are detailed in Sections 3.3 and 3.4. Labdog is the first identified Laboratory Response System (LaRS), combining elements of classroom responses systems (CRS), e-portfolios, and learning-management systems (LMS) specifically for use in the teaching laboratory.

7.1.1 Theoretical bases

Cookbook chemistry. As discussed in Section 2.1, Labdog was built to prevent 'cookbook chemistry' (Domin, 1999b), wherein students follow the instructions of the lab script without understanding the scientific or logistical rationale. Cookbook chemistry has been identified at both school (Abrahams and Millar, 2008) and undergraduate (Adams, 2009; Galloway and Bretz, 2016) levels. This is a problem which needs to be addressed with instructional design (Hofstein et al., 2013) to promote conceptual understanding in students (Domin, 1999b; Abraham, 2011).

Pedagogical basis. Labdog was informed by two central paradigms: constructivism (Section 2.1.4) and self-regulated learning (SRL; Section 1.3). Constructivism (Taber and Watts, 1997; Denton, 2012) states that students learn by combining observations, actions, and previously-learned material in order to construct a more detailed understanding of the science or concepts at hand. SRL is a fuzzy concept, which defines learning as the process of independently setting and then working towards a goal (Zimmerman and Moylan, 2009). SRL in science is the result of students' motivation, cognition, and metacognition (Schraw et al., 2006).

Question-based learning is central to the project. The use of questions on their own (Karpicke and Grimaldi, 2012), as well as a tool for formative assessment (Yorke, 2003; Black and Wiliam, 2009; Bennett, 2011) make questions a valuable instructional technology.

Technological basis. The role of technology in education is discussed in Section 2.2. Web-, or cloud-, based technologies are those which are installed on remote servers which users access through web browsers. Web-based systems are increasingly popular globally, and within universities (González-Martínez et al., 2015). Several researchers have reported adopting web-based systems in the laboratory setting, e.g. in the use of electronic lab notebooks (Bennett and Pence, 2011; Amick and Cross, 2014; Weibel, 2016). Labdog advances such case studies by creating a purpose-built tool, as opposed to bootstrapping an existing technology.

7.1.2 The research design and focus

The overarching research question guiding this work was:

> How does the integration of Labdog, a novel web-based Laboratory Response System (LaRS), affect the student experience and perception of learning in and from introductory-level practical-based chemistry education?

To answer this question rigorously, I adopted a mixed-methods approach (Section 3.1.1). More specifically, I adopted design-based research (DBR, Section 3.1.2), a research paradigm which encourages flexible, adaptive, and iterative evaluation practices which ultimately seek to improve teaching and learning in practice (Barab and Squire, 2004; Reimann, 2011; Anderson and Shattuck, 2012).

7.1.3 The work conducted

Following literature reviews, and a series of interviews with UoS UG students (Section 3.2), Labdog began development in January 2015. Three separate one-off pilot studies were conducted between October 2015 to January 2017, and a year-long trial took place in 2016/17 with the science foundation year (SFY) programme at the UoS.

In the pilot studies (Chapter 4) students recognised Labdog as easy to use, and beneficial to their learning - though certain user interface (UI) elements needed to be improved. However, a number of technical problems heavily negated the impact of Labdog on students' learning or understanding in the earliest pilots. Additionally, it became obvious that designing practical activities in Labdog, specifically the focus of questions, and definition of what makes a single 'step', impacted the student experience. When Labdog was seen as drastically separate to the practical work at hand, students would abandon the software in favour of completing the practical.

These pilots informed the development of Labdog, to which I responded by improving stability and UI. At this point, Labdog was suitable for a year-long trial in 2016/17 with the SFY cohort (Chapter 5). Labdog was used over two semesters in nine practical sessions and 14 discrete practical activities, though only data from eight sessions and 13 activities were carried forward for analysis, due to technical problems. Unlike previous pilots, I examined the data produced in 2016/17 to better understand the relationship between the use of Labdog and student-perceived learning in the laboratory (Chapter 6).

7.1.4 Summary of Findings

A lot, and varied, evidence suggests that the use of Labdog promoted student awareness of the molecular processes related to their actions in the laboratory.

Taking data from the longest-running evaluation, the SFY in 2016/17, student completion of Labdog was 78% and 79% across semesters one and two respectively. This number represents the number of students who started using Labdog and reached at least the penultimate step in the practical. Three sessions showcased significantly lower usage, which can be attributed to practicals with iterative and/or time sensitive steps, e.g. a titration, where students could not easily break their activities to use their mobile device.

Overall, the vast majority of SFY students saw the use of Labdog as worthwhile. Pedagogically, the use of Labdog to create a LaRS-enhanced model of laboratory instruction provides a number of educational and logistical benefits to students and educators. Students reported that Labdog helped them think more about the scientific rationale and reasons behind their actions and observations in the lab. The ability to provide and received real-time formative feedback was repeatedly identified as one of the most valuable benefits of using Labdog for educators and students.

A number of students reported that Labdog placed excessive demands on their time and attention during the laboratory. SFY students were required to simultaneously complete the practical in Labdog, fill out their lab notebooks, and submit post-lab coursework. This caused a number of students to identify Labdog as a cause of stress or anxiety, which could counteract or prevent the educational benefits. Furthermore, there were ongoing complains regarding technological problems.

A pre-existing quantitative instrument, the MLLI (Galloway and Bretz, 2015), identified several differences between SFY and UG students' learning in and from the practical environment. A number of statistically significant differences emerged, namely that SFY students reported being less confused about the underlying concepts, and thinking about their interactions at the molecular level. Differences in staffing, content, assessment, and facilities between UG and SFY environments make it difficult to attribute differences solely to Labdog, however. Fortunately, qualitative data from interviews with SFY students suggests that Labdog led to an improvement of both of these things.

Examining SFY students' responses to a subset of questions in Labdog further supports the notion that Labdog encouraged students to engage with the atomic and molecular concepts relevant to their actions. The same data also suggests that teacher feedback to misunderstandings present in student responses was often inconsistent.

Very little evidence emerged which suggests that the use of Labdog has been detrimental to students' learning or understanding. Students who raised concerns cited concerns about workload or about the use instability of the technology. Labdog can be considered a successful example of DBR-informed education research, producing both a meaningful novel contribution to knowledge on learning in the laboratory context, as well as producing a viable and usable novel technology.

7.2 Answering the research questions

In the beginning of this report I posed the following overarching research question:

> "How does the integration of Labdog, a novel web-based Laboratory Response System (LaRS), affect the student experience and perception of learning in and from introductory-level practical-based chemistry education?"

I divided this broader question into a series of five smaller and specific sub questions:

- **RQ1:** What is the student-perceived role of the laboratory in their chemical education?
- **RQ2:** To what extent, measured through student-engagement, can Labdog be considered a technical and logistical success?
- **RQ3:** What is the student-perceived impact of adopting a LaRS-enhanced model of laboratory instruction?
- **RQ4:** To what extent do students who use Labdog demonstrate engagement with, and understanding of, the relevant scientific concepts to a practical?
- **RQ5:** What instructional design principles for a LaRS-enhanced laboratory environment can be drawn from the research?

In the following section, I discuss each of these sub-questions. Afterwards, I draw this discussion into a cohesive response to the primary research question.

7.2.1 RQ1: The laboratory has a varied role in students' education.

Pre-existing research has examined the role of the teaching laboratory from the educator's perspective (Bruck et al., 2010; Bretz et al., 2013). A subset of literature, e.g. Galloway and Bretz (2016) and Samarapungavan et al. (2006), has considered this question from the students' perspective. To better understand the student-perceived role of the teaching laboratory I conducted a series of interviews with both UG (Section 3.2) and SFY students (Section 5.3.2).

UG interviews. I interviewed 19 UG students, with interviews lasting 20-45 minutes. Students were asked four questions about their perceived role of the laboratory, and three questions about their self-reported metacognition. Interviews were transcribed and subject to a Grounded Theory analysis (Glaser and Strauss, 2009). Three main findings emerged from the interviews:

1. Five types of student-perceived benefits emerged: improved procedural competencies, better insights into relevant chemistry, better understanding and learning, personal/affective experiences, and the development of professional skills.
2. A distinction between 'applying' and 'contextualising' chemical principles emerged, differentiated by active and passive language respectively. While some students actively sought to relate chemical theory to their work, others more passively considered how their experiences in the laboratory gave them 'hands-on' experience with seeing chemicals, e.g. knowing what a certain chemical looks like.
3. Half of the students stated the laboratory did help them develop their conceptual understanding.

In total, eight students reported that the laboratory helped advance their understanding, and seven stated that it helped to contextualise or put knowledge into practice. A smaller number of students stated that the laboratory provided personal satisfaction (5), or gave them the experience of a professional chemist (5). Very few students identified transferable skills like collecting data (2), solving problems (2), increasing confidence (2), or developing initiative (2).

Fifteen (83%) UGs reported the laboratory was a chance for them to see and potentially engage with the relevant chemical concepts. These students demonstrated linking the macroscopic observations to their submicroscopic scientific causes (Johnstone, 2000). As one of the UGs stated:

> "You could see the chemistry happening in front of you... [which made it] easier to understand... what's going on and why you're doing what you're doing"

Other students saw labs as preferable over lectures: "rather than just being able to know a fact and reel it off in an exam, you'd actually be able to apply it, know what it really means". These are clearly benefits indicating that the laboratory provides and requires a greater extent and detail of understanding from students (Krathwohl, 2002), who are applying concepts in new scenarios.

A divide emerged between students who actively and passively considered the knowledge. One active student complemented the lab because "it's reinforced a lot of what I've just done in first year, it gives me a chance to put that into practice and see 'okay I've learned about this and now I can actually do it' and go back and look at the theory behind it". Contrastingly, some students saw the role as much less active: "you need to know what colour should it be, should it be a solid, should it be a liquid, should it be hydroscopic so are you expecting it to stick to the sides of the glassware?".

Nine (47%) students stated directly that the laboratory has negligent to no benefit to their understanding or learning. For example, one student thought the laboratory developed the procedural competencies: "the actual teaching lab itself, I don't think it didn't really help much with the understanding more with just the practical". Another identified problems with the instructional design: "There wasn't really much of a learning style in the labs, it was more 'you have this script, do it, ask us questions if you're stuck or anything' so we didn't really learn too much".

This presents a varied view of the role of the laboratory. While some students already see the laboratory as intrinsically linked to the chemical concepts they are learning, others suggest a fundamental disconnect between the two. In agreement with the literature, the laboratory offers many potential benefits, perhaps suggesting at a greater need to design instruction around the most desired outcomes.

SFY interviews. A series of eight shorter (10-25 minute) interviews were conducted with eight SFY students in December 2016. These interviews too were also analysed using Grounded Theory. These interviews were designed to examine students' opinion of using Labdog, but also touched on the laboratory in general.

Similar to the UGs, SFY students valued the laboratory as it allowed them to apply (5, 62%) and visualise (3, 37%) the chemical theory relevant to their actions. Perhaps due to the smaller number and shorter duration, these interviews presented a smaller diversity of benefits and therefore they contribute in a less meaningful way to answering this research question.

Compared to UG students, SFY students were notably less likely to mention that the laboratory could improve their experience or expectations for a career in science. This could be due to a difference in skill, activity type, or distance from entering the workforce.

7.2.2 RQ2: Labdog became a technically stable technology.

DBR dictates that interventions need to be achievable by a teacher with little special knowledge or technical skill (Herrington et al., 2007). As Labdog is a completely novel technology, it is important to demonstrate and examine how, and if, it was used successfully from a technological/logistical standpoint.

I used student usage data as a proxy for stability. If Labdog were unstable, we would see low student retention, i.e. students would not consistently progress to the end of the practical. To generate this data I developed an analytic engine within Labdog that recorded individual student progression.

Specifically I used the percentage and number of students who reach the penultimate step. I presented this information for SFY piloting in 2015/16 (Table 4.1), and in A-level outreach events in 2016 (Table 4.2) and 2017 (Table 4.3). For the year-long trial in SFY I present data for semester 1 (Table 5.2) and 2 (Table 5.4). Note that for each of these investigations, students always had a paper copy of the lab script.

Over five practical activities in the 2015 SFY pilot, an average of 42.5% of students reached the penultimate step. I also identified specific steps and questions where student usage of Labdog fell the most. When asked to shift rapidly and continually between Labdog and the practical activity, usage would decrease. This was made worse in scenarios where students were given more than one question to answer before they were able to progress.

In the 2016 outreach events, 59.5% of 121 A-level student pairs reached the penultimate step of a natural product extraction over seven sessions. In nearly identical events in 2017, this fell to 51.1% of 137 students over eight sessions. This average was reduced by two low-usage sessions (35.3% and 33.3% respectively), where my absence meant Labdog was not formally introduced to students. When removed, the average increases to 56.5%, a comparable figure to the previous year. During both years there were ongoing problems with wi-fi access in certain parts of the laboratory

During semester one of the SFY year-long trial, an average of 78.23% students reached the penultimate step of seven practical activities over five sessions, compared to 79.63%

in six practical activities over three sessions in semester two. However, there was a reduction in the average number of students who started using Labdog in semester one (38.14) compared to semester two (31.17), due to students withdrawing from the SFY. This higher usage of Labdog within the SFY, compared to other pilts, can likely be attributed to the more consistent environment of the SFY.

The majority of students were able to consistently use and engage with Labdog. In the one-off nature of the outreach events, students tended to need an introduction to Labdog and encouragement to use it from the beginning of the session. By comparison, when integrated more wholly into the instruction of the SFY, students were more likely to engage with Labdog until the end of the activity. Given this engagement from students, specifically the high retention in the year-long SFY pilot, Labdog can be considered a stable and often-used technology.

7.2.3 RQ3: Many students stated that Labdog benefited their learning.

The reported impact of Labdog on student learning was measured in surveys in the pilots (Chapter 4), and the year-long SFY study (Section 5.3). Additionally, SFY students were interviewed by a member of teaching staff in 2016/17 (Section 5.3.2). The evidence consistently suggests that the use of Labdog has a positive impact on students' perception of learning.

Pilot surveys. A survey given during the 2016 Twilight events, showed that approximately three quarters of respondents (n=56) rated Labdog's ease-of-use as $\geq 4/5$ on a subjective scale, and 63% of students rated Labdog as $\geq 4/5$ in terms of helpfulness. Students were asked the open question 'how do you think Labdog helped you during this session', and of 42 responses, 23 were positive, three were negative, and 16 were mixed. Most frequently (20), students mentioned that Labdog helped them learn, giving comments such as "It encouraged us to learn a deeper understanding of the chemistry behind our experiment and allowed us to venture deeper into the depths of the scientific realm". Ten students reported that using Labdog slowed them down during the sessions, and a few (4) felt that it was simply too much work.

An almost identical survey was conducted in the 2017 Twilight events. From 88 responses, students reported a largely positive opinion: with 93% rating the ease-of-use at $\geq 4/5$, and 74% rating it likewise in terms of usefulness. When asked the open question if Labdog helped or hindered students during the session, responses were overwhelmingly

(63) positive with very few (2) wholly negative responses, and several (9) mixed. Many (26) students felt Labdog presented them the practical in a clear way, and 22 stated that it helped them think about the chemistry behind their actions. A smaller, but still noteworthy, number (12) identified that Labdog lead them to a better level of understanding.

