The Design and Evaluation of Laboratories in Engineering Education: The Purpose-First Approach

Joshua Burridge

B. Tech (Computing Studies), MIT (Software Engineering)

Supervisor: Professor David Lowe
Auxiliary Supervisor: Professor Judy Kay
Auxiliary Supervisor: Professor Keith Willey

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

This research reported in this thesis was supported
by the award of a Research Training Program
scholarship to the PhD Candidate.

School of Computer Science
Faculty of Engineering
The University of Sydney
Australia

2024

ABSTRACT

Laboratories are a key component in the practice of engineering education, typifying much of the identity of the discipline. Throughout the information age new technologies have disrupted these teaching and learning activities with innovations around access, observation, and pedagogy – however these innovations have seen only sporadic take-up and continue to be understudied in the literature. This thesis considers the techniques used in the design and evaluation of laboratories in engineering education as a potential inhibiting factor in this adoption. It investigates how these practices might be developed into more effective ways to answer the questions being asked of these teaching and learning activities, with overarching research question being:

> "How may laboratories in engineering education be effectively designed to meet the purposes their stakeholders require, and evaluated against these purposes?"

Prior research work to this thesis by the author indicated an issue with the standard practice of evaluating laboratories in engineering education. This issue can be described as the application of traditional research methods from the practice of *engineering* to the research of engineering *education*, particularly quantitative experimental designs which attempt to control the complexity in these contexts. The impacts of this misalignment were observed to be research outputs which were lacking in internal and external validity, limiting their ability to act as impetus for the adoption of innovations, and inputs to a collective understanding of the nature of laboratories.

This thesis followed an Educational Design Research (EDR) methodology to investigate this issue, to iteratively design and evaluate a set of artifacts intended to improve outcomes in the design and evaluation of these laboratories. EDR is an iterative methodology which seeks to generate both an artifact which is the object of the study, as well as design process learnings obtained through the iterations refining its design. The iterations this thesis followed incorporated literature review, synthesis, expert review, and empirical studies.

The literature review and synthesis were in large part driven by siloing observed in research practices in this discipline, with solutions to similar issues developed in other disciplines such as science, mathematics and nursing education not seeing significant breakthrough into engineering education research practice. Also incorporating pedagogic theory and cognitive psychology, and with a focus on research methodologies, the synthesis derived an initial set of artifacts to be applied to the design and evaluation of laboratories in engineering education.

These artifacts comprised an Approach incorporating an ontology, epistemology, and process, and a Framework providing structural and communicative support for applying the process. These artifacts were then iterated on with input and validation from engineering education researchers and practitioners in their own teaching contexts. The artifacts developed by this thesis following the EDR methodology are the Purpose-First Approach, and the PIPOV Framework.

The Purpose-First Approach is an embodiment of an ontological, epistemological, methodological, and process perspective on the design and evaluation of laboratories in engineering education. It is characterised by the statement:

> Designing, refining, evaluating, and reporting on laboratories in engineering education must all explicitly derive from the purpose(s) to which that laboratory will be applied.

It embeds the EDR methodology as a better-aligned methodology more likely to achieve improved outcomes in laboratory evaluation than current practice, as evidenced by examples from other disciplines in the literature and the empirical work of this thesis.

The PIPOV Framework is a supporting tool to aid in the effective application of the Purpose-First Approach by practitioners without an extensive background in the theories embedded in the Approach, in order to allow adoption quickly and easily. The Framework draws from prior laboratory characterisation studies in the literature and embeds contributions from constructivism/social constructivism and Constructive Alignment, and EDR, as well as extensive iterative refinement and empirical investigation throughout the work in this thesis.

The empirical work to iteratively refine and validate these artifacts were an expert review, a longitudinal collaborative study, and a pair of validation studies. The expert review involved one-on-one interviews with N = 9 engineering education practitioners who had previously published research on laboratories in engineering education, in which the artifacts were described and stepped through to generate discussion and feedback to improve their design. The longitudinal collaborative study involved the author being embedded with a team in another institution designing, implementing, and validating a new laboratory using an iterative process over approximately three years. Finally, the validation studies involved 2 sessions with teams designing the evaluation process for their laboratories, one new laboratory and one modified laboratory, using the generated artifacts.

The empirical work in this thesis supported by the literature synthesis indicates that these artifacts are effective in supporting the design and evaluation of laboratories in engineering education. Their design processes are intuitively familiar to engineering educators via their design practice, and their evaluation practice mandates and recommendations have been positively adopted by practitioners and observed to be effective in their own contexts.

This thesis provides a series of contributions to knowledge as outputs from the above development and empirical research. The literature review and synthesis conducted provides a multi-disciplinary body of knowledge in support of improving the maturity level of engineering education research on the back of work already completed in literature not typically cited in our discipline. An initial investigation of the scope of 'what is a laboratory' involving interviews with experts in laboratory research. The Purpose-First Approach and PIPOV Framework as artifacts are contributions as empirically validated tools for the design and evaluation of laboratories in engineering education. Following the EDR methodology, the design insights, decisions, motivations, and observations that were a part of the iterative process of developing these artifacts are also provided as a research output, for incorporation in parallel design work in other contexts, extension, or refinement of these artifacts themselves by others.

Engineering education stands on the cusp of an educational technology explosion – networked equipment, learning analytics, generative AI, haptic digital interfaces, augmented reality, and digital presence to name a few. Challenges in engineering education means the status quo is likely to be disrupted regardless of our ability to be ready for it. Current evaluation practice has not been effective at answering the questions being asked, and innovation in these practices is clearly needed to support the follow-on innovation in teaching and learning activity design. The Purpose-First Approach and PIPOV Framework provide an evidenced and practice-tested answer to this call.

ACKNOWLEDGEMENTS

There are so many people to whom are owed more thanks than I can possibly express in any reasonable time or space or the limitations of text. I will do my best to capture all that I can, and make no excuses for any omissions.

First to the supervisors who gave me the support to tackle a problem wherever that problem led, even though it was hard, out of the ordinary, and not the way things used to be done - Professors David Lowe, Judy Kay, and Keith Willey.

To my family, who supported me no matter how little sense the words I was saying made.

To Andrea and the triplets (and Kevin!), who became family.

To the denizens of the dry office and the wet office, for making sure to collect everybody when the time came to go from one to the other.

To all the people in the School and Faculty who helped along the way, whether teaching me or letting me teach.

To Tim, David 1, and David 2, for managing to deal with an entire world of separation and still build cool and valuable things.

To the academics who were just crazy enough to try something new.

And to the students who were just polite enough to go along with it.

STATEMENT OF ORIGINALITY

This is to certify that to the best of my knowledge; the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Joshua Burridge

List of Publications

The following publications involving work in this thesis are presented with the author's contribution.

Reference	Topic & Relevance	Contribution
J. Burridge and D. Lowe, "Breaking down the laboratory supertype conflation," in *Proceedings - Frontiers in Education Conference, FIE*, 2018, vol. 2018-Octob.	This paper was transitional between the earlier Masters research and this thesis, involving literature synthesis from the earlier work, but with some analysis, writing, and publication overlapping with the initial stages of this thesis. It examined a key limitation in laboratory examination which led to my questioning the appropriateness of experimental designs more broadly.	The synthesis and analysis were conducted by me with feedback, guidance, and proof-reading from my thesis supervisor.
J. Burridge and J. Kay, "Dark Data and Learning Analytics," in *Australian Learning Analytics Summer Institute Proceedings*, 2019.	This presentation was examining the idea of education data which was being collected by new computer-mediated/web interfaces but not being examined. This theoretical synthesis work led directly to the 'Defeating Hawthorne…' applied paper described below.	The synthesis and analysis were conducted by me with feedback, guidance, and proof-reading from an auxiliary supervisor
A. Hadigheh, J. Vulic, J. M. Burridge, T. Goldfinch, J. Thomas, and A. Opdyke, "Preliminary Evaluation of Immersive and Collaborative Virtual Labs in a Structural Engineering Unit of Study," *AAEE2019 Conf.*, 2019.	This paper presented empirical work on the design and evaluation of an immersive VR laboratory. The evaluation method used in this study was an early exposure for me in the application of qualitative methods towards the evaluation of laboratories, and while lacking the precision possible from quantitative methods, demonstrated their ability to give insights that could be iterated on.	I came onto the project to provide an engineering education laboratories research perspective to the collaborators, who included engineering education practitioners, and collaborators out of general education research. I contributed to the design discussions leveraging literature on non-traditional laboratories, and in partnership with the education research member of the team developed and conducted the in-class observation research protocol. I also contributed to the analysis of results and writing of the paper.

J. Burridge, D. Lowe, K. Willey, and J. Kay, "Defeating Hawthorne in tech-enabled education: Passive observation of student behaviour with a remote laboratory The Laboratory System," *AAEE2019 30th Annu. Conf. Australas. Assoc. Eng. Educ.*, 2019.	This paper extended the earlier literature synthesis work on dark data to provide an empirical example leveraging data that had been automatically collected but not analysed during the Masters research work. This case study was later used to inform discussion of white-boxing the evaluation approach to laboratories when describing or applying PIPOV.	I performed the software processes of log acquisition, cleaning, processing, and analysis to produce the analytics. I wrote the paper with feedback, guidance, and proof-reading from my thesis supervisor as well as auxiliary supervisors.
J. Burridge, "The Purpose-First Approach to Laboratory Design, Evaluation, and Research," in *Proceedings of the 31st Annual Conference of the Australasian Association for Engineering Education*, 2020.	This workshop presented the current state of the thesis artifacts and invited activity from participants applying it to their own laboratories. It is described in section 7.2.3.1.	I developed the workshop materials and gave the content presentation. I facilitated the work on laboratory contexts in the workshop with an auxiliary supervisor co-facilitating.
D. Reid, J. Burridge, D. Lowe, and T. Drysdale, "Open-source remote laboratory experiments for controls engineering education," *Int. J. Mech. Eng. Educ.*, vol. 50, no. 4, pp. 828–848, Oct. 2022.	This paper presented the empirical evaluation of the Edinburgh remote laboratory designed in collaboration with this thesis work. It is based on work described in Chapter 6.	I contributed to the design process for the laboratory as part of a long-term research collaboration between laboratory researchers at the University of Sydney and the University of Edinburgh. Development was primarily conducted by the Edinburgh team, with the Sydney members contributing expertise in design and evaluation research practices. The research was designed by the first author from the Edinburgh team relating to the interface they had developed on the equipment. My contribution to the data collection was advice on usability instrument selection, and qualitative questions. I iteratively contributed to the quantitative analysis with the first author leading and other members of the collaboration giving

		feedback. I designed the analysis of the qualitative results and iteratively contributed to this analysis with the first author with feedback from other team members. The first author led the writing of the paper, with contributions by me particularly to the qualitative component and all members giving guidance and proof-reading.
J. Burridge and D. Lowe, "Purpose-First & PIPOV: A framework to enable design-based research for engineering education laboratories," in *HERDSA*, 2022.	This presentation described the current state of the thesis artifacts along with empirical validation results to that point.	I developed the artifacts through the course of this thesis work with feedback from my supervisors. The presentation was written by me, with feedback, guidance, and proof-reading from my thesis supervisor.

AUTHORSHIP ATTRIBUTION STATEMENT

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Joshua Burridge

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor David Lowe

HUMAN RESEARCH ETHICS APPROVALS

This thesis involved several studies involving human participants. These studies were all approved by the University of Sydney Human Research Ethics Committee process, with the exception of the study involving students at the University of Edinburgh which were approved by the equivalent processes at that institution. The following ethics protocol numbers and titles from the University of Sydney are associated with this research:

- **2017/942**
  'Towards a research-informed design for immersive virtual reality (IVR) environments'
- **2018/229**
  'Sahara Remote Laboratories Integration'
  Retitled 'Student Experience of Engineering Education Laboratories' after amendment.
- **2019/1040**
  'Understanding Engineering Education Laboratories'

The "Student feedback on 'Controls 3' remote lab" data project associated with this research was reviewed and approved through the University of Edinburgh Research Ethics and Integrity self-assessment process as low-risk. The data from this project used directly in this thesis work was based solely on the public dataset version available at https://datashare.ed.ac.uk/handle/10283/4007, and so reciprocal approval through the University of Sydney was not required.

CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

SELECTED GLOSSARY & ABBREVIATIONS

This thesis involves the use of several highly overloaded terms that require precision in the way they are understood. This frontismatter provides a description not to present any authoritative singular meaning, but rather to flag for readers the way in they are used and should be understood in the context of this thesis. It also provides a central reference for abbreviations used.

- AAEE
  Australasian Association for Engineering Education
- ABET
  Accreditation Board for Engineering and Technology
- AJEE
  Australasian Journal of Engineering Education
- ANOVA
  Analysis of Variance
- AQF
  Australian Qualifications Framework
- AR
  Augmented Reality – see VR; Virtual Reality
- Asynchronous
  Involving a disconnect between some elements in respect to time. For example, one student doing some solo work with remote laboratory equipment at 4pm on a Saturday, while another does so at 1am on a Tuesday, and class is at 9am on Thursdays. Can be partial – students might be working at the same time as each other but not their teacher, or a class might submit instructions to a piece of equipment but not hear the results back for a few days.
- AT
  Assessment Task
- Batch Mode
  A type of laboratory use which involves the giving of instructions asynchronously and waiting for results, rather than having direct control of the equipment.
- Bloom
  Short for Bloom's Taxonomy, a framework for classifying intended learning outcomes.
- CA
  Constructive Alignment
  A framework for understanding and defining teaching and learning
- DBR
  Design-Based Research; see EDR
- Design Research
  See EDR
- EA
  Engineers Australia
- EDR
  Educational Design Research.
  An iterative research approach which embeds multiple component methods and instruments to triangulate answers to research questions progressively. Many different terms are used inconsistently to describe variations of this approach. Another notable

conflation can exist with the teaching of design, as in 'research about design education', which this thesis is not engaged in.

- EJEE
  European Journal of Engineering Education
- Engineering
  Broadly and largely arbitrarily defined for this thesis, with blurred borders between other disciplines such as science, biomedicine, project management and business, software development, etc. Defined primarily through faculty names (e.g. Faculty or School of Engineering), and through research publication venues.
- Engineering Research
  See EDR
- General Instructional Theory
  A research goal for EDR after identifying and elaborating on the local instructional theory across multiple contexts. The collective transferable learnings which can be used by readers of EDR research.
- hands-on
  Laboratories involving physical interaction with equipment – contrasted with computer-mediated
- HCI
  Human-Computer Interaction
- ICT
  Information and Communication Technology
- ILO
  Intended Learning Outcome
  What students should be able to do (alternatively know, understand, be competent in, etc.) after completing some component of study.
- IPO
  Input-Process-Output
  Systems analysis model for the application of a system to data, material, people, etc.
- JEE
  Journal of Engineering Education
- KIPPAS
  Knowledge and understanding, Inquiry skills, Practical skills, Perception, Analytical skills, and Social and scientific communication
  See reference [1]
- Laboratory
  This thesis includes an attempt to define and then scope this term in section 7.2.1. This thesis is scoped specifically to laboratories used for teaching purposes rather than research or manufacturing, and usually in an engineering context.
- local instruction theory
  The core element being progressively elaborated by the EDR process. Initially a learning hypothesis, which is detailed and validated iteratively through each phase.
- MR
  Mixed Reality – see VR; Virtual Reality
- NTL
  Non-traditional Laboratory
  Any laboratory which uses a different modality or design style to a 'traditional' laboratory.

Often defined as remote (conducted over the internet), simulated (no real equipment), or VR/AR/MR/XR enabled – but this is limited only by the available designs and technology. Generative AI is likely a new entrant to this list.