Two surveys were conducted during the 2015/16 SFY pilot: after sessions one and four. Students continually found Labdog easy to use, but there was a notable shift towards negative evaluation after the fourth session. When asked for open comments, four students criticised the UI design, and six critiqued the instructional activities within Labdog. Four students identified that using Labdog alongside paper-based lab books caused issues with concentration and stress.

SFY surveys. The SFY study in 2016/17 presented a chance to examine the impact of Labdog in an integrated and longer-running perspective, compared to previous shorter pilots. Three surveys were conducted during 2016/17: two in semester one, one in semester two. Aggregating from responses to a survey after the first (n=34) and fourth (n=36) sessions, 85.5% of students responded 'strongly agree' or 'agree' (abbreviated to S/A) that Labdog helped them feel "actively involved and engaged in what I was doing, not just following a set of instructions", and 72.2% of students responded S/A that Labdog improved how much they learned from the practical work. This included benefits to both procedural competencies (68.06% S/A), and to a lesser extent, scientific concepts (59.72% S/A).

A more detailed survey was constructed in semester two, which was informed by the previous survey and SFY interviews (discussed below). The survey investigated five areas: application of theory to practice, engagement with scientific principles, overall learning process, use of technology, and perceived understanding of the scientific principles. 22 students (of 36 in the cohort, 61%) responded to the survey, and I generated Cronbach's alpha scores for each of the five categories of questions. Strong (≥ 0.7) values were found for application of theory, engagement with concepts, and understanding of concepts, suggesting questions in these areas were investigating the same underlying factors. Students reported that Labdog helped them visualise the scientific principles of their actions, as well as consciously consider such principles during the laboratory. Furthermore, students reported that Labdog lead them to a better understanding of the chemical concepts, specifically through the answering of questions and the provision of formative feedback to their answers.

Students were given two open-answer items: to explain if Labdog was useful to them, and if using it was worthwhile overall. Responses for its usefulness were split between positive (5), negative (6), and mixed (7)[1]. The biggest student-reported barrier to Labdog being seen as useful was the extra workload required of them, e.g. one student stated that Labdog "took too much time and attention from carrying out the practical". Five students stated that the overall benefits to their learning were useful, while six also clarified that the questions made Labdog useful. The majority (14) of responses (n=21) to the question 'was using Labdog worthwhile?' were positive. Ten students stated that Labdog was worthwhile because of the questions or the benefits to their understanding. Those students who believe Labdog was not useful cited technical error (2), a belief that technologies should not be used in the laboratory (2), or the extraneous demands on time and effort as making Labdog not worthwhile.

SFY Interviews. At the end of semester one, seven SFY students were interviewed (Section 5.3.2). The interview covered the use and perception of Labdog, and perceived-impact of using Labdog on their learning process. The interviews presented further strong evidence that Labdog has led to more conceptually-focused learning in and from the laboratory. Again, students attributed this almost entirely to the question-based nature of Labdog and the subsequent formative interactions with demonstrators.

Several quotes from these interviews validated the intended pedagogical goal of Labdog. One student reported that because of Labdog "it has made me sit there when I'm doing things and think 'why am I doing that?'". Another exemplar quote extends on this theme: "you answer it whilst you're doing it, which is much more likely to be beneficial, because you can then understand while you are doing, and you're more likely to understand what you're going to do next or what you have done previously". The technological nature of Labdog enhanced this benefit, as stated by another student: "this questioning of what have you seen and why have you seen it. Asking those questions, which, I don't know if it was just written down in the schedule, but it just seems much more interactive in the Labdog format".

Not only was the act of answering the questions beneficial, but the real-time interface allowed demonstrators to spot misconceptions as they arose. As one student stated: "if you are having issues it is hard to ask for help. Where during chemistry it is like they actually come up to you and then you realise that you made a mistake, they

---

[1] Not all respondents provided an answer to these questions

correct it for you, And I feel like that is really memorable as well". This presents a unique benefit of Labdog as a web-based and real-time educational technology, which can collect information on all students synchronously.

Several students reported experiencing technical issues with Labdog, e.g. the were unable to submit photos when required. Though Labdog became more technically stable over the year, both in my experience and reported by the SFY students in the second survey, this seemed to establish a certain amount of poor will among students that became hard to address. Secondly, students repeatedly raised concerns about the demands which arose from having to use Labdog alongside paper-based lab reports. These concerns can both be addressed through continued technical and instructional development.

7.2.4 RQ4: Students demonstrated and reported engagement with chemical concepts within Labdog.

Nonetheless, students repeatedly reported that using Labdog led them to engage more with the relevant scientific concepts. To investigate this I examined a series of SFY students' responses to questions in Labdog from 2016/17 (Section 6.2). Secondly, I used a pre-existing survey instrument (Section 6.1): the meaningful learning in the laboratory instrument (Galloway and Bretz, 2015, MLLI), which measures student-reported learning in the laboratory environment. Both SFY and UG students completed the MLLI in December 2016 and April 2017.

Exploring students' responses to Labdog questions posed during practical sessions. I examined the questions posed to students in semester one and identified six distinct types: observation, calculation, definition, method, reaction, and other. Of the 36 questions posed to students during practical activities in semester one, one third (12) asked students to make relatively simple, but time-consuming, calculations. Students' responses to such questions are limited in scope, providing simple and discrete pieces of information, which may nevertheless demonstrate conceptual understanding. However, there was a need to identify questions which can more reliably demonstrate students' (mis)understanding. To do so, I started with the most simple and quick exploratory analysis: the length of responses. There was a statistically significant difference in the length of responses (number of words, ANOVA $P \leq 0.001$) between question types, supporting the implicit idea that different types of questions encourage students to provide different lengths of responses.

As a result of experience and reflection, by semester two there where only one question each for calculation, method, and definition types. In semester two, students were asked 12 questions over six practical activities, an average of 2 per practical compared to 5.1 in semester one. Furthermore, these questions were predominantly (9, 75%) about observing and explaining in the laboratory. This was a purposeful decision by the teaching staff, who wished to use Labdog to provide more detailed insight into students' understanding. This raises questions about the suitability of statistical analysis, as well as the specific value of word-counts as an insight into the quality of responses.

To further examine the quality of student engagement with the questions, I conducted an exploratory analysis of the responses to three questions asked in semester two. These questions were all of the observation/reaction category - a new type of question identified in semester two, characterised by the unified need to observe and explain the same question. The questions related to i) the reduction of silver during the reaction between Tollens' reagent and an aldehyde (N=28 responses), ii) the reduction of copper during the reaction between Fehling's solution and an aldehyde (N=30 responses), and iii) an explanation of how pH correlates to acidity (N=37 responses).

In each question, the vast majority of responses demonstrated an understanding and engagement with the most fundamental process: 92% of students identified that silver was being reduced, 80% that copper was being reduced, and 83% identified the stronger or weaker acid. To this end, most students demonstrated an ability to relate an observation to the theory, e.g. the appearance of a precipitate as an indicator that some reaction had taken place.

Conducting an exploratory analysis of the questions revealed that only around half of responses included causal or procedural links. For example, only 50% of students in question one attributed the formation of a precipitate (i.e. the macroscopic) to the reduction of silver (i.e. the sub-microscopic or chemical process). In the very similar question about copper reduction and precipitation, no students linked the formation of a precipitate and the reduction of copper. This is despite the fact that 46% of responses recorded the formation of a red precipitate, and 53% correctly identified the reactant and product. Contrastingly in the third example question, 89% of responses made some attempt to link pH and the strength of an acid, with 67% explaining the difference in acidity in relation to the substance's liklihood to dissociate into the conjugate base and proton.

A notable minority of responses were rejected by teachers using Labdog's interface: 21.4%, 30%, and 27% for the questions respectively. Feedback was split between a request to provide more specific detail, e.g. the species involved in a reduction, and feedback which asked for more detail of the chemical process or details, namely in relation to dissociation and acid strength. Lack of specificity in the response is characterised by a student stating only the first part of the reaction: "Cu^{2+} gains electrons from the aldehyde" - without stating what was produced. Similarly, the response may lack detail, as a student may provide the most simple/direct response to a question at hand, e.g. "The pH values are greater in ethanoic acid compared to those of hydrochloride acid", without providing explanation or qualification.

I identified a number of non-rejected responses with similar characteristics to those rejected by the demonstrators. Although the absolute numbers were relatively small I identified an additional 6, 7, and 3 responses respectively. In the first two questions these were dominated by students who gave detail of only the reactant *or* product of the reduction. Within the third question, three did not relate acidity to dissociation. This suggests that educators need to provide more consistent feedback through Labdog, especially when multiple demonstrators are working simultaneously. More consistent and thorough feedback will in turn work to create an environment where students are consistently required to demonstrate detailed understanding, if they know any errors will be noted and addressed by educators. Additionally, this is only a measure of the feedback which took place in Labdog, and it is possible that feedback took place outside of Labdog.

The MLLI. The MLLI is a pre-existing and validated survey instrument (Galloway and Bretz, 2015) which examines learning in and from the practical environment. The MLLI was given to both UG and SFY students two times during 2016/17. UG students completed it as online survey in pre-lab activities, and SFY as a paper survey during workshop sessions. Unlike the original researchers' survey I used categorical responses to items, and not the continuous slider-scale between 1-100 - so as to keep the methods consistent with all other student surveys conducted in the research.

There were a number of differences between SFY and UG responses (Table 6.2). In semester one, four questions returned a significant difference, compared to two in semester two. The distribution of responses to these questions is shown in Table 6.3.

Four of the items which returned a significant difference likely cannot be attributed wholly to Labdog. These items were about students worrying about getting 'good' data, getting 'stuck' but continuing, making mistakes and continuing, and being fascinated by the equipment they used. Differences in these items can more reasonably be attributed to differences in content and assessment methods between the cohorts. The SFY has a more informal atmosphere and is taught in a general teaching laboratory which lacks diverse or specialist equipment. Differences in instruction styles and assessment methods create further confounding differences.

Nine items on the survey were conceptually relevant to the nature of Labdog (highlighted in blue in Table 6.2). Two of the questions had statistically significant different distributions between SFY and UG students: their reported focus on procedures over concepts (Figure 6.1), and confusion about the underlying concepts (Figure 6.2). UGs report a greater focus on the practical procedures over the relevant scientific concepts, compared to SFY students. Additionally, SFY students were consistently less likely to report confusion regarding the related chemical concepts to a laboratory, which was close to significant in semester 1 ($P = 0.056$).

A complex narrative emerges from various other items in the MLLI. When asked if students considered 'what the molecules are doing' during a reaction in the laboratory, there is a notable growth in percentage of SFY students agreeing $\geq$ 40% between semesters ($S_1 = 55.9\%; S_2 = 86.3\%$), while UGs show a relatively high but stagnating response to the same question ($S_1 = 79.8\%; S_2 = 77\%$). Similarly, over the two semesters, SFY students become increasingly likely to agree ($\geq$ 40%) to interpreting their results at the molecular level ($S_1 = 77.47; S_2 = 86.5\%$), compared to a lower and more consistent UG report ($S_1 = 76.7; S_2 = 78.74\%$). These results suggest that the use of Labdog consistently over time may encourage students to focus on the molecular domain of chemistry and that other chemical instruction may not encourage such focus. This provides further evidence that Labdog has had positive effects on students' education.

Despite a greater focus on the chemical concepts related to the practical work, the data suggests that UGs are more likely to understand the 'reasons behind the procedure', see the procedures as 'simple to do', and develop greater confidence in the lab. Therefore, while Labdog may help students focus on the chemical processes, it has an unclear effect on confidence and perceived ability. This is likely confounded by an un-measured difference in the ability, confidence, and self-efficacy of students as they enter their respective years.

The research has presented a mixture of self-report and observational data, providing a varied picture of student engagement with molecular processes, both with and without Labdog. Given the data presented both here and in the response to RQ3, it would be difficult to suggest that the use of Labdog did not result in students giving greater conscious attention to the conceptual aspect of their actions. That is, it had students relate the sub-microscopic and the macroscopic. This could no doubt be improved by carefully considered design, in both focus and number, of questions in Labdog, as well as more consistency in educator feedback. It is important to consider these findings in the broader context of SFY and UG study.

7.2.5 RQ5: Instructional principles for a LaRS-enhanced laboratory

The previous questions described and evaluated how, and if, a LaRS-enhanced model of laboratory teaching affects student learning. The purpose of such evaluation, and this overarching research, is to develop the global research understanding into the use of practical work in chemical education (Wang and Hannafin, 2005). Herrington et al. (2007) state that there is a need to develop theory that contain:

> “substantive and procedural knowledge with comprehensive and accurate portrayal of the procedures, results and context, such that readers may determine which insights may be relevant to their own specific setting” (p.7)

Given the almost countless possible contexts that Labdog could be deployed in, across subjects, abilities, and geographies, I present a research-based discussion of two instructional principles which be more universally applied for educators who may wish to use Labdog, or some other LaRS system. From this, future educators will hopefully be able to apply good design principles to their own context.

As I have iterated over the course of the research, the LaRS-enhanced model of laboratory instruction has shown promise, but several conceptual and practical barriers have been raised. Although there are many potential points of discussion, I wish to focus on two primary design principles:

1. Encourage flow state.
2. Design around Cognitive Load Theory

7.2.5.1 Encourage Flow State

Across both the pilots and the year long trial, students suggested that they resented being interrupted from ‘getting on with the practical’. Furthermore, there was a notable reduction in student engagement with Labdog when students had to frequently task switch between using Labdog and performing the practical. In order to better understand the cause and effect of this problem, I consulted the research literature on psychological focus and engagement. This lead to the identification of ‘flow’. I could find no research literature relating flow to the teaching laboratory, making this discussion both exploratory and extremely valuable.

Those in a ‘flow state’ are extremely engaged with a problem or activity, experiencing intense focus and concentration, and losing track of other (sub)conscious awareness, e.g. time, hunger or fatigue. Flow is the complete investment of attention into a current activity. It represents a state of extreme engagement, motivation, and cognitive performance where individuals are engaged in a problem or activity, working at peak cognitive performance to solve it.

The phenomenon first emerged in the study of creative individuals, e.g. painters or dancers who would become incredibly fixated on their work or activity only to be largely disinterested once they became disengaged (Csikszentmihalyi, 2000). This applies as much to research scientists in the laboratory as it does to dancers in a studio. Most people, at some point, experience flow, and can relate to the feeling of being so engaged by an activity that all else seems to fade from consciousness. Flow is related to, and arguably the result of, attention: what information students bring into their awareness, how intensely they focus on it, and for how long (Csikszentmihalyi, 2014).

It is therefore unsurprising that flow is associated with increased learning and academic performance (Egbert, 2004; Schüler, 2007). As a form of intrinsic motivation it can be linked to the wider learning literature, and SRL specifically (Pintrich and De Groot, 1990; Schunk, 2008).

Research around flow is relatively new. Experimental evidence for flow is affected by a number of methodological and practical limitations, namely in inducing flow in controlled conditions, and measuring an intrinsically subjective experience (Moller et al., 2010). The increasing use of well-designed experimental conditions is also producing a rapidly evolving understanding of flow, which contradicts previous understandings or findings, e.g. Harris et al. (2017). This does not detract from the benefits of flow, but rather to the specifics of inducing it.

Nevertheless, the flow state is made up of six characteristics (Nakamura and Csikszentmihalyi, 2014):

1. Intense and focused concentration on what one is doing in the present moment;
2. Merging of action and awareness;
3. Loss of reflective self-consciousness, i.e. loss of awareness of oneself as a social actor;
4. A sense that one can control one's actions, i.e. a sense that one can in principle deal with the situation because one knows hot to respond to whatever happens next;
5. Distortion of temporal experience, typically a sense that time has passed faster than normal;
6. Experience of the activity as intrinsically rewarding, such that often the end goal is just an excuse for the process.

As opposed to the often-studied areas of athletics or creativity, education presents a more challenging domain in which to apply and study flow (Fong et al., 2015). Where leisure activities are implicitly engaging and voluntarily, many educational activities are obligatory. In general, flow is less common and harder to study in non-leisure activities (Abuhamdeh and Csikszentmihalyi, 2012). This difficulty is attributed to the higher intrinsic motivation and goal orientation present in leisure activities (Abuhamdeh and Csikszentmihalyi, 2012). As a result, students with higher self-determination and self-regulation are more likely to have a higher quality experience in challenging educational settings (Bassi and Delle Fave, 2012). This shows a return to the concept of SRL, where an individual's actions and experience in a context setting are determined notably by motivation, goals, ability, and metacognition (Zimmerman and Moylan, 2009). This brings the study and induction of flow into the wider context of SRL, legitimising its involvement with this research.

By considering how flow can be encouraged, both the process and outcomes of learning can be enhanced (Engeser and Rheinberg, 2008). I could find no published research relating flow to the science laboratory context in specific. It has been studied in relation to procrastination and self-regulated learning (Kim and Seo, 2013), creative writing

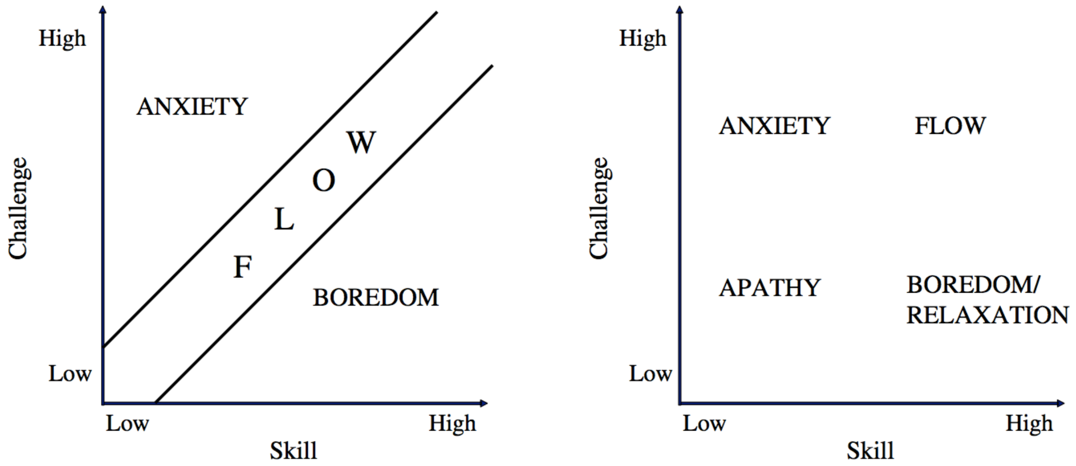


Figure 7.1: The linear and quadric models of the challenge-skill relationship in flow, taken from Engeser and Rheinberg (2008)

(Kaufman, 2002), language learning (Egbert, 2004), engagement with science lessons (Linnansaari et al., 2015), the use of educational technology (Pilke, 2004) - especially the use of gamed-based education, i.e. gameification (Lim et al., 2006; Pavlas et al., 2010).

The findings of this research are relevant to laboratory instruction in general. Three instructional principles which educators should bear in mind, informed by literature, are:

1. Balance the challenge of questions with students' skill level;
2. Use formative feedback to maintain student focus on the practical;
3. Space and group Labdog activities to minimise interruption to students' workflow.

Balancing challenge with skill. Flow requires the material presented to students to be "just-manageable" (Nakamura and Csikszentmihalyi, 2014, p.92). This notion of balancing skill with challenge, i.e. 'balance', is a well-studied aspect of the flow state. There are two ways to perceive the relationship of balance and flow, represented graphically in Figure 7.1. The left graph represents the the earlier linear model (Csikszentmihalyi, 2000), in which "people must continually engage in new challenges to match their increasing skills, and they must perfect their skills to meet the challenges" (Hektner and Csikszentmihalyi, 1996, p.4). This relationship scales linearly, in that higher skilled individuals need proportionally more complex tasks to achieve flow.

The evolution of this model, presented in the right of Figure 7.1, asserts that flow can only be achieved by highly skilled individuals performing more highly challenging tasks (Csikszentmihalyi and Csikszentmihalyi, 1992). Under this model, lower skilled individuals do not experience flow, even under conditions which would do so under the linear model. Instead, such individuals will experience apathy (Whalen, 1997).

In a meta-analysis of 28 studies, Fong et al. (2015) examined the quantitative evidence for the impact of balance on flow. The authors found “adequate support that matching skill and challenge is robustly related to feelings of flow or optimal experience” (p.440). In investigating eight moderating factors, the authors found that the impact of balance on flow was diminished with younger, i.e. under thirty years-old, individuals and in more individualist, i.e. not group-focused, settings. Both of these are true for the research I have conducted: where most students were young, and working in pairs.

One potential explanation for the importance of balance in older individuals is that they become increasingly experienced, and therefore have greater, or more broad, set of skills (Wolfe and Kolb, 1980). However, Fong et al. report an unclear, potentially non-linear, relationship between age and balance-flow, because of the limited sample data. If balance is more important for older students, then for many educators who work with novice students, there is a less pressing and intricate need to balance skill and challenge, but rather rely on the internal motivations of the students.

Leisure and education are fundamentally different research contexts (Abuhamdeh and Csikszentmihalyi, 2012). Therefore, balance is probably best viewed in a preventative sense, i.e. should be sought to prevent non-engagement (anxiety, boredom, apathy), rather than to promote flow. It is unlikely that balance alone is enough to promote flow, and instead educators should prevent it being another barrier to engagement with problem-solving. Unfortunately, this presents a moving-goalposts situation, where an individual student’s intrinsic motivation, goal-orientation, and preference for greater challenge all influence their response to varying levels of balance (Bassi and Delle Fave, 2012). This makes it complex to design for cohorts with which the educator is not familiar.

Pragmatically, however, there is a need to present actionable decisions. First and foremost, educators should avoid posing questions which are too complex, and drastically exceed students’ ability. Educators should be wary of the challenge of completing the practical procedure itself. In semester one 2016/17 we underestimated this, as characterised by asking a great number of questions throughout a practical, born from an underestimation of the intrinsic complexity of completing the practical.

Interestingly, it is unclear if different types of challenge, e.g. procedural versus conceptual, can be treated as separate or related. Here it may be wisest to adopt a precautionary approach, by presenting questions which involve as few concepts as possible and limiting the complexity of the questions which involve numerous concepts. This is the most moderate compromise, which may especially benefit students as they simultaneously learn new practical techniques and conceptual material. Targeted research into this presents an area of interesting future research.

Ultimately, the difficulty of the question comes from the discretion of the educator. The difference in syllabus, abilities, and practical activities make specific cross-context advice impossible. Adopting an iterative approach to evaluation in this research allowed insight into such factors, and it is wise for educators to do the same: reviewing students' responses, and listening to their experience.

Feedback and interruptions Flow requires “clear proximal goals, [and] immediate feedback” (Nakamura and Csikszentmihalyi, 2014, p.92). Feedback should provide individuals with information about their progress in an activity, and how it could be adjusted to be improved (Abuhamdeh et al., 2005). In the laboratory context, students should be aware if their practical activity is (not) going as planned, and what they need to do in order to finish the activity.

Feedback is an essential component of good instructional design universally (Norman, 2013). The benefits of targeted formative feedback are well studied and established in education research (Hattie and Timperley, 2007; Black and Wiliam, 2009; Harks et al., 2014). Instant and unambiguous feedback is essential to promoting flow state - individuals need to be accurately aware of their progress (Csikszentmihalyi, 2000).

In the traditional laboratory, students could use observations and measurements as feedback. For example, are students seeing what they expect to see, or is their data showing what they expect it to? This, implicitly, requires that students have some idea of what to expect, i.e. the ‘clear proximal goals’ in either their observations or the data. Such goals could be set prior to the laboratory, e.g. in pre-labs or lectures, during the laboratory itself, in the lab script, or instruction from the educators.

Labdog already provides an overview of student progress from their home-screen dashboard, showing a percentage completion of both pre- and in-lab activities. More immediately in the lab, educators are able to use students' responses to questions to provide real-time feedback. This can include written or verbal feedback delivered in-person. Real-time data and feedback were developed in 2016/17, and the SFY students reported that this functionality and interactions were incredibly valuable in developing their understanding. However, an exploratory analysis of a subset of semester two questions revealed that educators would only feedback on approximately half of the incorrect/inappropriate responses.

Educators should consider using questions in Labdog which can facilitate feedback. This could be a case of simply reporting an observation, or promoting the use of the 'request help' function if students feel that they're no on the right track. This needs to be paired with the clear statement of proximal goals, i.e. students should know what they should observe or record. Such information can easily be placed in the stepwise instructions presented to students in Labdog.

Additional functionality could also be added to Labdog, e.g. adding conditional progression through a practical. For example, Labdog could present students with a checkbox or button asking them if they (did not) see or record some specific value or observation. Students who do not appear to be on track could then be redirected back to a predefined step in the practical. Making educators aware of when this happens could lead to in-person interaction. Alternatively, students could be presented with preprepared troubleshooting advice.

If educators have to approach students to provide feedback regarding a previously given answer, there is a potential that this could interrupt students, thus interrupting attention and flow. Pragmatically, this is likely, given the student-teacher ratio, and the time requirements of thoroughly addressing a student's misconception. Students must be drawn out of their current activities to revisit some previous step which they may have forgotten. Specific research is necessary to address the exact impact of this, which will provide more specific information on the cost of compromises between real-time and delayed feedback. At present, educators should use the existing functionality in Labdog to filter the responses which come through on the real-time dashboard, providing feedback on only the most conceptually or procedurally pertinent questions.

The spacing or clumping of activities. In order to enter the flow state, students need to focus on *just* the problem at hand. This means allowing them to pursue an individual problem to an appropriate conclusion before they are asked to perform related, but distinct, activities. For example, a student may be asked to record data as part of a problem, but the analysis or plotting of such data is in itself a separate activity.

For example, when students were asked to measure the temperature of their solution in the 'coffee cup calorimetry' experiment. Students had to record data every thirty seconds for approximately five minutes and were simply not able to do so while using Labdog. In part, this may be due to the dual use of mobile phones as stopwatches and for Labdog, but was largely due to the splitting of attention between the two.

This lead to a second problem, however. As noted in the twilight events: when students started to enter the flow state when conducting the practical work, they would not readily return to Labdog. This is evidenced by a significant decline in student engagement with Labdog. In this instance, this was exacerbated by leaving the workbench to use a rotary evaporator.

Here, I suspect, students were entering the flow state and not willing to leave it. It is reasonable to assume that they did not want to re-start using Labdog because they did not want to have to invest time and effort to simply catch up to where they already were. This runs counter to being in the flow-state. As such, Labdog needs to be integrated into the problem-solving activity. This involves acknowledging when students need to 'just get on' with their actions, and allowing them to do so. Fundamentally this comes down to what constitutes a single 'step' in a practical. Students should not feel penalised for returning to use Labdog after they have completed a section of their work. In such instances where prolonged time is spent away from the software, posing no questions, or a simple summary question may be most appropriate.

Additionally, educators should identify moments where there is a lull in activity, for example while something is heating or cooling. These are moments when students' flow would be broken anyway. Such limits can be inherent to the activity, e.g. crystallisation, caused by logistic constraints, e.g. a limited equipment, or could even be artificially placed by educators, e.g. simply instructing a student to wait for longer than necessary for a crystallisation to occur.

7.2.5.2 Design around Cognitive Load Theory

Cognitive load theory (CLT) is a model for understanding the limits and nature of human cognition, specifically as someone engages with a task. There are two broad systems of memory (Baddeley et al., 2009): working memory (WM) and long-term memory (LTM).

WM is the system used when handling and interpreting information in the moment, e.g. when a student is actually in the laboratory.

The capacity of the WM is severely limited, and it is generally accepted that it can only handle 4 ± 1 items of information at once (Cowan, 2010). Furthermore, any information left unused in the WM, is lost within about 30 seconds (Paas et al., 2010, p.117). Despite this incredibly limited capacity, WM is the gateway through which information moves into LTM. LTM itself can store and organise incredible volumes of information, without such constraints for size or time. This transaction of information between working and long-term memory is likely one of the most conceptually simple notions of learning.

When the WM's capacity is exceeded, so-called 'cognitive overload', learning is severely hindered (de Jong, 2010). CLT is the study of how and when this happens. Initially developed by Sweller (1989, 1988), CLT has established itself as both experimentally demonstrable and pragmatically useful in designing effective instruction (Van Merriënboer and Sweller, 2010; Wang et al., 2011; Paas and Ayres, 2014).

The importance of CLT to practical activities was mentioned by Millar and Abrahams (2009), who stated "tasks that strongly involve the domain of ideas are likely to have significantly higher demand than those which simply aim to allow students to see, and remember, an observable event" (p.64). Recent education research has also examined how CLT affects students in the blended learning environment. This interest is driven by how easily multimedia can cause cognitive overload, due to the richness and diversity of information they present (Brunken et al., 2003; Mayer, 2003). The susceptibility to cognitive overload from multimedia stems from two fundamental assumptions of human cognition: dual-channel (Baddeley, 1992), and the dual-coding (Paivio, 1990). The dual-channel assumption asserts that auditory and visual information are treated as two separate types of information by the brain, even when they are intuitively linked, presenting information about the same idea at the same time. The dual-coding assumption states that both auditory and visual data are handled differently by the brain. Instructional environments which have students constantly transition between information presented visually and audially can therefore easily lead to cognitive overload.

Cognitive load is a well-researched and nuanced area of understanding. In the following pages, I summarise this complexity into actionable research which could minimise the chance of cognitive overload from the use of Labdog.

Our current understanding describes cognitive load from three sources: intrinsic, extraneous, and germane (Van Merriënboer, 2005). Respectively, these relate to the complexity of the material to be learned, the interpretation of the instructional material(s), and the act of learning in itself.

Therefore, I wish to focus on the cognitive load from three sources:

1. Conducting the practical procedures;
2. Intrinsic Load: The use of Labdog, and interpretation of the resources within it;
3. Extraneous Load: The complexity of the concepts being presented.