- PF
  Purpose-First Approach
  The guiding principle for the artifact developed for this thesis: "Designing, refining, evaluating, and reporting on laboratories in engineering education must all explicitly derive from the purpose(s) to which that laboratory will be applied." See 7.1.4.
- PIPOV
  Purpose, Inputs, Processes, Outcomes, Validation
  The name for a visualised framework which this thesis developed as an artifact in service to the Purpose-First Approach. See 7.1.5.
- PRISMA
  Preferred Reporting Items for Systematic Reviews and Meta-Analyses
  A structured approach to reporting what and how many resources were included or excluded at each step in a systematic review.
- Proximal
  A term for laboratories where the students using the laboratory are proximal, i.e. in the same room as, the equipment and/or interfaces they are using. Might also apply to other participating elements, such as the students being proximal to their teacher. Cf. Remote.
- Purpose
  A loaded term in this thesis as applied to laboratories. Captures the intended learning outcomes, but also any other reasons or motivations for a laboratory to be evaluated against, from any and all involved stakeholders – such as cost, space, student throughput, etc.
- RCT
  Randomised Controlled Trial
  A formal quantitative experimental design often used in medical research. Attempts to completely control for confounding and nuisance variables.
- Remote
  A term for laboratories where the students using the laboratory are remote, i.e. in a different room, building, country, etc. as the equipment/and or interfaces they are using. Might also apply to other participating elements, such as the students being remote from their teacher. Cf. proximal.
- RQ
  Research Question
- SCQF
  Scottish Credit and Qualifications Framework
- Simulated
  A term for laboratories which do not involve (or only partially involve) real equipment and/or real data. Computer simulation.
- SOLO
  Structure of the Observed Learning Outcome
  A framework for classifying intended learning outcomes.
- Synchronous
  See Asynchronous

- TA
  Teaching Assistant; Tutor
- TL
  Traditional Laboratory; see NTL
- TLA
  Teaching and Learning Activity
  A bounded component of a teaching program. Often a single class or an activity in a class, might also be an assessment process occurring in parallel to other TLAs, such as an ongoing portfolioing activity.
- UEQ
  User Experience Questionnaire
  An HCI instrument for investigating user experiences with interfaces.
- UI
  User Interface
- UX
  User Experience
- VEE
  Virtual Experiment Environment
  A term used to mean a remote laboratory – notably not virtual as this thesis defines it.
- VR
  Virtual Reality
  A deeply overloaded term – originally used in reference to any environment virtually simulated by a computer, particularly virtual 3d environments displayed on 2d computer screens. Now more commonly used in reference to headset-based immersive virtual environments involving head-tracking and simulated physical interaction with virtual actors and objects. Has several variants, of which VR might be this base version, or might be a collective term for all, and none used consistently. Augmented reality takes in real world data to leverage inside the virtual reality, mixed reality pairs real world vision with virtual elements, extended reality includes any number of extended data sources or interaction methods to blend virtual and real elements. Useful to understand for this thesis largely in the conflated interactions between terms such as ‘virtual labs’, ‘labs with VR/AR/MR/XR elements’, ‘virtual data’, etc., especially with different interpretations between different cited literature.
- XR
  Extended Reality – see VR; Virtual Reality
- ZPD
  Zone of proximal development

1 INTRODUCTION

Laboratories are a fundamental part of many engineering education curricula. They are mandated by many accreditation bodies [2], used in marketing engineering to students [3], and are the subject of significant budgetary and human investment [4]. Laboratories are often described as delivering 'practical' education to students, in a discipline that may be described as the practical investigation and manipulation of our built and lived environment [5].

However ubiquitous and critical laboratories are held to be, this thesis will show they are not an activity that is well understood in the literature, nor is the significant research efforts expended to understand and improve these teaching and learning activities effective. This thesis will first identify and explain these issues, then synthesise a proposed solution. In parallel with this synthesis, it will empirically refine and develop this solution, to better understand what benefits it might yield.

In 2005, Feisel and Rosa published what would become the most highly-cited research on laboratories in engineering education the Journal of Engineering Education (and most likely the entire discipline) has yet seen [6]. In this paper, they established some key points about these laboratories:

- Laboratories were a critical part of engineering education.
- Innovative technologies were drastically altering the laboratory space, and lack of understanding was hampering the effective adoption of these innovative technologies.
- Research on these laboratories had up to that point been very limited.
- What research had been conducted was greatly lacking in consideration of laboratory objectives.

It is now nearly 2 decades since Feisel and Rosa's research, and based on the work that led to this thesis the following statements will be shown to remain true:

- Laboratories remain a critical part of engineering education.
- Innovative technologies continue to drastically alter the laboratory space.
- Lack of understanding continues to hamper the effective adoption of these innovative technologies.
- Research on laboratories continues to be limited (though in different ways)
- What research is being conducted is greatly lacking in consideration of laboratory objectives.

Besides the obvious observation that the hampering of these innovative technologies is a problem that should be solved, the impacts of these continuing issues with laboratory research are significant. On the one hand, laboratories are being designed and run without a significant research foundation to identify best practices, to give insight into how to correct for given errors, etc.

On the other hand, while research on laboratories continues to be limited, this limit is no longer the lack of publication observed by Feisel and Rosa. Currently there exists significant bodies of published research on laboratories, however, this thesis will show it is limited by a general lack of internal and external validity. This limitation means that effort is currently being misdirected – and that greater value could and should be derived from ongoing research effort in this field.

The fundamental problem with current practice that this thesis observes is the attempted application of quantitative (pseudo-) experimental research methodologies under circumstances that inherently fail to support them. Experimental research methodologies are a core part of the engineering discipline – ironically, they are often what is being taught as part of the very engineering education laboratories being evaluated. However, these methodologies carry specific requirements, which when breached mean their mathematical basis is no longer valid. The educational contexts laboratories are assessed in routinely breach these requirements, and much of the published research on laboratories in engineering education does not properly address this invalidity. Even where these internal and external validity issues are correctly identified, the potential impact of the research effort expended is therefore greatly reduced. This invalidity present in individual research activities involving laboratories means that secondary or meta studies involving laboratories is thus unable to establish a valid foundation, leading to a general inability to identify and support answers to the questions being asked of laboratory research since Feisel and Rosa and earlier.

This thesis first dives into the nature of this problem – the nature of different levels of experimental design, the methodologies employed in current research, and the breaches that inherently occur. It then explores potential alternative approaches, and proposes a solution. Finally, it validates that proposed solution, which it titles the Purpose-First Approach, and the accompanying PIPOV Framework.

1.1 The Path to The Problem

This thesis found itself in a markedly different place to where it began. This is likely true to some extent for all theses, but in this case the journey to the eventual problem statement is a fundamental part of establishing why that question is so important, and what the potential impact of this research may be.

In a perfect world, this thesis would instead present and investigate some implementations of the fantastic innovative technologies available for laboratory development – remote access, integration of simulated and real data, novel interfaces – from VR/AR to haptic controls. These technologies present new opportunities that hold enormous promise on existing laboratory attributes, such as availability (remote equipment need not be scheduled for a specific class time), accessibility (new interfaces can support students with differing physical and mental requirements), budget (time-sharing, elimination of set-up/tear-down/clean-up, miniaturisation to save space, prevention of breakage due to user error – all of these can serve to provide the laboratory at a cheaper cost).

They further present incredible opportunities to do things students in a classroom have never been able to ‘do’ before – experiments too dangerous to allow students to conduct, such as involving explosives, large scale equipment, or biohazards; or experiments fundamentally infeasible to have students conduct, for example, a remote access experiment sent up to the International Space Station. Most relevantly, these technologies present the possibility of improved pedagogic outcomes – however, unlike the clearly demonstrable benefits to concretely definable and measurable attributes, these have not yet been established.

This thesis work grew out of an earlier research project to do just that – take an implementation of a laboratory leveraging innovative technologies (in this case remote access and the integration of both real and simulated data), and then evaluating it. The stumbling block for this research project was the repeated and unanswered question “how do you know if it is working?” – moving beyond the immediate “the equipment reacted as expected” to the far more complex “did the students learn what they needed to by using it?”.

As will be described in detail to follow, the current methodologies most often used in engineering education research are not fit for purpose when attempting to answer this question.

Incompatibilities between the requirements of the usual methodologies and the fundamental nature of education research have significantly inhibited the potential value of the research efforts being conducted as a result.

A common saying for the nature of research is Newton's "If I have seen farther, it is by standing on the shoulders of giants" [7] – it is the position of this thesis that the core issue with research on laboratory education is the lack of a necessary giant. That is to say, there is key pre-requisite knowledge and research that has not been integrated into the discipline, and that lack has left the research being conducted without a necessary foundation, and so left it fundamentally flawed.

Rather than taking the grandiose and likely futile approach of trying to be that giant, this thesis instead seeks existing giants serving similar purposes elsewhere. The general challenges present in evaluating laboratories have proven not unique to laboratories, and so the eventual problem this thesis tackles is identifying and attempting to integrate knowledge from outside current engineering education practice that may fill that gap, and lead to increased value and impact from ongoing laboratory research.

1.2 PREMISE, PROBLEM STATEMENT, AND RESEARCH QUESTIONS

While the identification and selection of a research approach was a defining component of this research work, it is necessary in this introduction to 'skip ahead' and frame the problem within the approach which was eventually chosen as the most appropriate. This approach is Educational Design Research, which involves a phased or iterative blending of research with contextual practice. This approach does not begin with fundamental research questions – rather it begins with a context and problem statement, and research questions are generated sequentially through different phases or iterations in service to these. This section will map out the flow of these questions and reference the subsequent chapters and sections where they arise. The overall flow of these research questions and their findings leading to the next research question in the following phase is also described holistically in chapter 5.

The problem statement for this research work is the need for rigorous and actionable methods for the design and evaluation of laboratories in engineering education. Stated as a scoping research question this is:

> "How may laboratories in engineering education be effectively designed to meet the purposes their stakeholders require, and evaluated against these purposes?"

Answering this overall research question involved a sequential list of contributing sub-questions:

1. How do the research methodologies used in evaluative research on laboratories in engineering education support or inhibit answering the questions this research poses?
2. What methodological alternative provides the appropriate characteristics to resolve the identified issues with experimental approaches in the design and evaluation of laboratories in engineering education?
3. Can Educational Design Research be successfully applied to the design and evaluation of laboratories in engineering education?
4. How can practitioners be supported in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?
   4.1. Do the Purpose-First Approach and PIPOV Framework support practitioners in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?

These contributing sub-questions each map to different iterative phases. Their sequence and connections is described holistically in chapter 5, and then individually in turn through the subsequent chapters.

1.3 Research Contributions

This thesis presents the following contributions to knowledge:

- Literature synthesis drawing knowledge from multiple external disciplines and collecting this into a coherent body of knowledge on laboratories in engineering education and how they might be designed and evaluated.
- A systematic review on practical inquiry methods applied to laboratories in engineering education, section 2.6.5.4
- Identification and definition of the 'research contortion' effect on research in engineering education, section 3.2.
- Describing and defining the Educational Design Research methodology through a structure of named research traditions and distinguishing literature in order to support its investigation in spite of the mixed terminology issues.
- The Purpose-First Approach and PIPOV Framework as artifacts supporting the design and evaluation of laboratories in engineering education.
- A lens for describing the scope of what is considered a laboratory teaching and learning activity.
- Design decisions made in the process of designing and refining the artifacts, and the literature and empirical observations which motivated this as an Educational Design Research output.
- Identification of white-boxed design and evaluation of teaching and learning activities incorporating mid-way observation and intervention via learning analytics as a critical opportunity in engineering education laboratory practice

In recognising that the current practice of laboratory evaluation is ineffective, and that some new and more effective practice should instead be used, it is perhaps unsurprising that the methodology used by this thesis is itself non-standard when compared to more traditional engineering education theses. As will be described in the methodology discussion to follow, this thesis takes a meta-methodology approach that mirrors the iterative design methodology it endorses. As such, the research contributions are intertwined with the overall design and development work of this thesis, and must be explicitly extracted and articulated. The detailed findings and what they contribute will be a feature of each relevant phase, but the source of these contributions in the abstract are as follows.

1.3.1 Literature synthesis

This thesis uses literature synthesis as a core technique for the generation of new knowledge. This synthesis includes examination of the existing literature on laboratory evaluation research in engineering education based on multiple pre-extant intersecting systematic literature reviews, as well as a review of papers using literature search and snowballing. Beyond laboratory evaluation research, further literature beyond engineering education is incorporated to synthesise potential solutions to identified challenges.

1.3.2 Empirical understanding

Leveraging artifacts created by the above synthesis, empirical studies were then conducted to investigate the conclusions drawn and refine the artifacts iteratively. These studies include the expert review to iterate on the designs of the artifacts, longitudinal work in applying the evaluation approach the artifacts are intended to communicate, and validation studies to investigate the feasibility of their use in supporting this approach. A contextual study was also conducted to attempt to establish a definition of laboratories in engineering education.

1.4 Positionality Statement

This thesis work is heavily influenced by the observation that context matters and cannot be meaningfully controlled for or extracted from the rich and interconnected environments of education. This is similarly the case for the empirical work itself that I conducted. Positionality statements are a common way to give the informed reader a window to this complexity by the reflexive description of this context by one who is conducting and reporting on research. Rowe [8] in [9] defines positionality as

> "the stance or positioning of the researcher in relation to the social and political context of the study (the community, the organization, or the participant group)"

Throughout this thesis the empirical work often involved my own observations made while embedded with laboratory designers in a collaborative space. In this context the critical stance or positioning is my relation to the social and political context of the educators designing and evaluating laboratories in engineering education.

My identity was in a state of flux during the time of this PhD thesis work. I was undergoing and continue to undergo a transition from software engineer, to engineering and IT educator and researcher. My undergraduate degree was a Bachelor of Technology (Computing Studies) which focused on the craft of software development rather than theoretical computer science. This degree was entirely taught online at an early time for that style of university education (late 2000s/early 2010s). After working as a software developer on web development and photographic systems, I studied a Master of Information Technology (Software Engineering) (Research Pathway), and it was during this degree that I first began working on laboratory design and evaluation, as part of a research project which initially leveraged my web development background, but quickly migrated into a crash course on educational laboratories, pedagogy, and evaluation. The questions raised as part of this research project led directly into this thesis project for deeper investigation.

Alongside this research work, I began teaching at university. Initially during the Masters degree as a tutor and content developer writing material and assessments in software engineering classes, progressing to lecturing drawing on my degrees and industry experience coinciding with the start of this thesis work. Throughout this time, I have been acting as a tutor, leading tutoring teams, designing teaching material, and lecturing in software engineering classes. Towards the end of this thesis work I transitioned to become a teaching academic coordinating classes, and at the time of concluding this thesis writing I have taken a teaching and research role providing capability development of academics in engineering and IT education and researching into engineering and IT education.

My motivations behind investigating the design and evaluation of laboratories in engineering education is due to a capability gap I observe between what is known to be possible in terms of technology and pedagogic innovation, and what is practiced. Research continues on exciting innovations, but the penetration of these innovations continues to appear limited to specific research projects by individual academics rather than triggering widespread quality improvement across the discipline. For example, the remote laboratory system I worked on during my Masters degree was originally developed from 2009-2012, but had been largely stagnant since the ending of funding, and was last updated in 2016. Something was inhibiting the adoption of these technologies, and my speculation was that the design and evaluation practices were the most likely cause.

My experience working with engineering educators in designing and evaluating laboratories and interacting in general has been one of an 'outside insider'. The discipline of software engineering in my personal experience is held separate from other types of engineering, often taught by a school of computer science rather than engineering schools as it was at my home university. The common techniques, practices, and ways of thinking between the 2 related disciplines however makes 'shop talk' flow very easily, and in conversations with engineering education practitioners as part of this thesis work any barriers to discussion were overcome quickly. The 'outside insider' role was one that helped greatly in observing engineering educators in their practice – it was normalised for me to ask questions about the specific engineering being done, and participants were quick to expand jargon or assumed knowledge as I was not within the same disciplines, so when asking research-motivated questions these were part of normal conversation flow and not intrusive.

The lens I use to interpret and analyse the observations made is described in detail as a search for and synthesis of an appropriate methodology, but in brief I used Educational Design Research as a research process and methodology embodying a post-positivist ontology and epistemology, with constructivism and Constructive Alignment as my primary pedagogic theory perspective.