I will discuss each of these in turn in the rest of this chapter. My intention is to provide actionable goals for practicing educators who wish to use Labdog, or some other LaRS-enhanced model of teaching. This also explains my decision to largely ignore Extraneous load. We must assume that practicing educators often have little control over their syllabuses and time-frames. Especially at HE level, the concepts must be inherently challenging or complex to grasp.

The cognitive load associated with procedures

The laboratory presents a ripe opportunity for cognitive overload. Yet, the laboratory also presents a context which is notably distinct from textbooks or online environments: it is dependent on time and location, combining the following of procedures, mastery of manual skills, and conceptual understanding of any observations made. As such, the historical CLT research may be of limited applicability:

> "[CLT] assumes that most knowledge does not have to be gained from experience but can be borrowed from other people (which has an obvious evolutionary advantage) and thus it follows that cognitive load theorists spend much of their time considering how that knowledge should be structured when it is presented to learners as well as in which activities learners should engage when acquiring information" (Paas et al., 2010, pp.116-117)

This early research suggests a more theoretical grounding. While theoretical understanding is indeed essential, the laboratory involves "vast information overload for students and, therefore, actual learning (in terms of understanding) is minimal" (Reid and Shah, 2007, p.179). The laboratory environment presents a curious mixture of knowledge being gained from both experience and being "borrowed from other people". Labdog can be seen as a tool to mix these two domains, by scaffolding and directing students' construction of pre-determined knowledge at pre-determined experiences. This fits within the broad principles of constructivism (Karagiorgi and Symeou, 2005), while still presenting a novel context to apply knowledge. As such, theory should not realistically be expected to predict or account for all possible variations in an inherently complex system.

With this in mind, I wish to examine a previously identified finding from this research. Within each of the twilight pilots a minority of students mentioned that Labdog slowed their working down. This is typified in one student's comment that they would prefer to continue the practical activity,"rather than paused to answer the questions". Or, as Elise, a SFY student, stated in a 2016/17 interview: Labdog can be "tedious at times... because you're trying to fill out Labdogs [questions] and steps, and sometimes it's just a waste of the quality of your time in the labs". In the second semester SFY survey in 2016/17, one student stated that there were too many "unnecessary steps, many of which were later removed". Though these do not represent the thought process of the majority of students, it is clear that students' cognitive resource or capacity can be overloaded, distracting them from the practical work.

Johnstone and Al-Shuaili (2001) believe a significant cause for overload can be attributed to the resources needed for students to perform procedures, and that "It is essential to establish the manipulative skills that they can 'go on auto-pilot'" (p.43). This idea builds on earlier research which established that students frequently respond to cognitively demanding activity by using shallow cognitive strategies, such as the recipe following (Johnstone and Wham, 1982). Especially for the Twilights, students would be performing new procedures with new equipment which involved relatively advanced concepts for them. Though SFY students may have been able to move the performance of procedures onto 'auto-pilot', this should not be expected of the Twilight students.

This draws attention to the importance of designing for novice-level chemists. Instructional design for novices needs to be explicit, providing direct guidance (Kirschner et al., 2006), as opposed to the use of project-, problem-, or inquiry-based laboratory instruction (Domin, 1999b; Polman, 2000; Hunnicutt et al., 2014). What works for experienced students is often detrimental to novices, largely due to the cognitive load required for novices to perform cognitive or procedural activities which may be assumed or taken for granted in more knowledgeable or advanced students (Kalyuga and Renkl, 2010). Educators need to acknowledge the cognitive demands placed on a student even when completing a relatively simple procedure, such as operating a rotary evaporator or handling hazardous materials. Labdog should not be confused for a tool which was intended to, or could, reduce the cognitive load associated with these tasks.

Previous researchers have argued that this difficulty rationalises the removal of laboratory-based work in education (Hawkes, 2004). Those researchers who have sought middle ground advocate for pre-lab activities as a tool to reduce cognitive load in the laboratory environment. (Carnduff and Reid, 2003), computer-based simulations (Kennepohl, 2007), or some combination of the two (Winberg and Berg, 2007).

I limit the advice I give here to iterative guidelines for improvement, as DBR defines improvements as evolutionary, not revolutionary - aiming for evidence-based iterative improvements in instruction (Reimann, 2011). From a pragmatic standpoint, iterative improvements and advice are more likely to help the average educator or instructional designer who cannot afford the resources necessary to revolutionise instruction, learning, and laboratory design. This research, therefore, looks at how Labdog, or other LaRS-like technologies, could be better integrated into the existing teaching environment.

Intrinsic load: the material being learned.

Complex concepts, e.g. those which are particularly large, inter-related, or nuanced, come with a larger intrinsic load. As such, students must expend considerable cognitive resources in handling, interpreting, consolidating, or applying these ideas. In order to manage the cognitive expenditure, educators may more heavily scaffold, i.e. support and structure (Rosenshine and Meister, 1992), students through learning activities. This may include educators providing components of an equation, or giving hints to relevant concepts or ideas.

The SFY laboratory instruction was carefully designed so that students would not be expected to explain any previously unseen chemical concepts or ideas. The SFY laboratory was always intended to demonstrate or apply concepts introduced previously in workshops or lectures, the so-called ‘confirmatory model’ of practical design (Abraham, 2011). This is not to say that such cognitive activities do not present intrinsic cognitive load: students are still expected to observe, interpret, and apply conceptual ideas (Domin, 1999b; Zimbadi et al., 2015).

Although several concepts may be almost universally difficult for students to grasp, individual students will have their own individual conceptual strengths and preferences. This may be impacted by a student’s prior knowledge (Kalyuga and Renkl, 2010), or quality of instruction (Tobias, 2010). Therefore, assuring that there is a sufficient student-educator ratio in the laboratory will allow such scaffolding to be provided on a one-on-one basis if needed. Conversely, providing such scaffolding within the instructional medium itself, e.g. integrating a series of hints within questions, could more easily provide such support to a greater number of students.

It will likely be easier for educators who are familiar with a particular student cohort, or with a greater familiarity with teaching a syllabus, to predict where the intrinsic load will be highest. Educators inexperienced in either of these regards could, therefore, benefit from tutelage or consultation with more experienced members of teaching staff.

While it is important to provide direct and clear instruction, caution needs to be taken against treating every student as if they were a novice. Kalyuga et al. (2003) present this idea of expertise reversal: "instructional guidance, which may be essential for novices, may have negative consequences for more experienced learners" (p.24). In the context of multimedia instruction, providing learners with animations broken down into individual steps was beneficial to novice learners but not necessarily expert ones (Spanjers et al., 2011). It appears that learners approaching new material require time and opportunity to engage with, and apply, their new knowledge in order to consolidate it (Wittwer and Renkl, 2008).

Pre-lab activities can be used to reduce the intrinsic cognitive load during a laboratory session, by ensuring students encounter the concepts before the laboratory (Reid and Shah, 2007). As (Van Merriënboer et al., 2003, p.9) state: "if the information is studied beforehand, a cognitive schema may be constructed in long-term memory that can subsequently be activated in working memory during task performance". In my experience with the SFY, I saw the more consistent implementation of pre-labs coincide with a shift towards relevant cognitive tasks to outside of the laboratory itself (Wilson and Read, 2017).

Extraneous Load: the use and interpretation of Labdog.

Unlike the previous points of discussion, extraneous cognitive load has a limited direct impact on learning, i.e. the retrieval or use of mental schema (Van Merriënboer, 2005). Extraneous cognitive load arises when students have to interpret "the instructional material used to present the content" (de Jong, 2010, p.106). Therefore this needs to be reduced through a combination of instructional design from educators, and user-design from a software perspective.

Unfortunately, the distinction of extraneous load can be difficult, as Paas et al. (2004) writes: “a cognitive load that is germane for a novice may be extraneous for an expert” (p.2). Sweller (1989) lists three principles of information presentation which can increase extraneous cognitive load:

1. **Split attention:** the cognitive resources required to simultaneously interpret two separate sources of information, e.g. audial and visual. Even if identical information presented in different formats, more cognitive load is required to interpret them (de Jong, 2010);
2. **Modality:** the resources required to combine information from multiple sources. This relates
3. **Redundancy:** the resources necessary to identify and negate identical information represented in multiple ways.

There are many ways to represent chemical concepts and processes, and moving between them can cause much difficulty for novice chemists (Johnstone, 2000; Talanquer, 2011; Taber, 2013). Presenting students with accurate and also easy to interpret representations of concepts is an important and complex instructional task. Previous research has looked at how traditional instruction, e.g. textbooks, have induced or exacerbated extraneous cognitive load. For example, Nyachwaya and Gillaspie (2016) examined the representations of chemical concepts in five general chemistry textbooks. An average of 86% (79-90%) of all diagrams were classified as ‘representational’, i.e. they “helped students understand associated concepts with explicitly provided links to the content and could be considered ‘instructionally useful’ representations.” (p.65). An overage of 8.6% (5-14%) of representations were purely decorative, however, which are more likely to add extraneous cognitive load (Cook, 2006). The authors also specifically identify and criticise diagrams which are not labeled, indexed, or referenced within the text, or those without an extended caption.

Like Labdog, textbooks can carefully order and structure the information presented to students, e.g. giving graphical, mathematical, textual explanations when they are needed or relevant. The ability of students to interpret this information and apply it to their context, e.g. observation and results, is heavily related to learning in the laboratory context (Domin, 1999a; Hofstein, 2004; Reid and Shah, 2007; Abrahams and Millar, 2008). However, identifying and isolating individual chemical principles can be complex, given the highly interrelated nature of chemistry (Taylor and Coll, 2001). Careful consideration, therefore, needs to be placed on both the relevant concepts, and how they are addressed before, during, and after the laboratory. The use of electronic lab notebooks (Van Dyke and Smith-Carpenter, 2017) or e-portfolio (Ring and Ramirez, 2012) are methods which some educators have used to address these concerns.

There is a need to filter the deluge of possible information which could be presented to students, lest it causes too much extraneous cognitive load. Educators should identify the most important concepts to be learned, including the most important procedures and concepts. The laboratory itself is cause for more concern than pre-labs, where students can more freely break or split-up work. As Labdog is simply a digital transcription of extant paper lab-scripts, educators who use it are no longer constrained by the space and single-media of paper. Educators are suddenly able to include more, and richer, learning resources. The information presented to students should not be repeated (Sweller, 1989), should have clear reason or benefit to students, and should be presented clearly and consistently (Mayer and Moreno, 2003).

Fortunately, no evidence for extraneous load in the use of Labdog emerged during the research, though none was explicitly looked for. This may be attributed to the fact that students were still given paper lab script during the labs. It could also be that the extant lab scripts were of relatively high quality, having benefited from annual review and revisions, being refined through ongoing experience with the SFY specifically.

7.2.6 Answering the Primary Research Question

Each of the above sub-questions can be combined to answer the overarching research question:

> “How does the integration and design-based research of Labdog, a novel web-based Laboratory Response System (LaRS), affect the student experience and perception of learning in and from introductory-level practical-based chemistry education?”

The numerous and varying data collected have strongly suggested that the use of Labdog has had a positive impact on students’ learning in the laboratory. Labdog has demonstrated itself as capable of ensuring that students engage with the underlying concepts related to practical work, and to facilitate formative feedback and interactions between educators and students in and during the laboratory. Labdog can, therefore, be seen as a successful attempt to prevent cookbook chemistry - a long-standing and diminutive experience on students’ learning from and in the laboratory environment

(Domin, 1999a; Johnstone and Al-Shuaili, 2001; Lunetta et al., 2007; Hofstein and Mamlok-Naaman, 2007). What is more, it does so while retaining the focus on closed-ended laboratory activities, which are prolific in entry-level chemistry (Domin, 1999b; Hofstein, 2004). This success has been made almost entirely possible by the freedom and iterative nature of design-based research, which prompted constant evaluation and redesigns over the three years from the conception of Labdog to the end of the year-long trial with SFY students.

In the following pages I will focus on the answers to this question from both parts: the use of design-based work, and the impact of Labdog on student experience and learning in the laboratory environment.

Design-based research and software development. Throughout this research, I have often presented Labdog in somewhat presumptive terms, i.e. taken its existence as assumed. Due to the scope and length limitations of a doctoral thesis document, I have dedicated little discussion to the technical aspects of this project simply because it is of largely separate to the educational impact of its use. I wish to briefly expand on this point and its implications for the broader process of research.

A more 'traditional' educational research project may concern itself with implementing and measuring the use of a pre-existing technology or instructional methodology. This kind of work has a valuable place in demonstrating the potential benefits of educational technology (Section 2.2.2). However, such research has gained a negative reputation for ignoring the importance of realistic and pragmatic improvements necessary to improve both software and teaching (Reeves et al., 2005).

The DBR paradigm has both facilitated and guided this research. DBR represents a relatively new, though growing, subset of education research and is faced with the expected growing pains as it finds a more integrated place within education research as a whole (Barab and Squire, 2004; Reimann, 2011). More carefully controlled and targeted research is feasible now that Labdog exists as a stable and accessible technology, as issues of scalability, usability, and sustainability become increasingly relevant (Fishman et al., 2004). Yet I cannot iterate strongly enough how such traditional approaches would have constrained or limited the result or product of this work (Herrington et al., 2007).

During the initial development of Labdog, I was faced with an educational problem with no clear existing solution. The result is both the specific generation of Labdog and the presentation of the broader idea of a LaRS. Doubtless something could have been constructed from existing tools, I imagine the use of VLE surveys or tests could provide a functional interface. By its nature, this research represents novel advancements in our understanding of the interaction between teaching, learning, and technology in the teaching laboratory environment.

Student perceived learning and experience in the laboratory. In interviews, UG chemistry students presented a range of purposes or benefits of the teaching laboratory. Students were split evenly between seeing the laboratory as a chance to develop conceptual understanding, or as a chance to see or witness chemical concepts. A number of students were also clear that the laboratory actively did not help them develop their conceptual understanding, e.g. due to social or anxiety-related reasons.

Interviews with eight SFY students after one semester of using Labdog suggest that Labdog helps students consciously engage with and thinking about the scientific principles behind their observations and actions. Similar findings come from the existing MLLI survey, which in which SFY student reported focusing on the underlying molecular actions of their actions. Custom evaluative surveys from the 2016/17 SFY cohort, as well as SFY students in 2015 and A-level students in 2016 and 2017, provide similar evidence. The results suggest that Labdog has often promoted students to consider the chemical concepts related to practical work. It is, therefore, reasonable to suggest that Labdog changes the student focus in the laboratory more towards the underlying scientific or chemical principles.

It does, however, make the laboratory more demanding, and if not integrated well, Labdog can contribute to cognitive overload and ultimately frustrate students and negatively impact the cognitive and affective benefits of the laboratory. SFY students reported that Labdog added to their extent workload in the laboratory. This was likely especially true when students had to complete some complex procedure, specifically if it was away from the workstation.

7.3 Future research

Discussion of the findings of this work has raised a number of interesting areas for future research. I have identified four relevant and recurring questions which could direct future work specifically. In this final section of the thesis, I present each question and a possible research design.

7.3.1 Can evidence of deep and meaningful learning be found with Labdog?

This work has made notable progress in identifying how the use of Labdog can affect students' learning processes, specifically having students more focused on the underlying chemical concepts. This needs to be supplemented with targeted and specific information on the impact of using Labdog on students' understanding. This, therefore, poses the initial question: how does the use of a LaRS-enhanced teaching laboratory affect students' understanding of the scientific or theoretical concepts related to laboratory content.

The need to isolate and then attribute changes in understanding present the need for a more experimental methodology. Such research would require a valid and reliable tool to measure student understanding. This creates a need for a series of questions which target discrete chemical concepts and are able to accurately record student (mis)understanding. This is necessary when making comparisons between, or within, groups of students. Using a pre-existing chemical concept inventory is a possibility, e.g. Mulford and Robinson (2002); Pavelich et al. (2004), though it is possible that a series of custom questions could be developed. However, such an undertaking represents a significant amount of work and complexity (Krause et al., 2004; Schwartz and Barbera, 2014). It is possible that an equivalent tool could be developed to assess students' procedural competency, perhaps taking from the growing work in and around e-portfolios or digital badging (Towns et al., 2015; Hensiek et al., 2017). Additionally, proxies and artifacts, e.g. students' workshop question sheets, in-class questions, or exam marks could be used as low-effort indicators of understanding.

The use of an experimental methodology, ideally a randomised control trial, would provide the most traditionally strong data for the impact of Labdog, or some other LaRS, on students' understanding. This involves working with two comparable groups, where one would use LaRS-enhanced laboratory instruction, and the other would not. Both groups would need to have relatively comparable demographic structure and prior levels of understanding and experience.

During the research, students would need to perform identical practical activities, work in similar laboratory environments, and have access to similar educational resources and material, including teaching/demonstrating staff. It may, therefore, be necessary to work with secondary or tertiary education environments - given a greater instructional consistency and potential sample size. This will more than likely involve working with under 18 year-olds, which raises a series of logistical barriers around consent and data protection, though these issues do not represent significant blockers to students' work.

Measurements of understanding could also be taken at varying time frames: e.g. after a practical, at the end of a semester, or at the end of an academic year. Taking a statistically rigorous approach to identifying the degree and consistency of difference in understanding will provide a much stronger quantitative argument for the impact of LaRS-enhanced instruction than I have been able to provide in this research. Unlike the research conducted so far, this proposed approach needs to acknowledge the loss of qualitative data. The experience of students and educators has provided valuable information into the reasons behind the success and failure of Labdog so far, and therefore future work should be carefully designed to not neglect such work.

7.3.2 What should Labdog questions ask, and when?

As discussed in Section 6.2, there are a number of different types of questions which we asked SFY students in 2016/17. This work was based largely on the assumption that linking the submicroscopic to the macroscopic, i.e. submicroscopic to the observable, is the most beneficial and unique advantage of using Labdog (Johnstone, 2000; Talanquer, 2011; Taber, 2013). Though grounded in long-standing theory, this remains an assumption. Developing a clearer understanding of how instruction can be designed to have students most successfully link observations to theory can produced theories that really work, presenting clear value to practicing educators, and continuing the themes of DBR (Reimann, 2011).

The wording and positioning of questions likely have significant impact on the development of students' understanding. I have provided a detailed discussion of how cognitive overload can enable or prevent learning in the lab, and how an understanding of flow state can be used to promote deeper and longer-lasting engagement in problem-solving. Though this research has provided strong evidence that Labdog can benefit students' learning, it cannot say with certainty how best to design instruction and questions.

This research has already produced a corpus of students' responses to questions - which Labdog stores in a database. Future research could start by examining how and when students provide the most correct and detailed responses. Such findings have clear implications in helping educators better precipitate misconceptions. The analysis could consider specifically where within a practical the questions are placed, what order they are presented in, and how many occur in a single step.

Analysis of existing data could be paired with investigations into the student experience of answering questions. Performing think-aloud protocols during the lab, i.e. having students report on something as they do it (Jääskeläinen, 2010), will provide insight into when or if cognitive overload, distraction, or confusion are most likely to occur. Retrospective and self-reported data, e.g. interviews or surveys, could provide both breadth and depth in insights.

There are a number of related questions to this work. For example, do the same principles of question design apply in pre- and post-laboratory environments? Alternatively, how, if at all, should the content and type of questions change with the concepts and procedures of the practical? For example, should laboratories with a greater focus on physical chemistry have fewer or more simple questions?

This research does not depend as heavily on the rigor of positivism as does the first proposal. The work would almost certainly have to adopt a mixed-methods approach: combining content analyses of students' responses, exploratory analysis of the types of questions, and qualitative information on the student experience. There are existing methodological instruments to measure both cognitive load, e.g. Paas and Ayres (2014), and flow state, e.g. Engeser and Rheinberg (2008). The research should, therefore, focus on effectively combining the existing methods, rather than creating new methodological tools or instruments.

7.3.3 Is Labdog a viable technology in secondary or tertiary education?

This research was fortunate enough to take place in a university setting, which provides relative freedom across the curriculum, instruction, and assessment. This early work required a great deal of trial and error, which likely would not be afforded in the constraints of a school environment, where standards and pressures are often dictated at a national level. However, secondary and tertiary educational environments represent the vast majority of novice-level chemistry students at a national and international level. Work into this context is therefore essential to affecting wider change to practice.

A future research project could adopt a case-study based approach within a small number of schools or colleges. This early work should demonstrate how, and if, Labdog would operate within the logistical constraints of a school or college. This is a continuation of the DBR concept that Labdog should be usable by real teachers, without excessive training or technical experience (Herrington et al., 2007).

The work would start by identifying a subset of schools and teachers which are willing ‘early adopters’, i.e. they have the goodwill and resources necessary to use Labdog in practice. It is likely the availability of technological infrastructure will form a barrier to participation in the school environment. The need for one or two digital devices per student pair, and need for reliable wireless internet, may be unrealistic requirements for many schools at present. Furthermore, there are data protection and security issues concerning individuals under 18 years-old.

The standardised syllabuses at these stages mean that practical activities in Labdog could be readily used across contexts. Once a practical activity has been designed effectively it can be shared freely among teachers and schools. This reduces the effort required for teachers to participate in the work. It would be advisable perhaps to only with a small number of selected practicals, at least in early work, so that Labdog can be used in contexts where it is most suited to the activity at hand. This would also present another context to identify and correct technical problems with Labdog.

The use of a case-study approach would allow the researcher to focus on the experiences of the individuals and the specific context - both teacher and student. It would prioritise and acknowledge the context at hand, without the need to immediately develop or focus on theoretical issues. This would be particularly important for acknowledging differences in technological infrastructure, and the nature of the prior technological experience of the students and educators. The research should not counteract these differences, but rather acknowledge them. The research should mix a focus on the student and teacher experiences, and impacts on teaching and learning. For this reason, I would advise a mixed-methods approach.

7.3.4 How could Labdog be a tool for summative assessment?

Throughout this work, Labdog has been used formatively, i.e. it has not contributed at all to students’ academic grades. However the data it collects, in students’ responses to questions, and their activity in pre- and during-lab environments, could be used for summative assessment. Labdog collects an evidence base of students’ actions and understanding, in a single workspace. Labdog therefore presents a unified workflow to both students and educators wishing to design, conduct, and assess practical work. This offers clear logistical benefits, as well as facilitating accountability and transparency in assessment, both internally and between schools and awarding bodies.

Recent changes[2] to UK A-level syllabuses have seen an increased focus on educators creating portfolios of students' actions and understanding along unified criteria. This 'practical endorsement' model could benefit very clearly from a system like Labdog, which educators could use to store and submit evidence that students have adequately performed some practical activity.

From a research perspective, this would involve the development of such functionality in the software. At the most basic level this would see the integration of summative marking and scoring into Labdog, and the ability to export such data in a user-friendly way, e.g. as a spreadsheet. More complex features would see re-working of the user interface to the e-portfolio section of Labdog, and making them accessible to both educators and students. A more complex demand would be the integration of assessment criteria, and related rubrics, into Labdog itself. This would allow educators to collate evidence from students, and identify it as evidence or justification for a certain practical mark. Not only would this simplify the workflow of educators, it would facilitate communication within schools and external qualification bodies when asked to justify their marks.

Furthermore, this could provide more data for research. Examining the consistency of educator marking within criteria, for example, could prove interesting. At a more conceptual level, this data could be used to critique or support the use of e-portfolio systems. For example, does the evidence collected and presented in Labdog provide valid and reliable evidence for what it claims to measure?

Just as this doctoral work has, this proposed project mixes software development with educational research, and will therefore likely rely on a less rigid methodology. This work has clearly benefited from using DBR, which integrates very well with contemporary software development practices (Conboy et al., 2015). As the project moves away from the development of software, it makes sense to adopt a case-study approach, working with specific schools and awarding bodies which show goodwill towards the project. It is important that the work accounts for the perspective of students, educators, school administration/management, and awarding bodies. For this reason, adopting an open and qualitative research method, allowing themes to emerge from each stakeholder organically, is essential. Such information can be used in the specification or development of the necessary new features, e.g. in creating formal requirements, as well as in the evaluation and discussion of the implementation of Labdog.

---

[2] As of 2016/17

Chapter 8

Online Appendix Instructions

Due to the large volume, and digital-nature, of data used in this research a printed appendix would be both cumbersome and inaccessible to the reader. As such, an online repository of research data has been prepared, which can be accessed on Pure[1] the University of Southampton's online data repository. Due to the sensitive nature of the data, e.g. details of students, the dataset must be requested, for which one should provide both the ID (**33093342**) and a UUID (**eca9c6db-23a8-41d6-b8b8-165672789507**). This repository contains the following data, which I have grouped by type, as opposed to chronologically.

A note on file formats and software recommendations

I have used a mixture of open standard and proprietary file formats throughout this research. Where open formats are accessible through multiple pieces of software, proprietary file formats are designed exclusively for use with a single piece of software.

1. `.m4a` are open standard audio recordings, which I have used for my interview recordings, and can be opened using any media player.
2. `.md` are text files formatted using markdown, an open standard for structuring text documents.[2]

---

[1] `https://pure.soton.ac.uk`

[2] `https://en.wikipedia.org/wiki/Markdown`

3. `.docx` or `.doc` is a proprietary word-processing file formats for which are designed for use with Microsoft Office[3], though are accessible through alternative open-source software, e.g. Open Office[4].
4. `.nvpx` is a proprietary file format used by the qualitative data analysis software NVivo, and can only be opened through their software, which is available for both Mac and Windows[5].
5. `.xlsx` or `.xls` are proprietary formats for numerical spreadsheets which are accessible using Microsoft Office, as well as open source alternatives such as open office.
6. `.csv` is an open file standard for storing simple numerical data, and can be accessed easily using any common spreadsheet software, as well as programming language.
7. `.ipynb` are analysis notebooks for analysis conducted in the Python programming language[6], using the Jupyter Notebook[7] format to record the process and results of exploration and visualisation.
8. `.R` are files written in, and for use by, R, the statistical programming language[8] which I have used for both data analysis and visualisation,
9. `.RProj` represent data stored by R-Studio[9], an open source piece of software for data exploration, analysis, and visualisation in R. These files contain environmental variables, such as the results from analysis, custom functions, and visualisations.
10. `.RData` are logs of activity within the R console, used for data analysis and visualisation.

Interviews

I conducted two major rounds of interviews over the research: with undergraduates on the historic role of the laboratory (Section 3.2), and with science foundation year students (Section 5.3.2).

---

[3] `https://www.microsoft.com/en-gb/store/b/office?icid=CNavSoftwareOffice`
[4] `https://www.openoffice.org/`
[5] `http://www.qsrinternational.com/nvivo/nvivo-products`
[6] `https://www.python.org/`
[7] `http://jupyter.org/`
[8] `https://www.r-project.org/about.html`
[9] `https://www.rstudio.com/`

The appendix includes protocols for each of the interviews, as word processing documents.

As part of these interviews I collected audio data of the interviews themselves, transcribed them to text files (ultimately formatted to Microsoft Word documents), and analysed them using NVivo. Files from each of these stages are available.

Survey

I used surveys at three major points throughout this research: a custom evaluation with A-level students following their use of Labdog in the Twilight sessions (throughout Chapter 4), a custom evaluation survey with science foundation year students following their use of Labdog in the 2016/17 academic year (Section 5.3), and the use of a pre-existing quantitative survey: the meaningful learning in the laboratory instrument (6.1).

Where surveys were given on paper (foundation year completion of the meaningful learning in the laboratory survey, and foundation year evaluation), the appendix includes copies of the documents given to the students. Where surveys were exclusively electronic, the questions are recorded in the raw results files.

For each, the raw data is available as simple table-formatted spreadsheets (largely as `.csv` but also as `.xlsx`). For each of these surveys there is also a related R-Studio project or Jupyter notebook - both of these formats record and relate the processes and functions for visualisation

Labdog data

There is a large volume and range of data which was collected and stored on the Labdog database, which was a SQL relational database[10]. I have included both cleaned data, i.e. processed and joined, and raw data.

---

[10] `https://www.mysql.com/`

Cleaned Data

Cleaned data is that which I have collated and filtered in a meaningful way. I present cleaned data for:

1. **Responses to questions** - showing when students have responded to which questions, including the question, the response, and details of the step.
2. **Response counts** - numerical analysis of the number of responses
3. **Coding of Responses** - which represents content-analysis for students' responses to questions.

Raw Data

I have downloaded the raw data on a per-lab session basis, with one `.csv` file per session. These have been clearly labelled by number, and a prefix used to denote the content of each, which relates to the name of the database. These are as follows:

1. `pls` - **Pre-Lab Sessions**: where each row represents a pre-lab session;
2. `qnr` - **Questions Need Response**: where each row represents an instance where an educator or demonstrator has rejected a student response to a question.
3. `sc` - **Step Completes**: where each row represents the movement of one student to one step in a practical to another
4. `sqr` - **Step Question Responses**: where each row represents a students' response to a question
5. `ss` - **Student Sessions**: where each row represents the session which links an individual student to completing a practical
6. `steps` - **Practical steps**: where each row represents a steps in a practical (not delimited by lab session).

References

M. Abraham. What can be learned from laboratory activities? revisiting 32 years of research. *Journal of Chemical Education*, pages 1020–1025, 2011.

M. R. Abraham, M. S. Craolice, A. P. Graves, A. H. Aldhamash, J. G. Kihega, J. G. P. Gal, and V. Varghese. The nature and state of general chemistry laboratory courses offered by colleges and universities in the united states. *Journal of Chemical Education*, 74(5):591, 1997.

I. Abrahams. Does practical work really motivate? a study of the affective value of practical work in secondary school science. *International journal of science education*, 31(17):2335–2353, 2009.

I. Abrahams and R. Millar. Does practical work really work? a study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14):1945–1969, 2008.

I. Abrahams, M. J. Reiss, and R. M. Sharpe. The assessment of practical work in school science. *Studies in Science Education*, 49(2):209–251, 2013.