1.5 Structure of This Thesis

This thesis follows a meta-implementation of the educational design research methodology. As such, it is focused on the development and refinement of an artifact (the final artifact comprises the Purpose-First Approach and PIPOV Framework). This is presented in the following chapter structure:

- **Part 1**: Background, Methodology, & Synthesis
  This thesis work does not lend itself to a single literature review, being comprised of several iterative phases of review, synthesis, and re-review. These contributions from the literature occurred at several points, in particular being embedded into the selected methodology as ongoing synthesis contributions. As such these iterations are broken up in this thesis Part to show the foundations in the literature behind the design decisions made, and to highlight the synthesised contributions to knowledge by this thesis prior to the empirical work.
    - Chapter 2: Literature Review.
      This chapter presents the core literature review establishing the nature of research on laboratories in engineering education. In particular, it identifies the inherent flaws in current research practices, and examines the impact of these flaws on the potential benefits of this evaluative research.
      Several key papers are either seminal in themselves, or exemplars of trends in laboratory evaluation research, and form the exploratory foundation of the synthesis work on which this thesis is based.
    - Chapter 3: Critical Review.
      This chapter further goes through several key contributing papers as case study vignettes, identifying their contribution to the motivations behind the later artifact development.

This chapter also details the 'research contortion' effect observed in experimental study designs.

- Chapter 4: Methodology. This thesis can be considered a meta-methodology thesis, examining the manner in which research on laboratories is conducted and studying more effective options. The methodology selected for this process is itself tied to the synthesis work being conducted, and this chapter describes that design process and the eventual methodology of both the thesis, and the contributed artifact. This process is iterative, with several distinct phases in the design process. This chapter presents the phases that were followed, as well as distinguishing between the methodology being proposed, and the meta-methodology being used to investigate it.
- Chapter 5: Phases. This chapter outlines the overall narrative which is followed by the studies to follow, within the structure of Educational Design Research. Each phase and its associated sub-research questions are mapped out, their success criteria are identified, and the outcomes of each phase which motivated and defined the scope of each subsequent phase are described.

- **Part 2**: Empirical Studies
While a significant portion of this thesis is literature synthesis work, empirical work was undertaken both in the design stages of the artifact, as well as work on validation of the 'final' version. This thesis Part describes that empirical work from methodological and findings perspectives.
  - Chapter 6: Collaboration Study. This chapter presents the major long-term collaboration study that was conducted in parallel with the development of the Purpose-First Approach and PIPOV, each iteratively improving and feeding back into the other through the active design, implementation, running, and evaluation of a technologically enabled laboratory.
  - Chapter 7: Artifact Development. This thesis is driven by the creation of an artifact, with the contributions to knowledge existing either side of this in the synthesis work developing it, and the empirical work exploring its use. This chapter describes the artifact itself, which is an approach to laboratory evaluation, a framework for that evaluation to be conducted within, and accompanying material to support that work. It also describes the contributing literature behind each component of the artifact. The Purpose-First Approach comprises the education design research methodology, coupled with communication surrounding the specific motivations for using the methodology for laboratories in engineering education (the 'why'), as well as more targeted and hopefully more approachable details for 'how' practitioners can utilise this methodology for their own laboratories. The PIPOV Framework is a visual model and tool for understanding, designing, and communicating laboratories in support of the proposed methodology, to maximise both the approachability of the methodology, and the benefits published research may yield.
As new technologies and teaching approaches have been developed, the lines between 'educational laboratory' and other teaching have become blurred. This line was investigated empirically with experts in the engineering education space, as well as iteratively presenting the earliest forms of the Purpose-First Approach and PIPOV frameworks for design experimentation.
Further iterative design experimentation was undertaken with the artifact components with experts and novices in the engineering education space. These

workshops served to refine the design and presentation of the artifacts for usability by practitioners.

    - Chapter 8: Field testing
      I: New Design. This section presents the initial phase of an ongoing project, using the current form of the generated artifact. This project is being conducted during the design of a new laboratory.
      II: Modification. This section is in parallel with the new design study, and looks at the application of the generated artifact to a modification of an existing laboratory.

- **Part 3**: Discussion & Conclusions
    - Chapter 9: Summative Discussion. While each phase of the iterative design process from synthesis to empirical validations includes discussion of the research findings and what these mean to the artifact and to an understanding of laboratory evaluation, this chapter looks at these holistically, examining the artifact as it exists at the end of this thesis work, and what knowledge exists around it as a result.
    - Chapter 10: Contributions and Conclusion. This chapter extracts and articulates the specific research contributions of this thesis, to differentiate them from the iterative design work that the thesis undertakes, before concluding the thesis overall. It acknowledges that the iterative process this thesis follows has no inherent 'end', and looks at what next steps exist both in terms of refining the artifact generated, and what further questions this research work overall presents for laboratory evaluation.

Part 1: BACKGROUND, METHODOLOGY, & SYNTHESIS

2 LITERATURE REVIEW

When answering the question of determining effective ways to design and evaluate laboratories in engineering education, a significant amount of synthesis and design based on pre-existing literature has proven necessary prior to any empirical study being possible. As a result, this literature review and background to this thesis work falls into some distinct themes:

- An establishment of the research domain and scope, that being laboratories in engineering education, and how they are designed and evaluated, as well as identifying the need for further work motivating this thesis.
- An exploration of disciplines external to engineering education to seek methods or techniques for resolving the identified issues in engineering education, described from the perspective of engineering education research.
- Supporting material that provides the necessary components for the subsequent synthesis and study design.

2.1 LABORATORIES

Laboratories, if one defines laboratory to mean direct investigation and experimentation with natural phenomena, have been used in engineering education since the inception of engineering as a discipline, with the earliest documented engineering education being master-apprentice study performed in entirely practical settings [6]. While the use of laboratories and practical education has waxed and waned against more theoretical approaches to education in engineering over time, it is generally held to be critical to the effective training of new engineers, and often features as a requirement for the accreditation of engineering education programs [2], [10].

There are a variety of reasons to include laboratories in engineering education curricula – direct pedagogic reasons, affective reasons, or even simple expectation. These broad reasons can be broken down into a variety of more specific intentions for laboratories, perhaps the most widely cited of which comes from Feisel and Rosa's work establishing the 'Role of the laboratory in undergraduate engineering education' [6]. This seminal work outlined 13 fundamental categories of laboratory objectives:

1. Instrumentation
2. Models
3. Experiment
4. Data Analysis
5. Design
6. Learn from Failure
7. Creativity
8. Psychomotor
9. Safety
10. Communication

11. Teamwork
12. Ethics in the Laboratory
13. Sensory Awareness

There is an important limitation that this list, while broadly sourced and likely to be a good representation of laboratories based on the questions asked, were specifically *student learning outcomes*, rather than objectives more broadly, and were specifically the learning outcomes that laboratories were currently being used for, rather than what they *might* be used for. It nevertheless provides a good baseline for consideration of what sorts of activities should be considered when investigating 'laboratories'.

Considering the extensive history of laboratories in engineering education [11], [12], [13], [14], the date of publication of this work is remarkably more recent than this long history would imply - 2005. This work was triggered by perhaps the most extensive paradigm shift in laboratory education in engineering – the inclusion of computing and networked equipment to deliver laboratory education in ways that were previously impossible. Minor, relatively low-impact publications had been made throughout the history of engineering education on laboratories, and questions of what they were for and whether they were successful, e.g. [12], [13], but these did not trigger ongoing publication interest or investigation by the engineering education discipline into questions of laboratory purposes or effectiveness. Most importantly, they primarily asked questions rather than answering them, hence the necessity of Feisel and Rosa's work.

These potential changes to what a laboratory in engineering education 'was' meant that questions that had been left largely unasked for decades were quite abruptly of significant interest. This can most readily be seen by a comparison of Ma & Nickerson's 2006 literature review [15] against a more recent 2021 review by Nikolic et al [16], and 2022 bibliometric analysis by Raman et al [17], showing exponential growth in research outputs on technologically-enabled forms of laboratories. The most fundamental of these questions being 'how can we determine if this new innovation is useful'. The changing nature of what was possible in laboratories has become one of if not the most significant driver of research outputs on laboratory education in engineering.

2.1.1 Laboratories in science education

While the focus of this thesis is on laboratories in engineering education, there is significant overlap in both practice and published literature with laboratories in *science* education. Similarly to engineering education's 'common citation root' seminal publication being Feisel and Rosa, science laboratories have a significant common citation root in Hofstein and Lunetta [18], and their continuing work in this space [19]. Similarly to Feisel and Rosa, Hofstein and Lunetta's work was motivated by new types of laboratories, specifically computer simulations – though significantly less so than in engineering, with more consideration being given to alternative aspects of laboratory design such as critical thinking and affective responses. Speculatively, this may be due to the core driver being interest by these academics, rather than an accreditation body.

Many key papers do not make a distinction between these disciplines – for example Ma & Nickerson [20] and Brinson [21]. This thesis does so, but only for consideration of scope and time rather than attempting to establish any such border. It is likely that engineering accreditation is the primary distinguishing characteristic, and so Feisel & Rosa's work [6] can be considered to define some of the scope of interest for this thesis. Nevertheless, significant review papers and citations in common among the background work for this thesis draw heavily on science education, and it is likely much of the synthesis and research work of this thesis will have relevance in that discipline also.

2.1.2 Technologically-Enabled Laboratories

There is a 'traditional' view of a laboratory – students file into a room with different sorts of engineering apparatus. Oscilloscopes and breadboards, beakers and Bunsen burners, microscopes, and slides, set up by groups of students working under the direction of some educator to investigate some natural phenomenon. It is this traditional style of laboratory that essentially defined practical education for the majority of the history of engineering education [6].

Just as they have done in every other industry and discipline in the modern era, however, computing and communications technologies have led to incredible innovation in what is possible for laboratories in engineering education. Engineering is of course a discipline inherently bound to technological progress, and the *equipment* used in a laboratory will change over time with the changing nature of engineering practice – such as migrating from the slide rule to the calculator to the computer [22]. The changes that triggered the new interest in the nature of laboratories however were changes not simply to the equipment being used, or the phenomena being investigated, but the fundamental mechanisms available for teaching and learning.

There are many ways in which technology might change the way a laboratory is or might be delivered, and it is almost certain that the limits of these potential changes have not been reached. Some helpful categories of technological enhancement can however be identified to give a foundation for discussion:

- Simulation
- Remote Access
- Core Augmentation
- Peripheral Augmentation
- Dispersal

These categories are a convenience invented by the author – they are not mutually exclusive, nor are they intended to be exhaustive – they are simply being used to introduce the scope of new laboratories more approachably being developed, and to provide a basis for discussion. They are intended to be broad enough to be somewhat resilient in the face of new innovations rather than present a specific taxonomy.

2.1.2.1 Simulations

Simulation refers to the partial or complete replacement of the natural phenomenon being observed with an artificial facsimile. This is most often run as a computer program that a student can interact with to see the outputs for given inputs, for example the outcome of mixing certain chemicals.

Computer simulations have been a feature of education from as early as the 1950s [23], and have grown to encompass a wide variety of software applications and supporting companies. This is an area that straddles the borders between 'laboratory' and 'not a laboratory', with some computer simulations being relatively divorced from the natural phenomena they simulate, and others being an attempt to faithfully replicate a physical experiment, down to the laboratory room itself [24], [25].

2.1.2.2 Remote Access

Remote Access involves equipment that exists in reality, that a student is able to connect to over some form of ICT equipment and control remotely.

This type of laboratory has its origins in internet-controlled robots, with applications in laboratories originating in the 1990s [26]. By using robotic components to manipulate equipment and providing a web-enabled interface, students can conduct experiments with real equipment remotely. These remote laboratories have been enhanced over time through the addition of features such as cameras, batch processing vs live control, and many more [27].

2.1.2.3 Core Augmentation

Core Augmentation is a broad term capturing several different technologies that change or replace the equipment being used, or the student's interaction with it. Computer-mediated equipment, augmented reality, virtual reality, various haptic technologies – these all change the way a laboratory is delivered and how students undertake them, through changes in interface.

Some elements present in remote laboratories appear without the remote aspect – a computer-mediated interface such as LabVIEW being used to control local equipment [28]. These were a significant driver for remote access in the first place, given once the robotic controls and computer interface were in place, adding internet connectivity was the only barrier.

More modern innovations include more sophisticated new interface technologies. VR headsets may be used to give an immersive, head and/or hand-tracking visual experience [29]. Augmented reality may be used to superimpose artificial elements onto a real-world environment [30]. Haptic technologies may be used to introduce new ways of controlling and sensing vs the traditional mouse and keyboard [31].

2.1.2.4 Peripheral Augmentation

Peripheral Augmentation differs from Core Augmentation in that it involves changes not to the core experimental equipment, but rather changes to the 'other' components of a laboratory as a teaching and learning activity. For example, the ability to communicate with educators or other students over the internet [32], or the use of smart tutoring systems to provide support [33]. As in the first cited example this can often occur in conjunction with core augmentation, given the mediated interface and innovation triggers already exist.

2.1.2.5 Dispersal

Dispersal is an outlier on this list in that it was *not* enabled by computing technology, having rather been enabled by pre-computer communications technology many years prior. The ability to send laboratory equipment to students, rather than students to the equipment, is something that has been enabled by using radio [34] and TV [35] communications technology, seeing a resurgence with the advent of the internet [36]. This typically involves sending a laboratory kit to each student, or the student building one with materials from home, with some form of peripheral augmentation to allow for access to teaching material (e.g., a broadcast, or web-based material), as well as direct communication with educators and peers.

2.1.3 Laboratory attributes

Laboratories do not simply vary along lines of their technological enhancement, of course. Various attributes exist which vary between different laboratory designs. The most prominent paper investigating these attributes is Nickerson et al [37], which breaks down laboratory attributes into several categories:

- Experiment
- Experiment Interface
- Coordination
- Coordination Interface
- Lab Frame

- Lab Technology (called 'Laboratory Mode' in some publications, e.g., [38]
- Individual Differences

Prior work by the author [39] explored these categories and revised them into the following:

- Layers: Infrastructure, Implementation, Integration
- Interfaces
- Experiment
- Data Source
- Coordination
- Synchronicity
- Individual Differences

Layers recognise the different meanings that may be applied to the term 'laboratory' – whether this means the platform some laboratory is presented by (infrastructure), the equipment being used (implementation), or the laboratory as a holistic teaching and learning activity incorporating not only the equipment, but also the instructions, the educators, the students, etc. An example of infrastructure would be a web platform for remote laboratories, e.g., the Sahara system [40]. An implementation would be the equipment and software necessary to conduct experiments and make observations, being anything from a collection of chemicals and glassware to a programmed micro-computer connected to robotic equipment. It is with Implementation that this thesis is primarily concerned, but of course this may only be arrived at within the constraints of the infrastructure and implementation that is designed as part of this, or exists prior.

The interfaces refocus Experiment Interface and Coordination Interface of the source Nickerson et al model [37], and in particular acknowledging that there will be further coordination with teaching materials and other elements with which the students will be interacting. This may be physical contact, or some form of computer-mediated interaction.

Experiment includes the instructions students are given as part of the teaching and learning activity that is 'the lab', the degrees of freedom the equipment provide, the phenomenon under investigation, the accuracy of the equipment, and everything else relating to the actual experiments being conducted and what they are conducted on (notably excluding the manner in which they are interacted with).

Data source is an arguably more accurate term for Lab Technology or Mode, and refers to whether the laboratory involves data coming from real equipment, or simulated. Notably it does not distinguish between traditional and remote laboratories where both use real equipment.

Coordination looks at the interactions between learner-content, learner-instructor, and learner-learner, as described in [41]. These interactions each have an interface as described previously, and establish what is possible for the students to do in the laboratory outside of manipulating the equipment itself (for example, discussing with peers).

Synchronicity is extracted out of the interface components in Nickerson et al [37] as a top-level attribute due to its significant influence on the nature of the laboratory itself. The experiment and all coordinations previously listed may be synchronous or asynchronous in nature.

Individual Differences capture the myriad of different contributing attributes each participating student will bring with them to any teaching and learning activity. From the very obvious differing physical accessibility requirements, or different levels of prior knowledge, to more subtle differences such as mood, each may have an influence on the student experience of the laboratory itself.