S. Abuhamdeh and M. Csikszentmihalyi. The importance of challenge for the enjoyment of intrinsically motivated, goal-directed activities. *Personality and Social Psychology Bulletin*, 38(3):317–330, 2012.

S. Abuhamdeh, J. Nakamura, and M. Csikszentmihalyi. Flow. In *Handbook of competence and motivation*, pages 598–608. Guilford Publications, NY, 2005.

D. J. Adams. Current trends in laboratory class teaching in university bioscience programmes. *Bioscience Education*, 13(1):1–14, 2009.

P. Adams. Exploring social constructivism: Theories and practicalities. *Education*, 34 (3):243–257, 2006.

S. Adams Becker, V. Estrada, A. Freeman, and L. Johnson. Nmc horizon report: 2014 higher education edition. *Austin, Texas: The New Media Consortium*, 2014.

T. Alkhaldi, I. Pranata, and R. I. Athauda. A review of contemporary virtual and remote laboratory implementations: observations and findings. *Journal of Computers in Education*, 3(3):329–351, 2016.

A. W. Amick and N. Cross. An almost paperless organic chemistry course with the use of ipads. *Journal of Chemical Education*, 91:753–756, 2014.

T. Amiel, T. C. Reeves, et al. Design-based research and educational technology: Rethinking technology and the research agenda. *Educational Technology & Society*, 11(4):29–40, 2008.

T. Anderson and J. Shattuck. Design-based research: A decade of progress in education research? *Educational researcher*, 41(1):16–25, 2012.

A. Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.

A. Baddeley, M. Eysenck, and M. Anderson. Memory, 2009.

A. Bandura. The self system in reciprocal determinism. *American psychologist*, 33(4): 344, 1978.

A. Bandura. *Social foundations of thought and action.* Englewood Cliffs, NJ Prentice Hall., 1986.

S. Barab and K. Squire. Design-based research: Putting a stake in the ground. *The journal of the learning sciences*, 13(1):1–14, 2004.

M. Bassi and A. Delle Fave. Optimal experience and self-determination at school: Joining perspectives. *Motivation and emotion*, 36(4):425–438, 2012.

I. D. Beatty and W. J. Gerace. Technology-enhanced formative assessment: A research-based pedagogy for teaching science with classroom response technology. *Journal of Science Education and Technology*, 18(2):146–162, 2009.

I. D. Beatty, W. J. Leonard, W. J. Gerace, and R. J. Dufresne. Question driven instruction: Teaching science. *Audience response systems in higher education: Applications and cases*, page 96, 2006.

I. D. Beatty, A. Feldman, W. J. Leonard, W. J. Gerace, K. S. Cyr, H. Lee, and R. Harris. Teacher learning of technology-enhanced formative assessment. *arXiv preprint arXiv:0806.4924*, 2008.

H. Bembenutty. Three essential components of college teaching: Achievement calibration, self-efficacy, and self-regulation. *College Student Journal*, 43(2):562, 2009.

J. Bennett. *Teaching and learning science: A guide to recent research and its applications.* A&C Black, 2005.

J. Bennett and H. E. Pence. Managing laboratory data using cloud computing as an organizational tool. *Journal of Chemical Education*, 88(6):761–763, 2011.

R. E. Bennett. Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1):5–25, 2011.

S. W. Bennett and H. Wilkins. Assessment in chemistry and the role of examinations. *University Chemistry Education*, 8(2):52–57, 2004.

J. Bergmann and A. Sams. *Flip your classroom: Reach every student in every class every day.* International Society for Technology in Education, 2012.

D. Berrett. How ‘flipping’the classroom can improve the traditional lecture. *The chronicle of higher education*, 12, 2012.

J. L. Bishop and M. A. Verleger. The flipped classroom: A survey of the research. In *ASEE National Conference Proceedings, Atlanta, GA*, 2013.

R. A. Bjork, J. Dunlosky, and N. Kornell. Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64:417–444, 2013.

P. Black and D. Wiliam. *Inside the black box: Raising standards through classroom assessment.* Granada Learning, 1998.

P. Black and D. Wiliam. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1):5–31, 2009.

P. Black, C. Harrison, C. Lee, and B. Marshall. *Assessment for learning: Putting it into practice.* McGraw-Hill International, 2003.

L. Blasco-Arcas, I. Buil, B. Hernández-Ortega, and F. J. Sese. Using clickers in class. the role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education*, 62:102–110, 2013.

D. J. Blick. The purpose and character of laboratory instruction. *Journal of Chemical Education*, 32(5):264, 1955.

A.-M. Bliuc, P. Goodyear, and R. A. Ellis. Research focus and methodological choices in studies into students' experiences of blended learning in higher education. *The Internet and Higher Education*, 10(4):231–244, 2007.

G. M. Bodner. Problem solving: the difference between what we do and what we tell students to do. *University Chemistry Education*, 7(7):37–45, 2003.

G. M. Bodner and D. S. Domin. Mental models: The role of representations in problem solving in chemistry. *University Chemistry Education*, 4(1), 2000.

C. J. Bonk and C. R. Graham. *The handbook of blended learning: Global perspectives, local designs*. John Wiley & Sons, 2012.

J. G. Borkowski, M. Carr, E. Rellinger, M. Pressley, et al. Self-regulated cognition: Interdependence of metacognition, attributions, and self-esteem. *Dimensions of thinking and cognitive instruction*, 1:53–92, 1990.

D. J. Boudah. *Conducting educational research: Guide to completing a major project*. Sage, 2010.

J. D. Bransford, A. L. Brown, R. R. Cocking, et al. How people learn, 2000.

V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

S. L. Bretz. Novak's theory of education: Human constructivism and meaningful learning. *Journal of Chemical Education*, 78(8):1107, 2001.

S. L. Bretz, M. Fay, L. B. Bruck, and M. H. Towns. What faculty interviews reveal about meaningful learning in the undergraduate chemistry laboratory. *Journal of Chemical Education*, 90(3):281–288, 2013.

A. Brown. Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, motivation, and understanding*, pages 65–116, 1987.

A. L. Brown. Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences*, 2 (2):141–178, 1992.

P. C. Brown, H. L. Roediger III, and M. A. McDaniel. *Make it stick*. Harvard University Press, 2014.

R. C. D. Brown, J. D. Hinks, and D. Read. A blended-learning approach to supporting students in organic chemistry: methodology and outcomes. *New Directions*, (8):33–37, 2012.

L. B. Bruck, M. Towns, and S. L. Bretz. Faculty perspectives of undergraduate chemistry laboratory: Goals and obstacles to success. *Journal of Chemical Education*, 87(12): 1416–1424, 2010.

R. Brunken, J. L. Plass, and D. Leutner. Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1):53–61, 2003.

A. Bryman. Integrating quantitative and qualitative research: how is it done? *Qualitative research*, 6(1):97–113, 2006.

S. Bull and J. Kay. Open learner models. In *Advances in intelligent tutoring systems*, pages 301–322. Springer, 2010.

P. Burnard, P. Gill, K. Stewart, E. Treasure, and B. Chadwick. Analysing and presenting qualitative data. *British dental journal*, 204(8):429–432, 2008.

A. Burrows, J. Holman, A. Parsons, G. Pilling, and G. Price. *Chemistry*^{3}*: Introducing Inorganic, Organic and Physical Chemistry.* Oxford University Press, 2013.

C. Camerer, G. Loewenstein, and M. Weber. The curse of knowledge in economic settings: An experimental analysis. *The Journal of Political Economy*, pages 1232–1254, 1989.

W. S. Carlsen, S. Abell, and N. Lederman. Language and science learning. *Handbook of research on science education*, pages 57–74, 2007.

J. Carnduff and N. Reid. *Enhancing undergraduate chemistry laboratories: pre-laboratory and post-laboratory exercises.* Royal Society of Chemistry, 2003.

S. Cassidy. Self-regulated learning in higher education: Identifying key component processes. *Studies in Higher Education*, 36(8):989–1000, 2011.

L.-J. ChanLin. Technology integration applied to project-based learning in science. *Innovations in Education and Teaching International*, 45(1):55–65, 2008.

C.-H. Chen. Why do teachers not practice what they believe regarding technology integration? *The Journal of Educational Research*, 102(1):65–75, 2008.

I. Chopra, J. O'Connor, R. Pancho, M. Chrzanowski, and S. Sandi-Urena. Reform in a general chemistry laboratory: how do students experience change in the instructional approach? *Chemistry Education Research and Practice*, 18(1):113–126, 2017.

C. Christensson and J. Sjöström. Chemistry in context: analysis of thematic chemistry videos available online. *Chemistry Education Research and Practice*, 15(1):59–69, 2014.

I. Clark. Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2):205–249, 2012.

P. Cobb, J. Confrey, A. DiSessa, R. Lehrer, and L. Schauble. Design experiments in educational research. *Educational researcher*, 32(1):9–13, 2003.

L. Cohen, L. Manion, and K. Morrison. *Research methods in education*. Routledge, 2013.

A. Collins. Toward a design science of education. In *New directions in educational technology*, pages 15–22. Springer, 1992.

K. Conboy, R. Gleasure, and E. Cullina. Agile design science research. In *New Horizons in Design Science: Broadening the Research Agenda*, pages 168–180. Springer, 2015.

E. Cook, E. Kennedy, and S. Y. McGuire. Effect of teaching metacognitive learning strategies on performance in general chemistry courses. *Journal of Chemical Education*, 90(8):961–967, 2013.

M. Cook. Visual representations in science education: the influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, 90 (6):1073–1091, 2006.

B. Coştu. Comparison of students' performance on algorithmic, conceptual and graphical chemistry gas problems. *Journal of Science Education and Technology*, 16(5):379–386, 2007.

N. Cowan. The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, 19(1):51–57, 2010.

J. W. Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2013.

J. W. Creswell and V. L. P. Clark. Designing and conducting mixed methods research. 2007.

C. H. Crouch and E. Mazur. Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9):970–977, 2001.

M. Csikszentmihalyi. *Beyond boredom and anxiety.* Jossey-Bass, 2000.

M. Csikszentmihalyi. Attention and the holistic approach to behavior. In *Flow and the Foundations of Positive Psychology*, pages 1–20. Springer, 2014.

M. Csikszentmihalyi and I. S. Csikszentmihalyi. *Optimal experience: Psychological studies of flow in consciousness.* Cambridge university press, 1992.

T. de Jong. Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, 38(2):105–134, 2010.

C. Deacon, A. Hajek, and H. Schulz. Graduate teaching assistants' perceptions of teaching competencies required for work in undergraduate science labs. *International Journal of Science Education*, pages 1–20, 2017.

T. K. DeBacker, H. M. Crowson, A. D. Beesley, S. J. Thoma, and N. L. Hestevold. The challenge of measuring epistemic beliefs: An analysis of three self-report instruments. *The Journal of Experimental Education*, 76(3):281–312, 2008.

D. W. Denton. Enhancing instruction through constructivism, cooperative learning, and cloud computing. *TechTrends*, 56(4):34–41, 2012.

S. Deterding, D. Dixon, R. Khaled, and L. Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.

J. Dewey. *Logic, the theory of inquiry.* New York: Holt and Co., 1938.

D. L. Dinsmore and M. M. Parkinson. What are confidence judgments made of? students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, 24:4–14, 2013.

D. S. Domin. A content analysis of general chemistry laboratory manuals for evidence of higher-order cognitive tasks. *Journal of Chemical Education*, 76(1):109, 1999a.

D. S. Domin. A review of laboratory instruction styles. *Journal of chemical education*, 76(4):543, 1999b.

R. J. Dufresne and W. J. Gerace. Assessing-to-learn: Formative assessment in physics instruction. *The Physics Teacher*, 42(7):428–433, 2004.

J. Dunlosky and T. Nelson. Importance of the kind of cue for judgements of learning (jol) and the delayed-jol effect. *Memory and Cognition*, 20(4):374–380, 1992.

M. W. Easterday, D. G. Rees Lewis, and E. M. Gerber. The logic of the theoretical and practical products of design research. *Australasian Journal of Educational Technology*, 32(4):125–144, 2016.

J. S. Eccles and A. Wigfield. Motivational beliefs, values, and goals. *Annual review of psychology*, 53(1):109–132, 2002.

A. J. Edwards. Preface. *Contemporary Educational Psychology*, 11(4):305 –, 1986. ISSN 0361-476X. doi: http://dx.doi.org/10.1016/0361-476X(86)90025-1. URL http://www.sciencedirect.com/science/article/pii/0361476X86900251.

J. Egbert. A study of flow theory in the foreign language classroom. *Canadian modern language review*, 60(5):549–586, 2004.

I. Eilks and B. Byers. The need for innovative methods of teaching and learning chemistry in higher education–reflections from a project of the european chemistry thematic network. *Chemistry Education Research and Practice*, 11(4):233–240, 2010.

S. Engeser and F. Rheinberg. Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3):158–172, 2008.

C. Erzberger and G. Prein. Triangulation: Validity and empirically-based hypothesis construction. *Quality and Quantity*, 31(2):141–154, 1997.

A. P. Fagen, C. H. Crouch, and E. Mazur. Peer instruction: Results from a range of classrooms. *The Physics Teacher*, 40(4):206–209, 2002.

B. Fishman, R. W. Marx, P. Blumenfeld, J. Krajcik, and E. Soloway. Creating a framework for research on systemic technology innovations. *The Journal of the Learning Sciences*, 13(1):43–76, 2004.

S. T. Fiske, J. G. Cromley, T. Perez, and A. Kaplan. Undergraduate stem achievement and retention: Cognitive, motivational, and institutional factors and solutions. *Policy Insights from the Behavioral and Brain Sciences*, 3(1):4–11, 2016.

J. H. Flavell. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10):906, 1979.

V. J. Flood, F. G. Amar, R. Nemirovsky, B. W. Harrer, M. R. Bruce, and M. C. Wittmann. Paying attention to gesture when students talk chemistry: Interactional resources for responsive teaching. *Journal of Chemical Education*, 92(1):11–22, 2014.

C. J. Fong, D. J. Zaleski, and J. K. Leach. The challenge–skill balance and antecedents of flow: A meta-analytic investigation. *The Journal of Positive Psychology*, 10(5): 425–446, 2015.

M. Ford. Disciplinary authority and accountability in scientific practice and learning. *Science Education*, 92(3):404–423, 2008.

J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris. School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1):59–109, 2004.

S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

C. Frith and U. Frith. Theory of mind. *Current Biology*, 15(17):R644–R645, 2005.

K. Fulton. Upside down and inside out: Flip your classroom to improve student learning. *Learning & Leading with Technology*, 39(8):12–17, 2012.

P. I. Fusch and L. R. Ness. Are we there yet? data saturation in qualitative research. *The Qualitative Report*, 20(9):1408, 2015.

K. R. Galloway and S. L. Bretz. Development of an assessment tool to measure students' meaningful learning in the undergraduate chemistry laboratory. *Journal of Chemical Education*, 92(7):1149–1158, 2015.

K. R. Galloway and S. L. Bretz. Video episodes and action cameras in the undergraduate chemistry laboratory: eliciting student perceptions of meaningful learning. *Chemistry Education Research and Practice*, 17(1):139–155, 2016.