2.2 Laboratory Evaluation

The primary focus of this thesis is the design and evaluation of laboratories in engineering education. This is a topic that has had sporadic treatment in the literature, showing a varied level of depth and quality across different studies. There are three general categories that this research can be divided into: practitioner reports, comparative studies, and aspect studies.

Aspect studies are interested in some more specific attribute than the laboratory as a whole, for instance gender, or teamwork. They may be a more tightly scoped study looking at an aspect of a single laboratory, or they may be cross-sectional looking at the same aspect across multiple laboratory contexts. This thesis largely does not examine issues with aspect studies except where the aspect is being treated as an intervention on a holistic laboratory, or point of comparison between different laboratories, though it does use them as inputs for design decisions involving those aspects.

Practitioner reports are the local evaluations of laboratories by practitioners discussing what they implemented and (in good quality work) the evidence they have for its benefit. Practitioner reports show the highest variation in quality, ranging from a primarily technical discussion of how the laboratory was developed with anecdotal evidence of quality, or perhaps a simple affective survey for students (i.e., measuring how much students enjoyed the laboratory), as in [42], to more structured evaluative research as in [43], [44]. Practitioner research is critical to local improvement of educational practice, as well as forming the basis for wider-scale studies [45].

Comparative studies most often look at one or more of the technological enhancements described earlier vs some alternative implementation. These styles of studies were made particularly popular by the comparative aspects of the Feisel & Rosa paper [6] which contributed to the spark of research over the last 20 years – in large part due to the comparative nature of the initial issue which drove that paper's inception, namely the possible accreditation of remote and simulated laboratories and whether they 'matched' the outcomes of traditional approaches.

2.2.1 The Supertype conflation

One habit of this style of comparative research which was also investigated as part of the author's earlier work was the conflation of different laboratories based on a single primary differentiating attribute. These conflated 'supertypes' are proximal or hands-on labs, remote labs, and simulations [46]. While this does represent a flaw in research design and a gap in the overall knowledge of the discipline as a result, they are useful archetypes to appreciate given their proliferation throughout research in this space, particularly comparative research.

Proximal laboratories are what one may term 'traditional' laboratories – they comprise physical equipment manipulated directly by students in a laboratory room. The interface is hands-on without computer mediation between students and the equipment, and all coordinations are synchronous.

Remote laboratories are computer-mediated laboratories with real equipment operated at a distance. Students may be proximal to each other (e.g. in a room full of computers), or distant to each other as well. Coordinations may be synchronous or asynchronous.

Simulations are computer-mediated laboratories with a computer model of the intended equipment that exists only as a simulation. Again, students may be proximal or distant to each other, and coordinations may be synchronous or asynchronous.

The prior work [46] identified the ‘missing pairs’ of mode and interface, with computer-mediated proximal laboratories and haptic control of remote and simulated laboratories being ignored by the extant research in this space. Further missing elements exist when considering the list introduced previously – core and peripheral augmentation are captured by the idea of computer-mediated interfaces on proximal laboratories, but dispersal types of laboratories are not captured at all. This is part of a problem that exists in the modern era of laboratories in engineering education – the existence of new laboratories has found the definition of ‘laboratory’ lacking, and ‘what is a laboratory’ is an open question that will be addressed in part later in this thesis.

2.2.2 Research contortion

A distinct but co-occurring aspect of evaluative research which the prior work and this thesis challenge is the contortion of laboratory elements to an artificial parity with other laboratory elements for the purpose of comparison. This phenomenon will be revisited after the background into why it occurs and what impact it will have on studies is covered from contributing disciplines of research, but as an introduction, this contortion manifests in the artificial limitation of usually an affordance of a laboratory mode or interface. For reasons of experimental design, something that is possible for one laboratory being compared to another is disabled, or prevented, or chosen not to be implemented, because it should be a ‘like for like’ comparison.

The most common example of this is artificially limiting the affordances of a computer-mediated remote laboratory to match the affordances possible using a hands-on proximal laboratory. For example, a computer-mediated remote laboratory can usually be accessed at any time of day, multiple times by the same or different students, with much shorter sessions times due to students not requiring time for setup or teardown. However, when comparing such a laboratory with a more traditional implementation, students might be required to use it only during a scheduled laboratory session, with the same time provided as for the traditional version. This contortion is not in itself a flaw in a study, but it places limitations on the possible conclusions that are not always respected – for example, making determinations such as ‘remote laboratories are better/worse than traditional laboratories’, which is often the type of question being posed [21]. This will be returned to when considering context for studies later in this chapter.

While this thesis explicitly rejects the use of these supertypes, the literature that is published often falls along these lines, and so it is through this lens that the existing work on evaluation of laboratories will be considered. Brinson [21] undertook the most significant review of this comparative style of research to date, particularly focusing on simulated and/or remote laboratories being compared to more traditional designs. In the initial work Brinson reviews the existing literature for comparative studies, and then follows this up with further characterisations of these originally reviewed papers. Brinson’s work and some specific observations of what it implies about the nature of comparative evaluations of laboratories in engineering education are considered in more detail in the Vignettes section 2.3, in this chapter we will be looking at the papers being reviewed and what these contribute to an understanding of laboratory evaluations.

Brinson’s work represents a continuation of Ma & Nickerson’s earlier review [20], with a greater focus on synthesis than the earlier general review. Several relevant observations may be made regarding how laboratory evaluations are conducted – the first and most significant for this thesis is that laboratory evaluations are primarily conducted as single-run studies utilising purely quantitative methods. This excludes those published works that do not conduct effective evaluation, being practice papers which include a cursory treatment of ad hoc educator observations or basic descriptive statistics. Based on this grouping, 41 of the 56 reviewed papers were these purely quantitative studies, 4 were ad hoc practice papers, and 11 included some form of qualitative investigation. The qualitative studies reviewed largely did no fall into established qualitative practice – rather, the majority collected open ended data and then examined this in an ad hoc fashion.

Nikolic et al [47] present a more recent review that focuses specifically on ‘new laboratory implementations’, a term incorporating both technologically-enabled laboratories as well as laboratories which leverage more teaching methodology-based changes such as problem-based learning or flipped classroom preparatory work. As established by the earlier reviews and Feisel & Rosa of course, this focus is likely to have captured the overwhelming majority of publications on laboratories at all, with nearly all of this work being triggered by changes in the first instance.

The review presents a similar picture to Brinson’s – work on laboratory evaluation falls into lower-quality ad hoc practice papers, which present the laboratory development with cursory observations, usually from basic student surveys, with more rigorous work involving quantitative methods, usually analysing student assessment submissions, and resulting grades. Only a tiny proportion of high-quality papers reviewed – 3 of 16 – included separate qualitative data, with only 1 of these 3 using it for its evaluation work.

2.3 Quantitative Experimental Designs

The core motivational premise for this thesis is that the current practices of evaluation used to study laboratories in engineering education are not fit for purpose. This is fundamentally tied to the types of evaluation study designs identified as common in the prior reviews – quantitative experimental designs. Before the issues with these designs are examined, it is necessary to describe the specific nature of quantitative experimental designs, and in particular what requirements they pose, and what conclusions may be drawn from their results.

Quantitative experimental designs are designs founded on the collection and analysis of quantitative data. That is, data which either begins as or is converted into numbers. These numbers have mathematical procedures applied to them, primarily from the statistics subdiscipline, which lead to certain findings. This thesis follows the definitions in Creswell & Creswell [48], but the same nature and requirements of quantitative research have been generally established across many disciplines from medicine to social science.

The core of quantitative research is thus – if you can collect data by manipulating and observing reality in particular ways, you may then perform mathematical operations to that data, and the resulting numbers tell you something about the nature of reality you observed. The numbers may yield findings such as ‘no significant difference’, i.e. telling you the study does not conclusively tell you anything, or it may yield significant findings that may be interpreted as some elements being correlated (i.e. this tends to occur with that), or causative (i.e. this causes that). The most important aspect of this for this thesis is that one may not rely on the mathematical operations that are involved in the data analysis conducted **unless the requirements for their use have been satisfied**.

These requirements can be complex, and the nature of possible mathematics can similarly be complex. Fortunately, the general categories of experimental research designs give a shortcut from ‘what requirements have been satisfied’ to ‘what may be relied on when conducting analyses.

2.3.1 Categories of experimental design

Creswell and Creswell [48] define 3 major categories of experimental design: true experimental designs, quasi-experimental, and pre-experimental. When examining these categories supporting textbook-style and critique material will be used collectively which each give a holistic reference for experimental designs. These bibliographic citations were input for sections 2.3.1 - 2.3.5, with specific citations for points identified outside these resources, [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60].

True experimental designs are those which satisfy all of the requirements for quantitative analysis, within the defined scope of their application. There are many requirements for a well-conducted and robust true experimental design, but the majority of these (mitigating bias in survey questions, avoiding coercion, disclosing methodologies used, etc.) are common across all experimental designs. The 2 that are specifically relevant to this discussion are the need for a control group which does not experience the element under study, and random assignment of participants between the control group and the 'intervention' group. The fundamental requirement for a control group is that confounding variables that may *also* vary between the control group and the intervention group must be controlled, in order for the numerical findings to be valid for the intervention itself, rather than being possibly due to the intervention or possibly due to some other uncontrolled variable. By ensuring random assignment of participants in each group, nuisance variables, which are those which vary between participants but not in a correlated manner, may also be controlled.

By ensuring these 2 requirements (along with the other more general requirements for high quality studies), a true experimental study may establish *causative* and *correlative* relationships between the intervention and the observed outcomes, depending on the specific implementation, such as pre- and post- testing, between subjects, etc. The 'gold standard' for experimental designs of this type is the double-blind Random Controlled Trial, or RCT. Usually associated with clinical medical trials, this design is seen as the only way to quantitatively control for the stochastic nature of studies involving humans and appropriate control for confounding and nuisance variables. While interesting, these specific variants within the true experimental design space are not at issue for laboratory evaluation – the misalignment that will be examined occurs much 'lower down' the hierarchy of design quality.

Quasi-experimental designs are experimental designs which fail to meet the random assignment requirement. They are essentially a step down in the hierarchy of experimental designs, requiring only that a control group be used to compare against the intervention group. Designs which meet only this lesser requirement have not controlled for confounding variables, and the study is said to be limited. Limitations of studies prevent a firm finding of causation or correlation, because the mathematical approaches used to determine findings take the validity of their inputs as assumed – the term 'garbage in, garbage out' applies, where the processes do not apply any controls or manipulations to deal with such limitations.

Limitations are primarily dealt with through reader controls. That is, the limitation is acknowledged and described by the study authors, and interpreted by an informed reader as to how impactful it is likely to be to the results. This is an inherently qualitative process, with the limitations necessarily not able to be strictly defined in numbers – as, for example, measurement error +/- brackets can be in life science research [61]. Limitations are not absent in the earlier true experimental designs, but they are inherently embedded in quasi-experimental designs, as the lack of random assignment which defines this category is itself a limitation.

Pre-experimental designs, described in other publications [62] and for the majority of this thesis as 'pseudo' experimental designs (to acknowledge that in educational studies they usually do not pre-empt or precede a 'true' experimental design), are those studies which lack a control group. Under the assumptions embedded in quantitative research methods, studies which lack a control group cannot 'pass through' the experimental groups to make findings about the intervention. Rather, the findings apply specifically to the groups under study, and it is then up to the informed reader, having read the acknowledged and described limitations, to qualitatively interpret these to make conclusions about the validity of the findings as they apply to the intervention under study.

It should be noted that not all quantitative studies are experimental – there are myriad possible quantitative study designs, and more being actively developed, but the two most relevant for this thesis are aggregations, and quantitative observational studies. Aggregations make use of usually experimental study components, to analyse what each of these studies might say as individual data points in a larger aggregate picture – for example PRISMA [63] from the medical research discipline, which defines a rigorous process for conducting and reporting on quantitative systematic reviews and meta-analysis. Observational study designs are quantitative study designs that lack an intervention or treatment – for example, massive scale population surveys. These studies are able to identify correlations, and may be combined with other methods or studies to explore potential causal explanations for these.

2.3.2 Internal & External Validity

The above experimental design categories establish what relationships a study *can* establish, but they do not define what the study is establishing relationships about, nor do they establish the size of a relationship that might be detected nor how confident a reader might be about its establishment. This is reliant on the detailed specifics of the study from a design and implementation perspective. The combination of the overall experimental design, in addition to these specifics of design and implementation result in the study's internal and external validity – encompassing terms defined variously in the literature such as reliability, replicability, generalisability, type 1 and 2 errors).

Internal validity can be summarised as how correctly and appropriately the study informs a reader about the nature of the reality that was under study. Issues with study design that arise from experimental designs below true experimental in the quantitative experimental design hierarchy occur here. That is, the inability for a quasi-experimental design to establish a causal relationship can be expressed as a lack of internal validity; though this would usually only be expressed as such if a study incorrectly attempted to claim a causal relationship – a quasi-experimental study which claimed to establish a correlation may be entirely internally valid.

This mismatch identifies the essence of internal validity issues – they arise out of a misalignment between the question being asked, and the study which is attempting to answer it. At its core, any number which is observed says 'something' about reality, even if in the worst case all that may be validly stated about the measurement is that that was what was written down. When the mathematical/statistical methods used in quantitative research are applied to data which is not adequately prepared to support the assumptions embedded in them, or when conclusions are stated about the eventual numerical results which rely on those unsupported assumptions, that is when internal validity is violated.

External validity is the ability of a study to say something about contexts outside of itself – this might be expanding from a sample of a larger population to the population overall, or translating an intervention from one group to another. External validity can lead to conclusions in the external context up to and no higher than the internal validity that is present. That is, a study which has strong internal validity that has strong external validity can make strong statements about the external context, whereas a study with weak internal validity (e.g., a pseudo-experimental study that invalidly makes causal claims) can only make weak statements even if it has strong external validity.

Achieving external validity involves further requirements on the design and implementation of experimental studies. In essence it requires the study to be translatable to other contexts, which requires the context under study to be properly understood, controlled, and reported, such that a reader might identify whether this is able to be translated into their own context, and to what extent. External validity also feeds into aggregations – highly specific studies can be aggregated with other highly specific studies to make statements more generally. This is the impetus for systematic review papers, particularly in medicine and social policy [63], [64].

2.3.3 Experimental designs in education research

There exists a significant dissonance between the nature of experimental research, and its application in education research. This dissonance is highly variable between disciplines, with the discipline of engineering being highly likely to apply quantitative experimental designs when conducting evaluation [65]. It is likely that this is a result of the nature of the *engineering* discipline, in contrast to the engineering *education* discipline, and the methods that are appropriate for use there. Borrego [65] presents the most direct examination of this dissonance within engineering education specifically, both in comparing perspectives on different methodologies, as well as examining why quantitative methodologies appear to be the assumption within engineering education research, as is supported by the review findings described above.

Engineers are trained to apply the quantitative scientific method on fixed and measurable natural phenomena. This lends itself perfectly to true experimental designs – and due to the lack of human stochastic behaviours there is no need for double blind RCTs for valid findings. Engineering educators and engineering education practitioners are highly likely to be trained in traditional engineering, and so when conducting educational research, for example the evaluation of laboratories they are teaching with, they draw from their existing toolkit and apply the methodology they know [65].

The dissonance occurs as a result of the educational context being significantly different to the traditional practice of engineering – though notably not *all* of engineering as will be discussed later. 2 steel beams with common material and structure can be experimented on and compared with a reasonable level of confidence. Random assignment is not difficult to achieve, a control group which experiences no intervention is simple to establish, and the findings are likely to be valid for all steel beams of a similar makeup. Humans however do not lend themselves nearly so effortlessly to experimentation – no 2 humans are at all as close in nature as 2 steel beams might be, and their reactions are stochastic – that is they vary randomly at the local level, perhaps with order appearing on some aggregate scale. This dissonance of a trained engineer applying engineering research practice to engineering education research practice represents a significant risk to the validity of the work being conducted and reported, with limitations being present that cannot be controlled in the way they might be using an engineer's original skillset [65].

One key point that is under dissent in the literature is the nature of control groups in education research. One position is that control groups cannot be established – because a control group in education would be a group that was not taught in any way. Furthermore, to meet the 'gold standard' required of a double blind RCT as in clinical trials would require a placebo – that is, students believing they were being taught somehow when they were not [66]. If this position was upheld, it would limit all quantitative experimental designs in educational research to the level of pseudo-experimental design, leaving them unable to establish causal relationships for interventions.

Instead of adopting this perhaps more purist stance, the position may instead be taken that a control group may be established for an intervention in educational research by way of a ‘before state’, that is teaching being delivered in the way that it was before some intervention was implemented [53], [67]. This then allows for quasi- and even true experimental designs to be implemented in educational research, but these may only achieve internal or external validity if the design is such that the surrounding requirements on nuisance and confounding variables were appropriately managed. This limitation has been repeatedly observed across education research, e.g. [68].

2.3.4 Internal and external validity controls in education research

One critical method for this management is random assignment – which again is required in order to meet true experimental design requirements. Random assignment in education is an incredibly tricky task for a variety of reasons. Taken at the individual level, which is necessary to meet the level of a true experimental design, it would be necessary to randomly assign students to different classes.

Many studies instead opt for non-random sampling on the individual level, with students selecting or being assigned to classes as normal, but with random assignment of the treatment or intervention between classes. Alternatively, the difference between the perhaps-artificial control and the treatment can be year-on-year, though this raises still further questions around the appropriateness of the false ‘control’ group – there will undoubtedly be confounding variables that also change year on year between the control and treatment groups in this case [66].

It is possible to bypass these issues and create as close to perfect an internally valid study design as is likely to exist, by extracting ‘the study’ from the educational context it exists within. This is a similar spectrum as exists in HCI studies, comparing studies conducted ‘in the wild’ field studies and ‘laboratory’ studies [69] – in order to avoid confusion with the term ‘laboratory’ these will instead be termed studies ‘within’ and ‘outside of’ their surrounding context respectively. To illustrate what this might look like for a laboratory study in engineering education, consider a laboratory study which is part of a regular degree undertaken by students as part of their studies. It exists within its intended context, and is impacted by that surrounding context. Alternatively, students (or even more broadly recruited participants) may be recruited into a laboratory session run independently from any overall degree work or other teaching context, and participating in learning solely within the laboratory.

2.3.4.1 Context in studies: lessons from HCI

There is a large body of applicable work on the impacts of within and outside of context studies in the discipline of human-computer interaction. In the simplest terms, studies conducted within context have far more difficulty establishing internal validity – due to the messiness of the surrounding context and general lack of control that may be exerted by the study designers. Studies conducted outside of context have the greatest internal validity due to the intensive controls that may be applied, but due to being decontextualised lack external validity – the alternative contexts that one might wish to extrapolate or transfer the findings into do *not* similarly lack context, and so it falls to the reader to identify what the potential impacts of this un-studied context might be were the treatment or intervention to be applied to a more realistic setting [70].

There is no specific outcome from this of one style or the other being better than the other – simply more or less appropriate for the questions they may be applied to and, critically, what is necessary when reporting and later interpreting the results of such studies. In the main, these tend to fall down lines of abstract vs concrete, with abstract questions lending themselves to the generalised format of studies without context, while concrete questions asking about specifics, particularly implementations, being more suited to studies within context [70].

This has implications for studies on laboratories – following this example of HCI, a discipline which in conjunction with its larger parent discipline of design could readily be considered a more mature research discipline than engineering education and particularly on laboratories, it is thus likely that studies on specific laboratory implementations should be considered within context. This has a more judgemental impact on studies which involve the contortion of like-for-like comparison described in the earlier supertype conflation in section 2.2.1. Where a specific implementation is under comparative evaluation, it is invalid to artificially limit the affordances of that implementation when evaluating it, unless the question being posed and the reporting of the findings takes that into account. Certainly, it cannot be said to be a true evaluation of the laboratory under study – rather it is at best a study examining whatever aspect was allowed to vary. No examples fitting this possible valid design were found in the literature when conducting the literature review for this thesis nor do they appear in the cited literature review publications – likely because laboratory aspects are interdependent, and the effects of controlling one aspect cannot be established to be entirely independent of any other aspect.

2.3.5 The Assumptions and Requirements in Experimental Designs

Based on the above analysis of the literature, the following is a summation that this thesis uses as the basis for investigating quantitative designs in education:

- Education involves stochastic behaviours: an individual outcome is not predictable, but the outcomes of groups of individuals with similar circumstances trend towards some mean value. As such any study must be applied to groups using statistical methods in order to generalise the results across some population.
- A ‘perfect’ study, or ‘gold standard double blind randomised controlled trial’, takes a representative sample of some population, randomly assigns that sample between a control group, and a group for each intervention under study (i.e. a study comparing Thing A and Thing B would require 3 groups), and has the treatment/intervention given to each group applied in such a way that neither the participants nor the researchers (or e.g. the doctors delivering some treatment) know at the time of treatment/intervention who is getting what (requiring a placebo for the control group). There must be no systemic variance in conditions between groups other than the interventions under study. The outcomes of interest must be tested before and after the treatment/intervention. If these requirements are upheld, then the study is valid and the results can give indications about the treatment/intervention, including a causal relationship existing, i.e. the treatment was responsible for the outcome occurring.
- There is no known way to deliver placebo teaching (part of Kember’s primary argument [66]) thus there can be no control groups in education. At a minimum this means the novelty effect is always in play as a confound [71].
- The educator in an educational experiment delivering some lesson (as the equivalent of a doctor delivering some medicine) cannot be blinded – they have to know what is being taught and how in order to do so (the other part of Kember’s primary argument [66]). Thus, there can be no blinding of the researchers – and arguably no blinding of the students assuming they can freely observe what is being done.

- Without the control group and double-blinding, the only remaining protection against confounding and nuisance variables is the randomisation, which is only a protection under the assumption that *there are no systemic variances between the groups that are not inherent to the treatment*. This represents the most critical flaw in education research designs applying a purely quantitative experimental design, because there are so many possible (arguably certain) ways in which educational context variables might systemically vary between treatment groups:
  - Taught by different educators with different skills.
  - Taught by the same educator, but the order means one group is getting the first delivery of a class, and the other group is getting the second delivery of a class.
  - Taught in different rooms, or differently timetabled times (e.g. one morning vs one afternoon).
  - In the case of longitudinal studies comparing e.g. a change in teaching delivery between semester 1 and 2 involves drastically more systemic variance.
- There are also severe practicality and ethical reasons limiting truly randomised assignment.
- The only way to deal with potential systematic variance without the above avoidance controls is to mitigate it by measuring the possible sources so it is known whether any occurred.
- Replication by others is necessary to provide confidence that experiments have been run properly. Replication requires the recreation of the experiment with matching context variables to ensure the same result.
- Recreating contexts in education is similarly impossible – teachers are not fungible; student cohorts are not fungible.
- There are perhaps trivial or outlandish experiments in education which avoid these concerns, but these do not align with the sorts of research questions being asked of laboratories in engineering education.

Investigating the consequence of these assumptions to the evaluation of laboratories in engineering education and in particular the measuring mitigation is the basis of the following empirical work that led into this thesis.

2.4 Experimental evaluation: Argumentum Ad Absurdum

This thesis work represents a continuation and expansion of the same root question that was asked in prior work by the author. While the above background is new to this thesis, the initial conclusion that limitations existed in the application of quantitative experimental designs when evaluating laboratories is common to both – notably Feisel and Rosa [6], Nickerson et al [37] and Brinson [21] provided the initial indication of this issue.

One possible response to the establishment of this flaw in current research has been raised in several reviews, but most recently and specifically for this context in both Brinson [21] and Nikolic [47]. In essence this response is 'conduct quantitative research such that the confounding variables are investigated and controlled for to maintain internal and external validity as per 'true' experimental design rules,' and report these designs comprehensively. This response was attempted in a laboratory evaluation to assess whether this would be appropriate and resolve the observed issues.

2.4.1 Truss Laboratory

The laboratory that was the vehicle for this study began as a commercially produced piece of equipment intended to be run as a proximal, hands-on laboratory - Figure 1. Prior to the author's involvement in the project, this laboratory was converted to its initial remote laboratory form through the application of Arduino boards connected to a local PC and basic electronic motors - Figure 2. This was developed leveraging the Sahara Remote Laboratories system [40] to allow remote access.

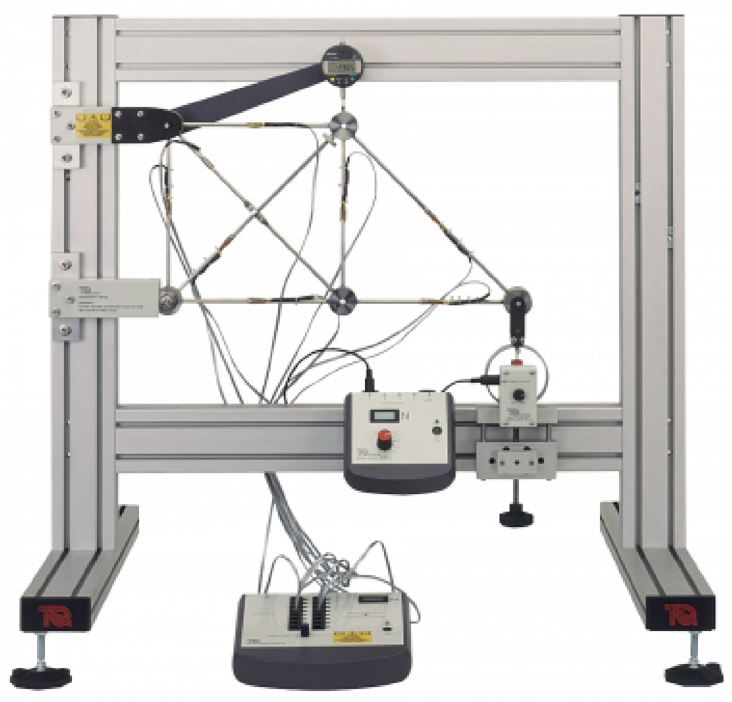


Figure 1:The unmodified, proximally operated truss [72] *via* [39]

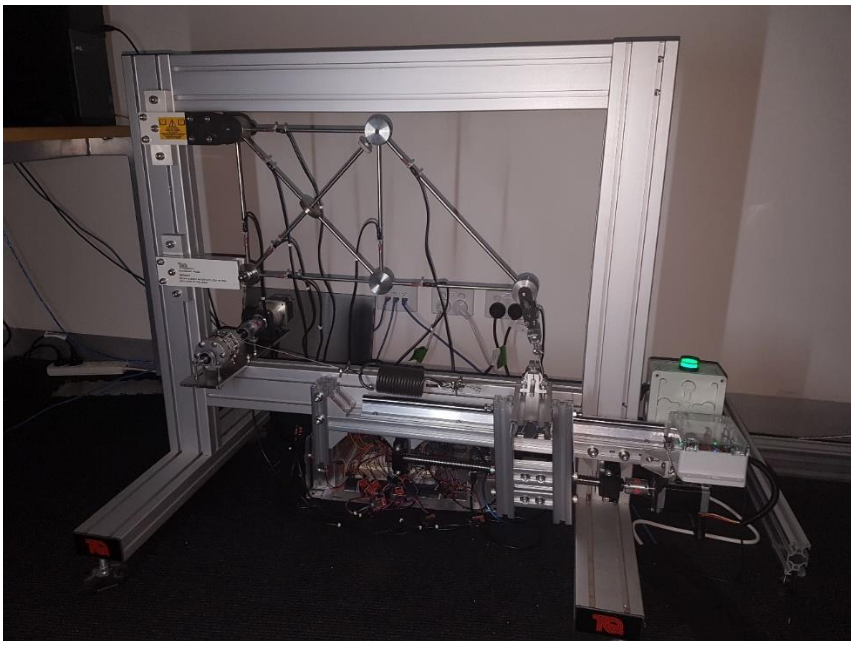


Figure 2:The truss rig during operation [39]

The truss was subsequently modified to include further degrees of freedom in the remote operation, and deployed by the author in an educational context for research. The specific details of this conversion are out of scope and may be found in [39], however the web conversion did support significantly more data capture far less invasively than the original proximal model. This deployment was conducted as a single-test pseudo-experimental design, intended to investigate the baseline responses of students participating in the laboratory. It comprised the students' use of the laboratory, completion of a laboratory worksheet, and several quantitative measures built into the laboratory web system for the study.

The purpose of the study can be broken down into a direct purpose, and a meta purpose. The direct purpose was to experimentally establish or reject a causal relationship between an observable student attribute (demographics, problem solving behaviours, or others), their behaviour with laboratories, and their assessment performance. The indirect purpose was to determine the validity of experimental designs to answering questions such as these.

By the letter, this experiment failed to answer the direct question it set out to ask, and no relationships could validly be established. Analysis of the initial constraints along with the observed behaviours of participating students is presented to support the prior conclusion that this failure was inherent to the nature of the question and methodology used, not the specific design of the laboratory study itself.

2.4.2 G*Power and multifactorial studies

The first challenge encountered when designing the study was the constraints placed on true experimental designs in order to fulfil the assumptions made by the mathematical formulae they comprise. These are described in the prior section 2.3.2 Internal & External Validity, but are applied in concrete form to this laboratory.

In order to answer the direct relationship question posed, internal and external validity would both be required. That is, a causal relationship or lack thereof must be identified (internal validity) while retaining the ability of a reader in another context to translate these findings to their own contexts to achieve/avoid the same results (external validity). There are some key mechanisms for doing so as previously noted:

- Random assignment of participants to groups comparing a difference in a variable of interest (independent variable) as correlated to some known variable of interest (dependent variable)
- Control of all variables not either independent or dependent (nuisance or confounding variables)
- Where these nuisance or confounding variables may not be controlled, they must be observed, and diverted into separate groups and a multi-variate approach used.
- Finally, if the maximum influence of a nuisance or confounding variable is known, and the observed relationship is larger, a causal relationship may still be established, however with damage to the knowable effect size.

Establishing a control group is left out as this would be necessary for establishing a *causal* relationship, and subject to the arguments previously presented in [66]. As will be shown, resolving this argument is unnecessary, as the lower internal validity of correlation may not be empirically met.

Consider the variable of age on academic performance. In order to establish a causal relationship between age and performance in a laboratory assessment, one must observe or control all variables of students other than age. For example, number of years of study can be assumed to correlate strongly with both age and performance, and this relationship is most likely causal. But in order to establish a causal relationship between age and performance one must either not allow years of study to influence the outcomes (e.g. use all the same years of study), or observe the values and mathematically control them (e.g. split up participating students by years of study and compare age vs performance within each group).

If none of the above measures are taken, then it mathematically may not be concluded that a causal relationship exists between the independent and dependent variables no matter their observed correlation. Further, this applies not simply to years of study, but to any variable that might influence the outcomes, e.g.,

- Instructor.
- Room the laboratory is conducted in.
- Time the laboratory occurs.
- Institution the laboratory occurs in.
- What the students had for breakfast that morning.
- Near-indefinite others.

It should further be noted that selecting a single state of a variable and holding it constant – for example the instructor delivering 2 versions of a laboratory, or the institution the laboratories occur within – maintains internal validity only at the expense of external validity. Knowledge may be gained about the behaviour and relationship between 2 variables under a given context, but no knowledge is gained about their behaviour and relationship were that previously controlled context to change. An institution might have a student culture that drastically supports the use of remote laboratories and have experimental outputs to show a positive relationship between remote laboratories and student performance, but another might have a different culture and reach directly conflicting conclusions.

There is one final method of handling threats to internal and/or external validity – to accept them, and to allow an informed reader to determine the likely impacts in their own context. This approach is often the one taken whether acknowledged as such or not – when a study fails to observe and report a given variable within an experimental design this is the default condition. The nature of these uncertainties is critical to qualitative designs and will be discussed further in the appropriate section below.

The influence of the above on the truss study was profound. In order to establish causal relationships merely between a single laboratory and a group of students, let alone the eventual goal of comparing multiple laboratory designs for effectiveness, the number of different variables to control for was immense.