K. R. Galloway, Z. Malakpa, and S. L. Bretz. Investigating affective experiences in the undergraduate chemistry laboratory: Students' perceptions of control and responsibility. *Journal of Chemical Education*, 93(2):227–238, 2015.

F. J. García-Peñalvo. What computational thinking is. *Journal of Technology Research*, 9:5–8, 2016.

D. R. Garrison and H. Kanuka. Blended learning: Uncovering its transformative potential in higher education. *The internet and higher education*, 7(2):95–105, 2004.

J. W. Gaynor and D. Brown. An online booking system encourages self-directed learning and personalization of study. *Journal of Chemical Education*, 89(8):1019–1024, 2012.

D. Gibson, N. Ostashewski, K. Flintoff, S. Grant, and E. Knight. Digital badges in education. *Education and Information Technologies*, 20(2):403–410, 2015.

B. Glaser and A. Strauss. The discovery of grounded theory: strategies for qualitative research. 1967.

B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: Strategies for qualitative research.* Transaction Publishers, 2009.

S. M. Glynn, P. Brickman, N. Armstrong, and G. Taasoobshirazi. Science motivation questionnaire ii: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, 48(10):1159–1176, 2011.

J. A. González-Martínez, M. L. Bote-Lorenzo, E. Gómez-Sánchez, and R. Cano-Parra. Cloud computing and education: A state-of-the-art survey. *Computers & Education*, 80:132–151, 2015.

J. I. Goodlad. A study of schooling: Some findings and hypotheses. *Phi Delta Kappan*, 64(7):465–70, 1983.

G. Goodman. *The Educational Psychology Reader*, chapter Critical Thinking: How Good Questions Affect Classrooms, pages 3–13. Peter Lang, 2014.

S. Gorard and C. Taylor. *Combining methods in educational and social research.* McGraw-Hill Education (UK), 2004.

K. J. Graham, B. J. Johnson, T. N. Jones, E. J. McIntee, and C. P. Schaller. Designing and conducting a purification scheme as an organic chemistry laboratory practical. *J. Chem. Educ*, 85(12):1644, 2008.

L. Grant and S. Jenkins. Practical skills of new undergraduates: Report on research workshops delivered on behalf of the gatsby charitable foundation. 2011.

C. Grbich. *Qualitative data analysis: An introduction.* Sage, 2012.

N. P. Grove and S. L. Bretz. A continuum of learning: from rote memorization to meaningful learning in organic chemistry. *Chemistry Education Research and Practice*, 13(3):201–208, 2012.

S. Guri-Rosenblit. ‘distance education’and ‘e-learning’: Not the same thing. *Higher education*, 49(4):467–493, 2005.

B. D. Haig. Grounded theory as scientific method. *Philosophy of education*, 1:1–1, 1995.

B. Harks, K. Rakoczy, J. Hattie, M. Besser, and E. Klieme. The effects of feedback on achievement, interest and self-evaluation: the role of feedback’s perceived usefulness. *Educational Psychology*, 34(3):269–290, 2014.

D. J. Harris, S. J. Vine, and M. R. Wilson. Is flow really effortless? the complex role of effortful attention. *Sport, Exercise, and Performance Psychology*, 6(1):103, 2017.

J. Hattie and H. Timperley. The power of feedback. *Review of educational research*, 77 (1):81–112, 2007.

S. J. Hawkes. Chemistry is not a laboratory science. *Journal of Chemical Education*, 81 (9):1257, 2004.

J. M. Hektner and M. Csikszentmihalyi. A longitudinal exploration of flow and intrinsic motivation in adolescents. 1996.

S. Hensiek, B. K. DeKorver, C. J. Harwood, J. Fish, K. O’Shea, and M. Towns. Digital badges in science: A novel approach to the assessment of student learning. *Journal of College Science Teaching*, 46(3):28, 2017.

C. F. Herreid and N. A. Schiller. Case studies and the flipped classroom. *Journal of College Science Teaching*, 42(5):62–66, 2013.

J. Herrington, S. McKenney, T. Reeves, and R. Oliver. Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. 2007.

D. Hestenes, M. Wells, and G. Swackhamer. Force concept inventory. *The physics teacher*, 30(3):141–158, 1992.

L. A. Hickman and T. M. Alexander. The essential dewey, volume 1: Pragmatism, education, democracy. 1998.

A. Hofstein. The laboratory in chemistry education: Thirty years of experience with developments, implementation, and research. *Chemistry Education Research and Practice*, 5(3):247–264, 2004.

A. Hofstein and V. N. Lunetta. The role of the laboratory in science teaching: Neglected aspects of research. *Review of educational research*, 52(2):201–217, 1982.

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

A. Hofstein and R. Mamlok-Naaman. The laboratory in science education: the state of the art. *Chemistry education research and practice*, 8(2):105–107, 2007.

A. Hofstein, M. Kipnis, and I. Abrahams. How to learn in and from the chemistry laboratory. In *Teaching chemistry–A studybook*, pages 153–182. Springer, 2013.

House of Commons Science and Technology Committee. Third report. science education from 14 to 19. 2002.

K. R. Howe. Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational researcher*, 17(8):10–16, 1988.

S. S. Hunnicutt, A. Grushow, and R. Whitnell. Guided-inquiry experiments for physical chemistry: the pogil-pcl model. *Journal of Chemical Education*, 92(2):262–268, 2014.

E. Hutchins. *Cognition in the Wild*, volume 262082314. MIT press Cambridge, MA, 1995.

R. Jääskeläinen. Think-aloud protocol. *Handbook of translation studies*, 1:371–373, 2010.

B. Johnson and L. Christensen. *Educational research: Quantitative, qualitative, and mixed approaches*. Sage, 2008.

R. B. Johnson and A. J. Onwuegbuzie. Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7):14–26, 2004.

A. Johnstone and A. Al-Shuaili. Learning in the laboratory; some thoughts from the literature. *University Chemistry Education*, 5(2):42–51, 2001.

A. Johnstone and H. El-Banna. Capacities, demands and processes–a predictive model for science education. *Education in chemistry*, 23(3):80–84, 1986.

A. H. Johnstone. Macro and microchemistry, 1982.

A. H. Johnstone. Why is science difficult to learn? things are seldom what they seem. *Journal of computer assisted learning*, 7(2):75–83, 1991.

A. H. Johnstone. Teaching of chemistry-logical or psychological? *Chemistry Education Research and Practice*, 1(1):9–15, 2000.

A. H. Johnstone and A. Wham. The demands of practical work. *Education in chemistry*, 19(3):71–73, 1982.

D. F. Jolley, S. R. Wilson, C. Kelso, G. O'Brien, and C. E. Mason. Analytical thinking, analytical action: Using prelab video demonstrations and e-quizzes to improve undergraduate preparedness for analytical chemistry practical classes. *Journal of Chemical Education*, 93(11):1855–1862, 2016.

S. Kalyuga and A. Renkl. Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, 38(3):209–215, 2010.

S. Kalyuga, P. Ayres, P. Chandler, and J. Sweller. The expertise reversal effect. *Educational psychologist*, 38(1):23–31, 2003.

Y. Karagiorgi and L. Symeou. Translating constructivism into instructional design: Potential and limitations. *Educational Technology & Society*, 8(1):17–27, 2005.

J. D. Karpicke and P. J. Grimaldi. Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, 24(3):401–418, 2012.

D. Katchevich, A. Hofstein, and R. Mamlok-Naaman. Argumentation in the chemistry laboratory: Inquiry and confirmatory experiments. *Research in Science Education*, 43(1):317–345, 2013.

J. C. Kaufman. For whom the words flow: A review of writing in flow. *The Journal of Creative Behavior*, 36(3):221–224, 2002.

C. M. Kelley and L. L. Jacoby. Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and language*, 35(2):157–175, 1996.

D. Kennepohl. Using computer simulations to supplement teaching laboratories in chemistry for distance delivery. *International Journal of E-Learning & Distance Education*, 16(2):58–65, 2007.

E. Kim and E. H. Seo. The relationship of flow and self-regulated learning to active procrastination. *Social Behavior and Personality: an international journal*, 41(7): 1099–1113, 2013.

P. A. Kirschner, J. Sweller, and R. E. Clark. 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, 2006.

N. Kornell, P. J. Klein, and K. A. Rawson. Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(1):283, 2015.

D. R. Krathwohl. A revision of bloom's taxonomy: An overview. *Theory into practice*, 41(4):212–218, 2002.

S. Krause, J. Birk, R. Bauer, B. Jenkins, and M. J. Pavelich. Development, testing, and application of a chemistry concept inventory. In *Frontiers in Education, 2004. FIE 2004. 34th Annual*, pages T1G–1. IEEE, 2004.

J. Kruger and D. Dunning. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121, 1999.

G. L. Lavigne, R. J. Vallerand, and P. Miquelon. A motivational model of persistence in science education: A self-determination theory approach. *European Journal of Psychology of Education*, 22(3):351, 2007.

C.-Y. Law, J. Grundy, A. Cain, R. Vasa, and A. Cummaudo. User perceptions of using an open learner model visualisation tool for facilitating self-regulated learning. pages 55–64. ACE, 2017.

Y. Le Corff, V. Gingras, and M. Busque-Carrier. Equivalence of unproctored internet testing and proctored paper-and-pencil testing of the big five. *International Journal of Selection and Assessment*, 25(2):154–160, 2017.

R. Levinson. *Teaching science.* Routledge, 2005.

C. Lewis. Driving factors for e-learning: An organisational perspective. *Perspectives: Policy & Practice in Higher Education*, 6(2):50–54, 2002.

M. Lichtman. *Qualitative Research in Education. A user's guide. Third Edition.* SAGE publisher, 2013.

C. P. Lim, D. Nonis, and J. Hedberg. Gaming in a 3d multiuser virtual environment: Engaging students in science lessons. *British Journal of Educational Technology*, 37 (2):211–231, 2006.

J. Linnansaari, J. Viljaranta, J. Lavonen, B. Schneider, and K. Salmela-Aro. Finnish students engagement in science lessons. 2015.

F. Lockwood. *Open and distance learning today.* Routledge, 2013.

V. N. Lunetta, A. Hofstein, and M. P. Clough. Learning and teaching in the school science laboratory: An analysis of research, theory, and practice. *Handbook of research on science education*, pages 393–441, 2007.

J. R. MacArthur and L. L. Jones. A review of literature reports of clickers applicable to college chemistry classrooms. *Chemistry Education Research and Practice*, 9(3): 187–195, 2008.

J. A. Mangels, B. Butterfield, J. Lamb, C. Good, and C. S. Dweck. Why do beliefs about intelligence influence learning success? a social cognitive neuroscience model. *Social cognitive and affective neuroscience*, 1(2):75–86, 2006.

S. Markic and I. Eilks. A case study on german first year chemistry student teachers beliefs about chemistry teaching, and their comparison with student teachers from other science teaching domains. *Chemistry Education Research and Practice*, 9(1): 25–34, 2008.

R. C. Martin. *Agile software development: principles, patterns, and practices.* Prentice Hall PTR, 2003.

R. E. Mayer. The promise of multimedia learning: using the same instructional design methods across different media. *Learning and instruction*, 13(2):125–139, 2003.

R. E. Mayer and R. Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1):43–52, 2003.

R. E. Mayer, A. Stull, K. DeLeeuw, K. Almeroth, B. Bimber, D. Chun, M. Bulger, J. Campbell, A. Knight, and H. Zhang. Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary educational psychology*, 34(1):51–57, 2009.

S. McClean, K. G. McCartan, S. Meskin, B. Gorges, and W. P. Hagan. Reflections on "youtesttube. com": An online video-sharing platform to engage students with chemistry laboratory classes. *Journal of Chemical Education*, 93(11):1863–1870, 2016.

R. McDaniel and J. R. Fanfarelli. A digital badging dataset focused on performance, engagement and behavior-related variables from observations in web-based university courses. *British Journal of Educational Technology*, 46(5):937–941, 2015.

D. McDonald, A. MacDonald, and C. Breslin. Final report from the jisc review of the environmental and organisational implications of cloud computing in higher and further education. *University of Strathclyde and JISC*, 2010.

B. Mehlenbacher, M. K. Autry, and A. R. Kelly. Instructional design for stem-based collaborative, colocated classroom composition. *IEEE Transactions on Professional Communication*, 58(4):396–409, 2015.

M. B. Miles and A. M. Huberman. *Qualitative data analysis: An expanded sourcebook*. sage, 1994.

R. Millar and I. Abrahams. Practical work: making it more effective. *School Science Review*, 91(334):59–64, 2009.

B. Miri, B.-C. David, and Z. Uri. Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in science education*, 37(4): 353–369, 2007.

A. C. Moller, B. P. Meier, and R. D. Wall. Developing an experimental induction of flow: Effortless action in the lab. *Effortless attention: A new perspective in the cognitive science of attention and action*, pages 191–204, 2010.

J. T. Morgan and C. Wakefield. Who benefits from peer conversation? examining correlations of clicker question correctness and course performance. *Journal of College Science Teaching*, 41(5):51, 2012.

Mozilla. Open badges, 2017. URL `https://openbadges.org/`.

K. R. Muis. The role of epistemic beliefs in self-regulated learning. *Educational Psychologist*, 42(3):173–190, 2007.

K. R. Muis and G. M. Franco. Epistemic beliefs: Setting the standards for self-regulated learning. *Contemporary Educational Psychology*, 34(4):306–318, 2009.

D. R. Mulford and W. R. Robinson. An inventory for alternate conceptions among first-semester general chemistry students. *J. Chem. Educ*, 79(6):739, 2002.

J. Nakamura and M. Csikszentmihalyi. The concept of flow. In *Flow and the foundations of positive psychology*, pages 239–263. Springer, 2014.

P. M. Napoli and J. A. Obar. The emerging mobile internet underclass: A critique of mobile internet access. *The Information Society*, 30(5):323–334, 2014.

C. F. Nguyen. The eportfolio as a living portal: A medium for student learning, identity, and assessment. *International Journal of ePortfolio*, 3(2):135–148, 2013.

D. Nicol and D. Macfarlane-Dick. Rethinking formative assessment in he: a theoretical model and seven principles of good feedback practice. *C. Juwah, D. Macfarlane-Dick, B. Matthew, D. Nicol, D. & Smith, B.(2004) Enhancing student learning though effective formative feedback, York, The Higher Education Academy*, 2004.

D. J. Nicol and D. Macfarlane-Dick. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in higher education*, 31(2):199–218, 2006.

G. Nilsson, E. Fok, J. M. Ng, and J. Cooke. A web-based spectroscopy tutorial for the inorganic chemistry laboratory, 2010.

D. Norman. *The design of everyday things: Revised and expanded edition.* Basic Books (AZ), 2013.

G. M. Novak. Just-in-time teaching. *New Directions for Teaching and Learning*, 2011 (128):63–73, 2011.

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

J. Nyachwaya and M. Gillaspie. Features of repreentations in general chemistry textbooks: a peek through the lens of the cognitive load theory. *Chemical Education Research and Practice*, 17:58–71, 2016.