- No direct variable controls (e.g. selecting only domestic enrolment males speaking Spanish at home aged 19 in their first year of study etc. etc. etc.) could be used, else external validity to any class not comprised solely of those individuals would be lost. Further, this was not practical for the study itself.
- Thus, all potentially influencing variables must be observed and controlled for.
- No experimentally established effect sizes of any variables of interest have been identified, for the same reasons they must be controlled for in the first place, and so no effect size control is possible either.

This inherently leaves multivariate study design as the only mathematically sound approach, to this or to any other study investigating causal relationships between variables in laboratories in engineering education.

Mathematically establishing multivariate studies may be performed using the software G*Power, which identifies the necessary participant group sizes to achieve the required p-value significance for a given expected effect size for a given number of variables needing to be controlled for and contrasted. This software immediately shows the absurd conclusions required to achieve the desired goals – limiting expectations to a mere 8 variables of the myriad possible, and a target detectable effect size of 0.1, a sample size of 2,283 students would be required [73], [74] via [39]. This is of course wildly *under*estimated, as the total number of potentially confounding variables in any classroom is far higher than 8.

2.4.3 Telescoping selection bias

Irrespective of the mathematical impossibility of achieving an answer to the direct question, and a seemingly purely mathematical conclusion to the meta question of 'no', the study proceeded with the reduced goal of identifying correlations of interest – not causal correlations, but identifiable correlations for future investigation, nonetheless. During this study, an attempt was made to observe as many variables regarding the students as could reasonably and ethically be achieved. In order through the sequence of participation in the study, these were:

- A meta-variable of sorts of students opting in to participate in the research on ethical grounds.
- Interactions with the laboratory equipment – digitally observed by the equipment itself for participating students without further student interaction.
- Laboratory assessment performance obtained for those students participating without further student interaction.
- An optional affective survey establishing student responses to the laboratory.
- An optional q-methodology study identifying problem solving behaviours.

A not unexpected result was observed during the study involving students' participation rates – the more questions are asked, action or thought is required, invasiveness increases, student participation drops. Of N = 132 students in the class, N = 121 agreed to participate and completed the laboratory, N = 29 further consented to any optional data collection, N = 14 completed all requested data. This phenomenon was termed telescoping selection bias, as it resulted in the final 14 participants having been selected along behavioural grounds rather than being a representative sample [39].

This telescoping selection bias must therefore apply anywhere attrition occurs with observation – thus requiring indefinite numbers of variables is likely to lead to indefinite attrition of participants. As a result, the original study concluded that a 'better' study design that resolves this attrition and can observe 100% of the potential confounds is unlikely in the extreme to be possible.

2.4.4 Dark Data

The above conclusion serves as the end of the prior work portion of this section 2.4. The following subsection made direct use of the collected data, but was completed subsequent to the conclusion of the earlier work and may be considered original analysis work of this thesis, with components previously published in [75] and [76].

One caveat to the expected observation that increasing the demands of study participation leads to a reduction in study participation is that increasing the observed variables that somehow *avoids* increasing the demands on participants will therefore be immune to this telescoping bias effect and thus ethical concerns notwithstanding may provide value without a participation cost. There are many sources of 'dark data' which either is or could be observed without participation demands, but the one of most interest in the context of non-traditional laboratories in engineering education is that of interaction data from computer-enabled equipment.

Subsequent to the analysis of the intentionally collected data, a review of the involved systems indicated value in the web logs being generated by control of the system. These web logs revealed:

- Correlations between queue size (number of users waiting in a digital line to use the laboratory equipment) and booking behaviour (users pre-booking timed access).
- Use and downtime patterns for the equipment.
- User behaviour with the laboratory once assigned – didactic, 'following the instructions' behaviour heavily contrasting with exploratory, 'investigate the relationships between degrees of freedom in the equipment' behaviour.
- Correlations between the above behaviours and wait time for the equipment.
- Correlations between all of the above, proximity to assessment deadlines, and number of user sessions.

In each case these observations could be made post hoc, with zero demands placed on the participants for their collection. The results clearly indicated design decisions that should be made to improve the laboratory given the observable proxy for learning of exploratory behaviour [77]. They were conclusions that lacked external validity, however the methods presented provide an alternative style of external usability – replication. A practitioner in another context need not make a decision solely on the research presented and their own context – if the observation demands are low, they might replicate the observations made and determine any correlations within their own context independently. This observation became a key feature of the artifact work in this thesis and will be referred back to in later chapters.

The details of the specific methods and findings used in that laboratory, while incredibly useful for that laboratory and interesting for applications elsewhere, are not relevant further to the specific material required for this background and instead may be found in the relevant papers [75], [76].

2.5 Evaluation in Other Education Disciplines

The problem facing engineering education in evaluating laboratories is far from unique. Even within engineering education itself it is likely that methods used to evaluate laboratories will map closely to methods used to evaluate other teaching and learning activities. There are however disciplines of education research that have a far more robust approach to evaluation than engineering education, which can inform the path engineering education might take in resolving the issues observed.

2.5.1 Science

Science education in this context presents almost bi-modal characteristics – as has been observed above research evaluating science laboratories is often entirely conflated into research evaluating engineering laboratories, with exactly the same approaches taken, and thus exactly the same issues preventing the work from being effective in yielding improved designs and implementations.

On the other hand, one significant difference to evaluation between science and engineering is the perception that engineering is a purely tertiary level subject, drawing on primary and secondary education subjects such as science and mathematics [78], [79]. While this is arguably an incorrect distinction in terms of the nature of engineering and science, it appears to manifest in significantly more literature studying science education at the primary and secondary levels than engineering.

Using a recent systematic review of evaluation of laboratories in science education from Gericke et al [80] as a foundation, the increased breadth of available studies is made apparent. Notably this paper is one of many citing the seminal Hofstein and Lunetta paper [18] mentioned earlier as a common foundation to much of the work on laboratories that are explicitly 'science' and not 'engineering'. This systematic review aimed to "identify important aspects of how to successfully make use of laboratory work as a science-teaching strategy in secondary schools" [80]. Simply in the stated research question aspects of integration with wider teaching are stated explicitly in ways not encountered in the identified literature in engineering equivalents. There is also explicit scoping of the term 'laboratory', which will be relevant in the later section 7.2.1 of this thesis examining the same.

The study uses 3 theoretical frameworks:

1. With the aim of developing students' learning of science
2. With the aim of developing students' learning to do science
3. Regarding the level of inquiry that facilitates aims 1 and 2

That is, explicitly stating the purposes of the contributing laboratories, albeit as broad, abstracted purposes as is necessary in a broad meta review.

The study identified 39 studies likely to contribute to the research question within the identified frameworks, and coded them according to the study design being used. There are two key observations that highlight the differences in the engineering education tradition and explicitly science education tradition on laboratories. First, the study correctly differentiates 'quantitative' studies into randomised controlled trials, specific quasi-experimental designs, correlation studies (pre-/pseudo-experimental designs), and simple observed numbers. The nature of RCT and validity is not critiqued, though as previously mentioned in section 2.3.3 this is an unanswered contention that is often accepted in research designs. Secondly, of the 39 studies, 8 used mixed methods triangulation, and 12 used qualitative methods. Qualitative methods were defined differently to that observed in [21] – where Brinson implicitly defines qualitative research as 'research which fails to meet the requirements of quantitative experimental designs' (i.e. quantitative affective responses without experimental controls are considered a qualitative method), Gericke et al in [80] describe qualitative designs as follows:

> "naturalistic observation concerned with understanding the process of students' meaning-making, interviews to capture students' and teachers' experiences, and analyses of documents used or created adjacent to exercises, such as teachers' written instructions or students' laboratory reports." [80]

More than half of the available studies had the rich description inherent to qualitative studies that enabled the synthesis work of the review to reach past the external invalidity issues of experimental designs and derive several outputs. In brief these related to the connection between level of inquiry and learning goals in laboratories as indicated in their starting frameworks, connections between science knowledge and skills, transfer of understanding between activities, and assessment. It is likely that these conclusions have relevant implications to engineering education laboratories also, but most relevant to this thesis is the establishment of what meta research is enabled by more appropriate methodologies in 'coal face' studies.

2.5.2 Nursing and wider health

Nursing (and to a less commonly published extent medicine, surgery, and wider health practitioner education) is perhaps an unexpected source of insight into the evaluation of laboratories in engineering education. There is however a strong link between the mechanisms used to teach practical skills in nursing, and technologically-enabled laboratories in particular. Speculatively this is perhaps because nursing must grapple with the issue of how you allow students to practice their skills and learn from failure, when failure in a healthcare setting holds the potential for significant negative consequences for the patient.

As such, nursing has for a significant length of time relied on the use of 'simulations' for teaching practical skills. These simulations are not inherently computer simulations as would generally be assumed when discussing simulations in engineering education – though in recent work computer and robotic elements have been introduced. Rather, these simulations meet the dictionary definition of the word 'simulation', in that they simulate issues with patients that nursing students must respond to as if they were real [81].

These simulations can involve the use of a human actor, or perhaps a stand-in such as a CPR dummy or more sophisticated pretend patient. Many parallels exist between these practical exercises and engineering education laboratories, and it speaks to the siloed nature of this discipline that the significantly more mature body of work available is not more heavily drawn upon for insights that might be transferred [82].

Two relevant aspects of evaluative research on nursing simulations were discovered when reviewing the available literature from this discipline – first, a direct attempt at the purely quantitative meta-analysis procedure following PRISMA [63] that is often upheld as the 'gold standard' from medical research when leveraging double blind RCT experimental design; and second, a series of example stand-alone studies that show the effectiveness of qualitative and mixed method study methodologies at performing these evaluations and designs.

2.5.3 PRISMA meta-analysis

Hanshaw & Dickerson [82] present a meta-analysis of literature available on nursing simulations, though limited to 'high fidelity' simulations which most closely match 'real' experiences. The PRISMA reporting protocol was declared as the overall approach, and 20 studies were identified, all of which were quantitative as is generally assumed by this approach.

Of these 20, at least 8 involved pre- and post-tests or simple correlations (pseudo-experimental design), 8 involved quasi-experimental designs, 2 claimed to be RCTs, 1 was a 'blinded controlled intervention' which most closely matched a quasi-experimental design, and 1 was a mixed methods study – however only the quantitative component of this study was used for the analysis in isolation. Some literature also used multiple quantitative designs.

Most of the studies involved were declared 'low' to 'moderate' in quality, which is a PRISMA standard to more systematically define validity, particularly internal validity. This rating was due to the inherent weaknesses in those study design types and attempts to implement RCTs in education as described earlier, while also being influenced by the single-site or context nature of those studies. Regarding limitations to the external validity of the analysis, the paper states

> "Studies in this synthesis were heterogeneous, with varied study designs, sample sizes, and simulation topics, which limits generalizability. A majority of studies reviewed were limited to convenience sampling in a single site in one geographical location." [82]

These are similar challenges facing any attempt to synthesise between studies in the engineering discipline, as will be demonstrated in the vignettes section 2.3 to follow. This example is shown not as an exemplar to follow, but rather to highlight the extreme difficulties with undertaking experimental design and subsequent meta-analysis in educational contexts. Nursing and the wider health discipline is the origin of randomised controlled trials and the gold standard of quantitative analysis of stochastic human responses, with practitioners more likely to be familiar with the requisite controls and assumptions required to be upheld – and yet to date this success has not been replicated in the equivalent educational evaluation context.

This conclusion is supported by other studies on nursing simulation evaluation with a focus on quantitative paradigms. Prion [83] presents an imported framework from the wider education research discipline, Astin's Input-Environment-Outcome framework for identifying variables of interest [84]. These are the independent variables controlled for in a study, environmental variables (this thesis typically terms these 'context') which are potential confounds, and outcome variables which are the dependent variables to be studies. The general conclusion of this paper is a mirror of the initial premise of the earlier described truss experiment – that these variables must be captured in order for a study to be robust. While there is consideration given to the difficulty of measuring specific variables, no consideration is made of the compounding difficulty of measuring all of the requisite context.

Foronda et al [85] present an earlier review to Hanshaw & Dickerson's, with a broader scope and 101 contributing papers. While not following the PRISMA approach and reporting substantially less detail to interrogate as a result, this review similarly calls for strict adherence to true experimental designs:

> "The discipline of nursing is lacking robust research and evidence that supports simulation as a method for nurse education" [85]

Noting that this lack is on the basis that robust research and evidence must be quantitative experimentation.

> "Consider a pre-/post-test design or use a control group versus an intervention group to evaluate if change or learning occurred." [85]

It should be stated that this review did make conclusions that are more readily aligned with the general conclusions of this background, in that *what* is being studied should be more robust than merely affective responses to teaching, and to identify differences in behavioural and cognitive outcomes.

2.5.4 Alternative Methodological Approaches

The discipline of nursing education as stated appears to be far more mature than engineering education, particularly when comparing nursing simulations to engineering laboratories. As might be expected, significant progress has been made in developing approaches that more appropriately deal with the challenges observed above. Several exemplars of methodologies and their successful findings are presented as representative of this body of knowledge.

2.5.4.1 Linguistic ethnography

Frandsen & Lehn-Christiansen [86] presents a qualitative research study using linguistic ethnography analysing transcripts from 'debrief' sessions. Debrief sessions in this context are conversations conducted after a simulation to reflect on the activities conducted, their correctness, further learnings, etc. A theoretical framework from Illeris [87], describing learning and competence development, was used to situate the analysis and present findings.

Outcomes of this process which were able to be empirically validated by the methodology applied included the influence of time pressures of the sessions on student learning, categorical differences in student motivations for learning, and the contributions of real clinical experiences to simulated learning. These outcomes are uninhibited by internal validity concerns caused by limited quantitative studies, and while they do not present a number to transfer to some other context, the rich description including transcript observations allows a practitioner in some other context to seek parallels in their own observations as a method of transfer.

2.5.4.2 Design-Based Research

The Design-Based Research approach, a version of which became the approach used in this thesis, was originally encountered by the author out of mathematics education research – that description is left to the detailed treatment of that methodology in section 2.6.5. Upon seeking saturation from the literature synthesis portion of this thesis however, extensive use of this family of research approaches was identified in nursing, and particularly for use in designing and evaluating simulation activities.

Koivisto et al [88] present the development of the 'NESTLED' model, a framework for educating simulation facilitators. This development followed an iterative design-based approach, specifically Design-Based Research drawing from multiple sources, including Reeves [59], Barab & Squire [89], and Plomp [90], among others. This process follows specific iterative phases with focus shifting between each [88]:

1. Systematic literature review
2. Analysis of current education programs
3. Development of a prototype
4. Testing of the prototype
5. Analysis of the test and refining of the prototype

These steps 1-3 generated an initial model, which was continually developed through to stage 5. In addition to this model, transferable design principles were also generated. For example:

> "Simulation facilitators [should] fulfil their roles as coaches and experts instead of 'deliverers of knowledge'" [88]

The data leading to the designed model and transferable principles were embedded in the design process over each phase. Literature synthesis, expert collaboration, and empirical observations of practitioner and student participants combine to form a mixed-method triangulation environment that is continuously analysed and responded to rather than being 'held static' to maintain experimental conditions.

Macdiarmid et al [91] similarly present a design-based research project in nursing simulations, in this instance Educational Design Research in the tradition of McKenney and Reeves [92]. This project, which is ongoing and reported in between iterative phases, examines and is in the process of designing authentic learning environments in graduate entry nursing programmes. The publication comprises a design through to first student contact subset of the larger project, and details the future phases yet to be completed. This mid-point to an ongoing project reporting is a common feature of iterative designs which might be under study longitudinally for years.

Even with only a partial completion, the iterative nature of the approach means outcomes have already been observed and analysed as inputs to each subsequent phase. For example, identifying incoming issues from the current state of practice that should be designed against, such as students failing to engage with pre-reading materials, leading to the inclusion of audio podcasts in the new design. Incorporating existing literature recommendations and expert collaboration supported the use of an active design to the podcasts rather than a more common didactic content delivery.

2.5.5 Other Disciplines

The above represents a point of saturation in this literature review for broadly different approaches to evaluation of educational activities in different disciplines of study. Similar tensions between quantitative limitations and qualitative approaches were encountered across several further disciplines. Notable examples of these include mathematics education, which as mentioned will be described as part of the educational design research section below, pilot training with flight simulations that shares many parallels with nursing simulations, and pre-service primary and secondary school teacher education.

2.6 Alternative Evaluation Research Approaches

Given the above evidence that the methodology being used to evaluate laboratories in engineering education is at best misaligned with – if not entirely unable to answer – the questions it is being applied to, there is a need for consideration of other methodologies that the literature suggests will be more appropriate.

Quantitative methodologies all fall into a descending hierarchy of what is able to be established in terms of relationships between laboratory attributes and their resulting outcomes. True experimental designs as described above being the only methodology able to answer questions such as 'given this laboratory attribute design choice, what will be the impact on student outcomes'. Other quantitative methodologies exist, but all fall into the pre- or pseudo- experimental category as described above, and cannot on their own answer the questions being asked.

Qualitative methodologies vary much more widely in terms of their ability to answer questions – as they generally embed the inherent, stochastic uncertainties present in human studies. Rather than enforce or contort activities to meet methodological criteria as in true or quasi-experimental designs, they instead observe reality as it is, and report the necessary detail for that observation to be interpreted by an informed reader.

2.6.1 Methodologies, methods, & instruments

When discussing qualitative and qualitative-inclusive methodologies, it becomes more important to distinguish between 'methodology', 'method', and 'instrument'. When discussing quantitative designs, 'methods' can generally be summarised as 'mathematical procedures' – a rich and varied set of tools, but all falling within that description. Thus, the distinction between methodology (experimental design; true, quasi, pseudo, between subjects, RCT, etc.) and method (the type of data being collected, and the mathematical procedures applied to that data) is a relatively simple one to make. 'Instrument' is also an important element, referring to the mechanism used to obtain data, or the specific observations that are then fed into the mathematical procedures. These can include surveys or questionnaires, computer-recorded data such as web or interface interaction logs, or any observation which might be converted to numeric data, such as tallying when certain behaviours were observed by an instructor [48].

2.6.2 Qualitative research

Qualitative research is differentiated from quantitative research by its analysis of non-numeric data – or in some cases analysis of numeric data using non-mathematical tools. This includes methodologies and methods looking at large bodies of text such as interview transcripts or observation notes, and other sources of rich data such as written descriptions, document artifacts like lab guides, etc. [48].

An important side note here is that a school of thought exists in some quantitative research publications that qualitative research can be defined as 'research which does not meet the standards of rigour for qualitative, experimental research' – essentially adding a category to the hierarchy described above: true experimental, quasi-experimental, pre- or pseudo-experimental, and then qualitative research. This is due to the nature of 'qualitative observation' which exists as a term in the life or 'hard' sciences – where a 'qualitative observation' might be 'the water got hotter', where a quantitative observation might be 'the water increased in temperature by 17 degrees Kelvin'.

While this is valid for use in the original context, it is by no means an accurate definition of qualitative research as it is properly conducted, and it would be better to use a term such as 'informal observation' rather than 'qualitative research' in such cases. This is critical to ensure both the appropriate level of respect for well-established and empirically validated methodologies in the quantitative research disciplines, but also to identify that informal observations which do *not* meet the appropriate standards of rigour necessary for qualitative research do not carry the same empirical backing as those methodologies and should be considered independently on their own merits.

Qualitative research includes an extensive range of methodologies, with some widely-used categories being: [93]

- Ethnography/Autoethnography
- Grounded Theory
- Phenomenology
- Phenomenography
- Case study
- Action Research

Of these the most relevant to the questions being asked in this thesis was action research, and formed the basis for investigation moving forward.

2.6.3 Qualitative Research Methods

Methods in qualitative research can involve the same instruments as in quantitative research, as well as some more specific ones – questionnaires can include areas for participants to write free-form text, observations can be written in prose rather than tallies or binary yes/no responses, or interviews might be conducted involving full text transcripts; many other possibilities exist. The specific *methods* that are then applied to the data thus collected cannot be so easily categorised as in the case of quantitative methods [48].

Some methodologies call for a type of 'quantitisation' [94] turning qualitative data into discrete quantitative data through mechanisms such as coding, or having an expert review quantitative data and mapping rich and varied responses into specific elements that multiple responses might match. After quantisation, quantitative mathematical procedures can then be applied.

Other methodologies leave the data as qualitative data, but process and analyse it in other ways. Thematic analysis follows similar coding practices but then applies sense-making to identify themes present across multiple points of data [95]. Case study analysis, such as was informally used in the vignette analysis above, review individual cases as holistic data points and discuss them as atomic elements rather than converting and aggregating them prior to analysis. Ethnography, phenomenology, and grounded theory each call for different types of rich description and analysis [48]. In each case, the methods used will vary considerably based on the methodology selected, and it may not be immediately obvious where the border between methodology and method lies.

2.6.4 Mixed and Multi method studies

Just as there is purely quantitative research, and there is purely qualitative research, there are of course methodologies which call for a combination of the two. These methodologies are generally composed of a quantitative methodology and a quantitative methodology, though options including a qualitative methodology along with combined quantitative and qualitative *methods* also exist [96]. The terms 'mixed/multi' paired with 'methods/methodology' are not used consistently or entirely accurately in the literature [97] – this thesis tends to use 'mixed/multi methods' as a catch-all as a result, with any relevant specifics or distinctions left to be described in the discussion directly.

Perhaps the most significant deviation is due to the literal meanings of the constructed terms themselves – a study which includes a purely quantitative correlational component followed up by an RCT experimental design fits the literal definition of having used 'multiple' methods and thus might be reported as having been a multi-method study, and this is valid for some contexts e.g. [98]. This thesis derives some working definitions from [99], summarised as:

- Triangulation defines a process by which combined perspectives on a phenomenon are used to identify the common reality between them.
- Within-methods triangulation is any triangulation which uses either only quantitative methods, or only qualitative methods (i.e. the above literal definition of multi-method is rejected)
- Multi-method studies, or between-methods triangulation, involve triangulation between at least one quantitative and one qualitative method.
- Mixed method studies involve direct integration of at least one quantitative and one qualitative method which occurs within the processes of at least one of those methods, rather than triangulating between finalised outputs of both.
- The more detailed definitions of e.g. QUAN+qual etc. from [99] contribute to the more abstracted archetypes discussed below, but the specifics of this taxonomy are not directly considered.

The benefits that can be derived from mixed/multi-method studies when performed correctly are drawn from the strengths of one method complementing the weaknesses in the other. This might be correcting for an issue of validity, an issue of explanation or causal identification, or establishing a complex narrative. Several versions of these lists exist, such as the almost algebraic version mentioned above, to a varied prose description in [48, Pt. I.1.2.2.3], to a typology analysis in [100] which are synthesised below.

Validity issues that might occur in a single method study can be mitigated by the inclusion of some other method. For example, a quantitative experimental design which encountered some uncontrollable confounding variable might triangulate using a qualitative method to investigate the likelihood and extent to which that confound might have influenced any results. A qualitative method which experienced the potential for bias due to use of e.g. self-reported interview data might triangulate with externally observable quantitative data to investigate this potential bias. In both cases research questions may be answered through the sum of the methods where they could not be alone.

Explanation is a common motivator for studies beginning with quantitative experimental designs. For illustration, consider a quantitative study which observed a significant difference in some dependent variable given a treatment. Quantitative experimentation can establish that this causal link exists, but it has no power to explain 'how' or 'why' this effect occurs. Qualitative methods may be used to investigate the identified relationship to examine potential mechanisms that it may operate through. Explanatory follow-up may also be used in other circumstances, such as investigating why replication failed in some other context, or some other expected outcome was not apparent in the quantitative analysis.

Establishing a complex narrative is a motivator that unlike the previous examples need not be reactive to some output, but rather may be baked into the design of the study from its inception. Investigating some phenomenon that is known to be rich, complex, and nuanced may call for multiple perspectives and sources of data to be collated and analysed holistically rather than by independent method outputs. For example, the relationship between political affiliations and some economic or demographic changes may be examined through qualitative methods, and may be examined through quantitative experimentation, but a full picture is unlikely to be obtained from either method in isolation. Instead, multiple contributing data sources may be compiled, and the phenomenon analysed across each of these together – without one solely being an input to some other, but rather each informing the other.

2.6.5 Practical Inquiry Methodologies

Having introduced some qualitative methodologies and methods above, there is a distinction that arises between the majority of qualitative methodologies, and action research. Most qualitative research methodologies articulate a general process to be followed, and the methods that support this. Ethnographies embed the researcher and have them observe and take notes, interview, and construct the deeper narrative within the notes and analysis. Phenomenology takes observations, interviews, and other rich data and analyses it for themes or structures within the participants' subjective experiences. In most of these cases, the methodology calls for a structure of a question being asked, or a topic of interest to be explored, a set of potentially applicable methods and instruments to be applied to answering that question/exploring that topic, and those methods involving procedures to analyse data to come to some discussion and/or conclusions [48].

Action research is an outlier in this set of methodologies in that while it still calls for a research question or topic of interest, and then outlines a process to be followed when answering or exploring, it does not have any specific declared methods or instruments that might be appropriate. It operates closer to the epistemological level, compared to the methodological levels ethnography and other qualitative methodologies operate at, but orthogonally opposed. Where epistemologies define philosophical strictures to do with what counts as evidence, how that might be applied towards the creation of knowledge, and how that knowledge might be interpreted, action research instead looks at the concrete mechanics of how different phases of research ought to be conducted, and what the constraints on the research should be – or more appropriately, where those constraints should be drawn from.

This different level of abstraction at which action research operates also means that the methods it employs can thus vary widely. It could feasibly be possible to design an action research project which incorporated an entirely quantitative, black box true experimental design – unlikely to be a *good* design, but the nature of action research does not preclude it. So long as that experiment fit within an iterative and contextualised investigation or intervention in some social situation or aspect of inquiry, this would meet the general requirements of action research. More realistically, action research will wherever relevant and appropriate include quantitative methods, in conjunction with qualitative methods. Within the different steps of action research, nearly any other qualitative methodology might be embedded – a phenomenological study might be used to explore participants experiences of something, which then leads to some intervention, which then is further evaluated using some survey or observation. Thus, action research might more appropriately be categorised as a mixed or multi method methodology depending on the nature of the contributing methods.

There are several related methodologies which follow the same style as action research, with specific structural differences. Various sources have attempted to deal with the conflation and overlaps to do with the terms used for these – method, methodology, paradigm, style, design, approach, etc. [101], [102]. The term this thesis will be following is largely 'methodology', though the artifact this thesis will develop uses the term 'Approach' given that it includes more definition and details than the more generic/abstract methodology from which it derives and implements. The term this thesis uses for the collection of related methodologies to action research is 'Practical Inquiry' methodologies – that term comes from the inherent 'involves direct and embedded change to the aspect of reality under study' that is the common attribute of these methodologies, and has previously been used in connection to action research specifically in the literature, but does not appear to have yet been used in the collective sense – it is nevertheless an apt term for them and aids in framing the discussion, in spite of or perhaps ironically in light of the use of practical inquiry as a term for laboratory study. Some sources use 'applied research', but far more broadly, and not accounting for contexts where other qualitative methods may involve direct application to changing reality in situ [103].

2.6.5.1 Action Research

Action Research as alluded to previously is a process framework for research involving participant collaboration and the live modification of authentic reality. Multiple definitions exist with significant variance in scope and details, e.g.

- "A type of applied research designed to find the most effective way to bring about a desired social change or to solve a practical problem, usually in collaboration with those being researched." [104]
- "Action research is a form of collective, self-reflective inquiry that participants in social situations undertake to improve: (1) the rationality and justice of their own social or educational practices; (2) the participants' understanding of these practices and the situations in which they carry out these practices. Groups of participants can be teachers, students, parents, workplace colleagues, social activists, or any other community members – that is, any group with a shared concern and the motivation and will to address their shared concern. The approach is action research only when it is collaborative and achieved through the critically examined action of individual group members." [105, p. 5] via [106] and

For the purposes of this thesis then, Action Research is the direct application of empirical observation, analysis, and reflection – that is to say research methods – informing actions that are taken towards achieving a defined change to meet a purpose or value. Rigour is achieved not through statistical significance or the mandated application of qualitative coding and analysis methods, but rather through its authentic connection with reality and lived experience of the collaborators.

Notably, in education in many cases the ‘group’ might be one single educator as the participant, collaborating with themselves in the role of both educational designer, deliverer, and researcher. This scales in breadth and arguably rigour to large groups pairing dedicated educational researchers with participating educators also acting as collaborators.

To give an example of this approach, consider a university which accepts mature age students, which has observed lower performance and affective responses from this cohort. In collaboration with educators teaching classes including this cohort, concerted efforts are made to identify and correct the causes of these issues. This is conducted by incorporating synthesis from the literature to identify aspects of concern, investigating those aspects, selecting some for changes, and observing the results of those changes. This observation might be ad hoc, follow some other qualitative framework, or be quantitative. The observations are analysed, and the researchers/participants reflect on these to form conclusions. This process may then iterate or reach conclusions to the research questions posed. Outputs to this process vary from ‘mid-way’ learnings such as “an effective way to observe X behaviour in the cohort was Y” to final conclusions such as “by applying technique A to classes the cohort’s performance and experience was improved”. Practitioners in other contexts may then make use of these outputs by either helping them to perform their own changes through action research, or incorporating concluding recommendations directly.

Several specialisations or refinements to action research exist around particular topics focusing on the empowerment of participants aspect of the methodology – social justice, decolonisation and indigenous knowledges, etc [107]. These may largely be seen as compatible with action research as a methodology, while incorporating more specific literature foundations, ontological and epistemological perspectives, and specific scopes. Their presentation as methodologies in parallel rather than being considered ways of approaching and performing action research speaks to the lack of clarity in the literature around the hierarchy of ontology, epistemology, methodology, and method.

There are commonalities between action research and the other methodologies examined as types of practical inquiry methodologies. Action research in many respects acts as a default or entry point for this discussion, being the earliest defined of these methodologies in the literature and having in some ways inspired the development of the others. Breaking down the observed aspects of action research above into common and differentiating yields:

Common

- Practical inquiry research involves the direct application of research methods to real, in situ environments, which are not extracted into a laboratory setting, simulated, or otherwise pulled out of the authentic context within which they exist.
- Practical inquiry research involves collaboration with local expertise within the context under study.
- Practical inquiry research may call for iteration to progressively approach some target result without ever mathematically provably reaching it.

Differentiating

- Compared to other forms of practical inquiry, action research focuses heavily on achieving the changes to a specific context, reported within that context, rather than focusing on extracting generalisable learnings
- The weight of evidence from external literature to practitioner expertise is weighted more towards the latter in action research.

- While not necessarily being part of the definition depending on the source used, action research more heavily favours certain types of change purposes compared to other forms of practical inquiry – social justice, participant empowerment and emancipation, etc.

2.6.5.2 Design Experiments

Design Experiments is an unfortunately termed practical inquiry methodology that is difficult to separate from the quantitative term 'design of experiments', the overarching terminology for true, pseudo, quasi-experimental designs, etc. It is nevertheless a well-established approach to research, particularly in education, where it is used to answer questions around best practice, develop theories for teaching and learning, and making predictions for other contexts [108].

As a result of the terminology issues, sometimes called "a Google Problem" [101, p. 28], several competing terms have emerged in later years that largely map to the same style of research and will be treated here collectively [92, p. 18], [101], [109, p. 4] :

- Design research (this and similar also conflicting with 'research on/about design')
- Design-based research
- Design studies
- Design research project
- Development(al) research
- Formative research/evaluation
- Design science research
- Engineering research (particularly problematic in this thesis' context)

Notably the term 'Educational Design Research' is also strongly linked to this family of methodologies, however the details based on specific traditions of researchers will be treated separately below as this has stronger relevance to the work in this thesis.