J. Osborne. Practical work in science: misunderstood and badly used? *School Science Review*, 96:357, 2015.

F. Paas and P. Ayres. Cognitive load theory: A broader view on the role of memory in learning and education. *Educational Psychology Review*, 26(2):191–195, 2014.

F. Paas, A. Renkl, and J. Sweller. Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32:1–8, 2004.

F. Paas, T. Van Gog, and J. Sweller. Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, 22(2):115–121, 2010.

A. Paivio. *Mental representations: A dual coding approach.* Oxford University Press, 1990.

I. Parchmann, C. Gräsel, A. Baer, P. Nentwig, R. Demuth, and B. Ralle. “chemie im kontext”: A symbiotic implementation of a context-based teaching and learning approach. *International Journal of Science Education*, 28(9):1041–1062, 2006.

M. Pavelich, B. Jenkins, J. Birk, R. Bauer, and S. Krause. Development of a chemistry concept inventory for use in chemistry, materials and other engineering courses. In

Proceedings of the 2004 American Society for Engineering Annual Conference & Exposition. Paper, volume 1907, 2004.

D. Pavlas, K. Heyne, W. Bedwell, E. Lazzara, and E. Salas. Game-based learning: The impact of flow state and videogame self-efficacy. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 54, pages 2398–2402. SAGE Publications, 2010.

H. E. Pence. Moving chemical education into the cloud (s). *Journal of Chemical Education*, 93(12):1969–1971, 2016.

E. M. Pilke. Flow experiences in information technology use. *International journal of human-computer studies*, 61(3):347–357, 2004.

S. Pinker. *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century.* Penguin, 2014.

P. R. Pintrich. *The role of goal orientation in self-regulated learning.* Academic Press, 2000.

P. R. Pintrich and E. V. De Groot. Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, 82(1):33, 1990.

J. L. Polman. *Designing Project-Based Science: Connecting Learners through Guided Inquiry. Ways of Knowing in Science Series.* ERIC, 2000.

M. Prat-Sala and P. Redford. The interplay between motivation, self-efficacy, and approaches to studying. *British Journal of Educational Psychology*, 80(2):283–305, 2010.

M. Prenzel. The selective persistence of interest. *The role of interest in learning and development*, pages 71–98, 1992.

M. Prince. Does active learning work? a review of the research. *Journal of engineering education*, 93(3):223–231, 2004.

K. Purcell, G. Atfield, C. Ball, and P. Elias. An investigation of the factors affecting the post-university employment of chemical science graduates in the uk. *Warwick Institute for Employment Research, Warwick University, Warwick*, 2008.

C. Randles and T. Overton. Expert vs. novice: approaches used by chemists when solving open-ended problems. *Chemistry Education Research and Practice*, 2015.

T. C. Reeves. Design research from a technology perspective. *Educational design research*, 1(3):52–66, 2006.

T. C. Reeves, J. Herrington, and R. Oliver. Design research: A socially responsible approach to instructional technology research in higher education. *Journal of Computing in Higher Education*, 16(2):96–115, 2005.

N. Reid and I. Shah. The role of laboratory work in university chemistry. *Chemistry Education Research and Practice*, 8(2):172–185, 2007.

P. Reimann. Design-based research. In *Methodological choice and design*, pages 37–50. Springer, 2011.

T. Rienzo and B. Han. Microsoft or google web 2.0 tools for course management. *Journal of Information Systems Education*, 20(2):123, 2009.

E. Ries. *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses.* Random House LLC, 2011.

G. Ring and B. Ramirez. Implementing eportfolios for the assessment of general education competencies. *International Journal of ePortfolio*, 2(1):87–97, 2012.

P. Roberts, D. Maor, and J. Herrington. Scaffolding students' reflection in an eportfolio-based learning environment: Interaction with the technology. 2013.

A. J. Romiszowski. How's the e-learning baby? factors leading to success or failure of an educational technology innovation. *Educational Technology*, 44(1):5–27, 2004.

J. Roschelle, J. Knudsen, and S. Hegedus. From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In *Designs for learning environments of the future*, pages 233–262. Springer, 2010.

B. Rosenshine and C. Meister. The use of scaffolds for teaching less structured cognitive tasks. *Educational Leadership*, 49(7):26–33, 1992.

G. B. Rossman and B. L. Wilson. Numbers and words combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation review*, 9(5): 627–643, 1985.

D. R. Sadler. Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144, 1989.

D. R. Sadler. Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5):535–550, 2010.

J. E. Sale, L. H. Lohfeld, and K. Brazil. Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and quantity*, 36(1):43–53, 2002.

K. Salta and D. Koulougliotis. Assessing motivation to learn chemistry: adaptation and validation of science motivation questionnaire ii with greek secondary school students. *Chemistry Education Research and Practice*, 16(2):237–250, 2015.

A. Samarapungavan. Scientists' conceptions of science: a study of epistemic beliefs. In *annual meeting of the American Educational Research Association, San Francisco, CA*, 1992.

A. Samarapungavan, E. L. Westby, and G. M. Bodner. Contextual epistemic development in science: A comparison of chemistry students and research chemists. *Science Education*, 90(3):468–495, 2006.

M. S. Samuelstuen and I. Bråten. Examining the validity of self-reports on scales measuring students' strategic processing. *British journal of educational psychology*, 77(2):351–378, 2007.

J. Sandars. The use of reflection in medical education: Amee guide no. 44. *Medical teacher*, 31(8):685–695, 2009.

M. Savin-Baden and C. H. Major. *Qualitative research: The essential guide to theory and practice*. Routledge, 2013.

J. M. Scaramozzino. Integrating stem information competencies into an undergraduate curriculum. *Journal of Library Administration*, 50(4):315–333, 2010.

M. Schommer. Effects of beliefs about the nature of knowledge on comprehension. *Journal of educational psychology*, 82(3):498, 1990.

G. Schraw, K. J. Crippen, and K. Hartley. Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36(1-2):111–139, 2006.

J. Schüler. Arousal of flow experience in a learning setting and its effects on exam performance and affect. *Zeitschrift für Pädagogische Psychologie*, 21(3/4):217–227, 2007.

D. H. Schunk. Goal and self-evaluative influences during children's cognitive skill learning. *American educational research journal*, 33(2):359–382, 1996.

D. H. Schunk. Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational psychology review*, 20(4):463–467, 2008.

P. Schwartz and J. Barbera. Evaluating the content and response process validity of data from the chemical concepts inventory. *Journal of Chemical Education*, 91(5): 630–640, 2014.

M. K. Seery. Harnessing technology in chemistry education. *New Directions*, 9(1):77–86, 2013.

M. K. Seery. Student engagement with flipped chemistry lectures. *Spring ConfChem Flipped Classroom. Retrieved from http://confchem. ccce. divched. org/sites/confchem. ccce. divched. org/files/2014SpringConf ChemP1. pdf*, 2014.

M. K. Seery and C. O'Connor. E-learning and blended learning in chemistry education. *Chemistry Education: Best Practices, Opportunities and Trends*, pages 651–670, 2015.

M. K. Seery, H. Y. Agustian, E. D. Doidge, M. M. Kucharski, H. M. O'Connor, and A. Price. Developing laboratory skills by incorporating peer-review and digital badges. *Chemistry Education Research and Practice*, 18:403–419, 2017.

H. Shen. Interactive notebooks: Sharing the code. *Nature*, 515(7525):151, 2014.

S. D. Sieber. The integration of fieldwork and survey methods. *American journal of sociology*, pages 1335–1359, 1973.

M. L. Sinapuelas and A. M. Stacy. The relationship between student success in introductory university chemistry and approaches to learning outside of the classroom. *Journal of Research in Science Teaching*, 2015.

S. A. Sinex and T. L. Chambers. Developing online collaboration skills in the general chemistry laboratory. *Journal of Chemical Education*, 90:1244–1246, 2013.

I. A. Spanjers, P. Wouters, T. Van Gog, and J. J. Van Merrienboer. An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior*, 27(1):46–52, 2011.

M. Stieff, M. Ryu, and J. C. Yip. Speaking across levels–generating and addressing levels confusion in discourse. *Chemistry Education Research and Practice*, 14(4):376–389, 2013.

J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.

J. Sweller. Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of educational psychology*, 81(4):457, 1989.

K. S. Taber. Revisiting the chemistry triplet: drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, 14(2):156–168, 2013.

K. S. Taber. The role of" practical" work in teaching and learning chemistry. *School science review*, 96(357):75–83, 2015.

K. S. Taber and M. Watts. Constructivism and concept learning in chemistry: Perspectives from a case study. *Research in Education*, (58):10, 1997.

V. Talanquer. Macro, submicro, and symbolic: the many faces of the chemistry "triplet". *International Journal of Science Education*, 33(2):179–195, 2011.

M. Tavakol and R. Dennick. Making sense of cronbach's alpha. *International journal of medical education*, 2:53, 2011.

N. Taylor and R. Coll. Using constructivism to inform tertiary chemistry pedagogy. *Chemistry Education Research and Practice*, 2(3):215–226, 2001.

G. P. Thomas. Metacognition in science education: Past, present and future considerations. In *Second international handbook of science education*, pages 131–144. Springer, 2012.

A. Tiberghien, J. Vince, and P. Gaidioz. Design-based research: Case of a teaching sequence on mechanics. *International Journal of Science Education*, 31(17):2275–2314, 2009.

S. Tobias. The expertise reversal effect and aptitude treatment interaction research (commentary). *Instructional Science*, 38(3):309–314, 2010.

M. Towns, C. J. Harwood, M. B. Robertshaw, J. Fish, and K. O'Shea. The digital pipetting badge: A method to improve student hands-on laboratory skills. *Journal of Chemical Education*, 92(12):2038–2044, 2015.

D. Treagust, G. Chittleborough, and T. Mamiala. The role of submicroscopic and symbolic representations in chemical explanations. *International Journal of Science Education*, 25(11):1353–1368, 2003.

L. W. Trowbridge, R. W. Bybee, and J. C. Powell. *Teaching secondary school science: Strategies for developing scientific literacy.* Prentice Hall, 2000.

M. Undorf and E. Erdfelder. Separation of encoding fluency and item difficulty effects on judgements of learning. *The Quarterly Journal of Experimental Psychology*, 66 (10):2060–2072, 2013.

M. Vaismoradi, H. Turunen, and T. Bondas. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & health sciences*, 15(3):398–405, 2013.

M. Vaismoradi, J. Jones, H. Turunen, and S. Snelgrove. Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice*, 6(5):100, 2016.

A. R. Van Dyke and J. Smith-Carpenter. Bring your own device: A digital notebook for undergraduate biochemistry laboratory using a free, cross-platform application, 2017.

J. Van Merriënboer, Jeroen JG Sweller. Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2): 147—178, 2005.

J. J. Van Merriënboer and J. Sweller. Cognitive load theory in health professional education: design principles and strategies. *Medical education*, 44(1):85–93, 2010.

J. J. Van Merriënboer, P. A. Kirschner, and L. Kester. Taking the load off a learner's mind: Instructional design for complex learning. *Educational psychologist*, 38(1):5–13, 2003.

J. B. Velasco, A. Knedeisen, D. Xue, T. L. Vickrey, M. Abebe, and M. Stains. Characterizing instructional practices in the laboratory: The laboratory observation protocol for undergraduate stem. *Journal of Chemical Education*, 93(7):1191–1203, 2016.

C. Von Aufschnaiter and S. Von Aufschnaiter. University students' activities, thinking and learning during laboratory work. *European Journal of Physics*, 28(3):S51, 2007.

E. Von Glasersfeld. An introduction to radical constructivism. *The invented reality*, pages 17–40, 1984.

A. Vrugt and F. J. Oort. Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. *Metacognition and Learning*, 3 (2):123–146, 2008.

J. Walker, V. Sampson, S. Southerland, and P. Enderle. Using the laboratory to engage all students in science practices. *Chemistry Education Research and Practice*, 17(4): 1098–1113, 2016.

F. Wang and M. J. Hannafin. Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4):5–23, 2005.

H.-H. Wang, T. J. Moore, G. H. Roehrig, and M. S. Park. Stem integration: Teacher perceptions and practice. *Journal of Pre-College Engineering Education Research (J-PEER)*, 1(2):2, 2011.

J. D. Weibel. Working toward a paperless undergraduate physical chemistry teaching laboratory. *Journal of Chemical Education*, 93:781–784, 2016.

S. P. Whalen. Assessing flow experiences in highly able adolescent learners. 1997.

G. H. Wheatley. Problem solving in school mathematics. *MEPS Technical Report*, 84, 1984.

N. A. Williams, W. Bland, and G. Christie. Improving student achievement and satisfaction by adopting a blended learning approach to inorganic chemistry. *Chemistry Education Research and Practice*, 9(1):43–50, 2008.

T. Wilson and D. Read. Learning from labdog: Best practice for laboratory response system questions. In *Conference Proceedings. New Perspectives in Science Education*, page 299. libreriauniversitaria. it Edizioni, 2017.

T. M. Winberg and C. A. R. Berg. Students' cognitive focus during a chemistry laboratory exercise: Effects of a computer-simulated prelab. *Journal of Research in Science Teaching*, 44(8):1108–1133, 2007.

J. Wittwer and A. Renkl. Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43(1):49–64, 2008.

D. M. Wolfe and D. A. Kolb. Beyond specialization: The quest for integration in midcareer. *Work, Family, and Career, edited by CB Derr. Praeger*, 1980.

J. Xian and D. B. King. The effectiveness of general chemistry lab experiments on student exam performance. *Journal of Laboratory Chemical Education*, 5(5):95–107, 2017.

S. A. Yoon and C. E. Hmelo-Silver. What do learning scientists do? a survey of the isls membership. *Journal of the Learning Sciences*, (just-accepted), 2017.

M. Yorke. Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education*, 45(4):477–501, 2003.

R. N. Zare. On the love of teaching and the challenge of online learning: A few reflections. *Journal of Chemical Education*, 77(9):1106–1107, 2000.

E. Zell and Z. Krizan. Do people have insight into their abilities? a metasynthesis. *Perspectives on Psychological Science*, 9(2):111–125, 2014.

K. Zimbadi, A. Loyle-Langholz, J. Kibedi, and K. Colthorpe. Using inquiry-based practicals to promote students' critical evaluation of the scientific literature and maturation of their understanding of the nature of scientific knowledge. *International Journal of Innovation in Science and Mathematics Education*, 23(5):91–103, 2015.

B. J. Zimmerman. Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary educational psychology*, 11(4):307–313, 1986.

B. J. Zimmerman. Attaining self-regulation: A social cognitive perspective. 2000.

B. J. Zimmerman and M. Campillo. Motivating self-regulated problem solvers. *The psychology of problem solving*, pages 233–262, 2003.

B. J. Zimmerman and A. R. Moylan. Self-regulation: Where metacognition and motivation intersect. *Handbook of metacognition in education*, pages 299–315, 2009.

B. J. Zimmerman and D. H. Schunk. Self-regulated learning and performance. *Handbook of self-regulation of learning and performance*, pages 1–14, 2011.

A. Zohar and S. Barzilai. A review of research on metacognition in science education: current and future directions. *Studies in Science Education*, 49(2):121–169, 2013.