The fundamental motivation for the development of these methodologies is to leverage the benefits of a design process such as is found in engineering, manufacturing, usability, etc. for gaining an understanding of inherently unknowable phenomena. Particularly in education where much of this research has been developed, the issues with applying quantitative experimental designs examined above led to seeking an alternative methodology which could mitigate or avoid those blocking characteristics [101].

The core of a design experimental study is to take some complex system exhibiting some combination of stochastic behaviours, imprecise measurements or proxies, and/or uncontrollable confounds – i.e., a context not suitable for quantitative experimental designs such as education, and apply a formal design process to it. This is arguably something that has already been done outside of any research context as a normal practice for developing e.g. teaching and learning activities long before its development into a research methodology. This alignment with what is to many a common practice anyway is another motivator behind its development – the proposed difference being the formal documentation, rigour, reporting, and generalisation techniques that support taking internal, personal learnings about a context (the development of skill and expertise) and make them accessible to the wider discipline as validated and published literature [101].

This defining characteristic of design experiments from [108] is:

> "Design experiments are conducted to develop theories, not merely to empirically tune 'what works.'"

The distinction from the design 'work' that must be employed to deliver education frames the different design activities and provides a mechanism for identifying quality. These activities fit into phases which are broadly [108]:

- Developing an initial design
- Conducting the experiment
- Carrying out a systematic retrospective analysis

Written as such, these have no significant difference to a quantitative experiment which is designed, conducted, and analysed – the critical consequences emerge from established from design processes and their design experiment framings [101], [108]:

- The initial design of the experiment is not and must not be fixed in the face of mid-run evidence for change.
- Data collection includes the processes for the initial design and throughout the process – analysis of data, the processes by which this occurs, the decisions that are made as a result, are each themselves data.
- Theories regarding the 'why' behind hypothesised or actual results are an embedded and investigated component of the process.
- Unanticipated results often cannot be left as motivators for future work, but rather must be investigated as part of understanding the current design experiment itself.
- A common set of phrases from design which capture much of the above are:
    - "Progressive elaboration" [110], that is a design is not fully known at any given point and slowly emerges as it is used and investigated, and
    - "Satisficing" [111], [112], that is there is an inability to acquire perfect knowledge of any requirements or outcomes, nor is there achievable perfection in design, but rather that design involves an increasingly refined movement towards an increasingly clarified but never perfectly understood goal. Requirements can be 'satisficed' indicating that the design is 'close enough' within the ability of its quality to be measured.
- The systematic retrospective analysis involves the direct movement of the observations and conclusions local to the design itself to some wider context. Unlike quantitative conclusions which are generalised via their mathematical foundations, generalised conclusions drawn from a design experiment must explicitly deal with that migration of context away from the experiment itself – considering what was dependent on what circumstances, what inferences may be drawn and what evidence exists for these [101], [108].

One important element of design experiments is the nature of the data being collected. As mentioned when describing Action Research as an outlier vs other qualitative methodologies, practical inquiry methodologies including design experiments offer no mandated methods that should be used for data selection, collection, or analysis. The aspects they define are variously the framing that occurs (e.g. Action Research's empowerment of participants), and critically the phases that the research should follow. This can be seen in e.g. [108] which does not define the data to be used at all, and contrasted with [92, p. 28], Table 1.2, which identifies various quantitative and qualitative data. This presents a challenge when practicing these methodologies, as even the detail presented in [92] does not indicate or propose analysis techniques – with the exception of the quantitative methods which are stated as quasi-experimental designs, the other 'methods' as it claims are actually sources of data, with the complete method not discussed. This is reinforced by a presented quote:

> "Design research is not defined by its methods but by the goals of those who pursue it." [113] via [92]

Thus, as stated earlier practical inquiry methodologies can be super-methodologies containing what could otherwise be entire research projects as sub-methodologies. A design experiments study might embed some quantitative design, paired with a phenomenographic study, or a thematic analysis.

Some key aspects of design experiments are distinct to the Educational Design Research tradition examined below. The most critical is the nature of the context that a design experiment might exist within. The nature of the context along a spectrum of 'in the wild' to 'laboratory' studies [69], [70] is weighted far more strongly in favour of laboratory contexts in many design experiment styles than EDR specifically or Action Research. Design experiments must still be examined within a largely authentic context with real educator and student participants, but these contexts might be set up solely for the purpose of the experiment for the purpose of theoretical understanding, rather than being 'live' or 'production' reality as in the case for Action Research and EDR [101], [108], [109].

Generalisable theory being the explicit purpose of design experiments has the consequence of design experiment research leaning more towards fundamental research without as impactful or sustained a local change than Action Research and EDR. This is not an inherent aspect of design experiments as a methodology, but rather an opportunity it allows which Action Research and the specific tradition of EDR preclude; longitudinal studies where experimentation is conducted entirely in situ is still present in some design experiments [92].

2.6.5.3 Educational design research

Similarly to above for design experiments, Educational Design Research is an overloaded terminology that has competing definitions and is often used incorrectly or without connection to the prior literature on EDR as a methodology. As a result, this thesis uses a more verifiable approach of relying on practitioner traditions by way of direct literature citation.

The term 'traditions' is used in alignment with an earlier use in Creswell [114] vs [93], noting the change in term from 1st edition to 3rd. Like much of this research space terminology is particularly challenging, with tradition competing with approach, strategies of inquiry, variety, methods [93, p. 5]. I use the term 'tradition' to specifically call out that the source of truth I am using regarding a given methodology is not a derivable and provable mathematical construct as in the design of quantitative experiments, nor is it a term with a single dictionary definition matching its use in the literature with clearly defined and standardised processes and scope, but rather it is a fluid entity, that much like the research subjects it examines is itself being progressively elaborated, with best practices that are being refined and discovered over time, as evidenced by the varied discussions on terminology and practice in [92], [101], [108], [109].

The only feasible way of pinning down this fluidity into something concrete and communicable is to identify the literature source which may be consulted to see e.g. how the methodology works, where and how it identifies rigour, etc. This is communicated as "in the tradition of x", where x is a researcher or researchers and an anchoring citation provided. There is often significant overlap and co-authoring between the specific practitioners, for example van den Akker overlaps with McKenney overlaps with Reeves.

Several relevant authors of EDR practice were identified during the work on this thesis:

- van den Akker [109]

- McKenney [92]
- Reeves [59]
- Bakker [101]
- Cobb [108]
- Gravemeijer [115]
- Cochrane [116]
- Sandoval [117]

Note that this is not remotely intended as a comprehensive list of practitioners or developers of EDR – many seminal and prolific authors in the design research methodologies space which contribute greatly to each of the above works such as Collins [118] are omitted, and many individual EDR studies tend to draw from multiple sources. The intention is rather being to gain a broad sense of the variation in this space to thus be capable of making informed decisions when creating specific designs.

Educational Design Research in this thesis is founded on the tradition of van den Akker anchored by [119] and particularly [109], with contributions from McKenney, Reeves, Bakker, Gravemeijer, and Cobb, among others. This tradition has emerged primarily out of research practices in education design and evaluation research in the Netherlands, specifically mathematics education, with strong influences from research design experiments and related methodologies. The overall structure of Educational Design Research matches the general statements made regarding Design Experiments – with the same phases and focus on authentic practice. There are details to the construct framing of EDR that differentiate it and provide much more detailed guidance and structure, however. One of the stated goals of [109, p. 3] is to advance standards in EDR for improved quality, and so this provides a foundation for understanding the practices required.

EDR as described in [109] involves:

- An authentic teaching and learning context within which the research will occur.
- A local instruction theory: an understanding or narrative around what is going to occur through some specific teaching and learning activity(ies) teaching specific concepts, skills, outcomes, etc., which is a cognitive process the learners undergo or experience.
- A domain-specific instruction theory: an overarching framework for a domain or discipline of study (e.g. statistics) providing guidance on how students learn and understand the different elements within that discipline. It is generalised out of local instruction theories within its domain, and informs back into those local instruction theories. A domain-specific instruction theory may not yet exist in any given EDR study and must emerge through the synthesis of many studies contributing over many contexts.
- A 'designed artifact', which is progressively elaborated over the iterations of the study, which is the vehicle through which the local instruction theory may be observed [108] via [109, p. 107]. A designed artifact might be a teaching and learning activity plan, or a piece of educational technology, or a specified curriculum, etc.
- Iterative phases designing, implementing, observing, and reflecting on the authentic context, with the designed artifact and local instruction theory being progressively elaborated in tandem.

There is some difference in terminology depending on the scale of the research being conducted. This is most apparent in the 'local instruction theory' terminology, which applies to the design of teaching and learning activities [109, p. 17], and may vary when applied to different scales such as an overall curriculum [109, p. 67], though a local and generalised element of study will always exist. As this thesis is concerned with the design and evaluation of laboratories this scale is the most appropriate. In related work e.g. [92] the local instruction theory may not be explicitly termed as such, instead being referred to variously as e.g. research findings, theoretical contributions, etc., however this lacks the ability to distinguish between local, domain, and global findings. Local and domain-specific instruction theories are terminologies in heavy use in mathematics education [109], but are not inherently mathematical in nature.

The nature of the teaching and learning context depends on the specific goals of the research. In one design in the source material [109, p. 17] the goal is the generation and validation of a new local instruction theory, and that local instruction theory is itself an artifact to be iterated over. In another design [109, p. 52] the goal is the improved development of educational technology, and the local instruction theory is an indirect goal that informs how the edtech should be changed to yield better outcomes.

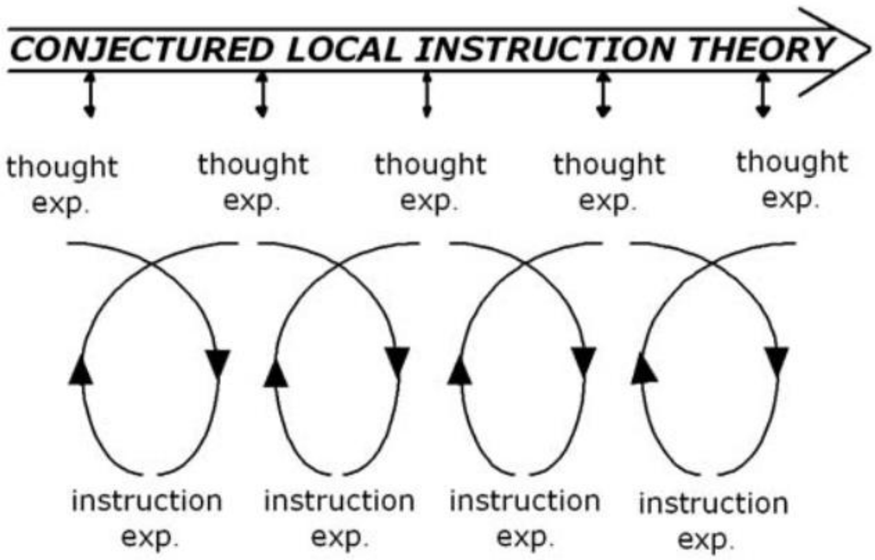


Figure 3: Iterative cycles in Educational Design Research [109, p. 28]

The process of conducting EDR involves iterations following the same phases as described for design experiments – as in Figure 3. An initial design of the artifact is developed, which involves conjecturing what the local instruction theory might be – this is in parallel with Action Research which similarly relies on the expertise of collaborating practitioners when creating designs. This initial design is implemented, delivered, run, etc., and observations are made during and after. Based on these observations, systematic reflection is made comparing the conjectured local instruction theory against the observed reality. Note the distinction between systematic retrospective reflection from the earlier discussion of design experiments and systematic reflection – EDR explicitly endorses continuous and context-situated reflection rather than waiting until the end of any given phase, shown in the diagram Figure 4 below with mini-cycles and macro-cycles. Some EDR work expands this further with meso-cycles [120].

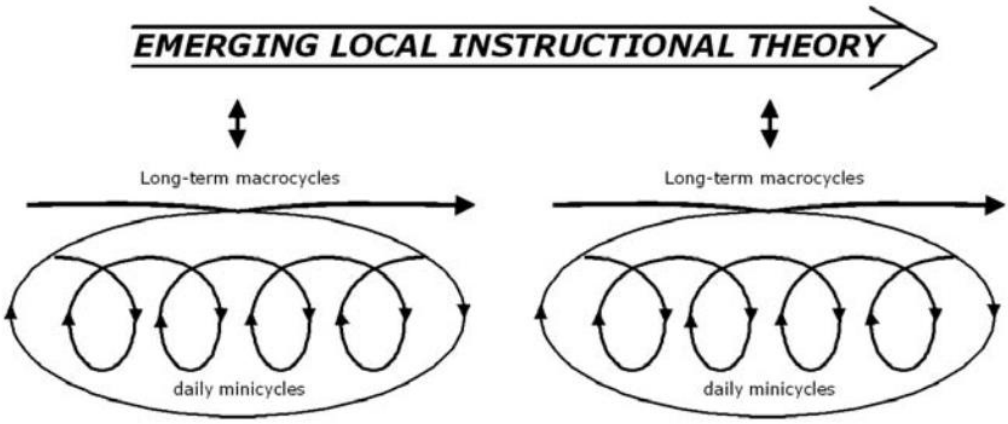


Figure 4: Macrocycles of iteration in Educational Design Research [109, p. 29]

As stated, practical inquiry does not mandate the use of any specific methods or sub-methodologies, even between the qualitative and quantitative paradigms – EDR similarly does not further specify this, placing process and framing constraints on the research design while being data identification, collection, and analysis agnostic. While the contexts to which it is applied preclude the use of causal quantitative experimental designs, correlation or descriptive quantitative designs may be components in an EDR study, as well as individual qualitative instruments or methods, and entire qualitative sub-methodologies. EDR often pairs appropriately with mixed/multi methodologies, as triangulation using varied data sources and methods is often necessary in response to aspects observed through the iterative process, as is the case in most exemplars in [109].

EDR is similarly agnostic regarding surrounding theories or frameworks. It specifies the process by which research phases should be conducted, identifies the artifact and instruction theories to be defined, but does not mandate any specific theoretical background to the research being undertaken, variously described in [109] as conceptual frameworks or ontological innovations. These theoretical backgrounds appear in EDR either as lenses through which the research data should be interpreted, or alternatively as instances of a domain-specific instruction theory being developed. EDR not specifying a lens of interpretation is a trait it shares with many qualitative methodologies, such as phenomenology [121].

2.6.5.4 Systematic Review

A Systematic Review of Practical Inquiry Research in Laboratories in Engineering Education

The purpose to this review of available methodologies is to identify alternatives to those that are commonly employed ineffectively to the design and evaluation of laboratories in engineering education. While these common approaches are reviewed above and will be examined through vignette case studies to follow, this was not a perfectly exhaustive review of all studies on laboratories in the discipline. Including a paper that will be left to the vignettes section 2.3 below for analysis, there are a handful of applications of at least EDR-adjacent methodologies to laboratories in engineering education.

Unlike the above review, which was conducted primarily via identifying high-impact papers and venues, citation chain snowballing, and requests to domain experts outside of engineering for relevant work in their disciplines, canvassing examples of practical inquiry research on laboratories was conducted through a formal systematic literature search. The following is an iteratively defined search query that was applied to the Scopus and Web of Science databases targeting the default search fields - article title, abstract, and keywords for Scopus, all metadata fields for Web of Science, notably excluding full body text in both:

"engineering education" AND ("laboratory" OR "laboratories") AND ("design research" OR "design-based research" OR "design based research" OR "design studies" OR "design research project" OR "development research" OR "developmental research" OR "formative research" OR "formative evaluation" OR "design science research")

This search query embeds the full list of terminology collated from sources describing design experiments and related methodologies presented above. No date ranges or other filters were applied to the searches, the search was maintained as an alert through end 2022. No venue quality judgements were made in the filtering process beyond inclusion in the indicated databases. N = 45 papers were identified for analysis below.

A separate query was also run for EDR specifically:

"engineering education" AND ("laboratory" OR "laboratories") AND "educational design research"

Which initially yielded no results – a subsequently alerted document is the subject of a vignette study in section 3.1.4. A series of more expansive queries were used omitting engineering education and/or including practical education and inquiry in place of laboratory, with no further engineering education literature found. Significant bodies of work in science and health education were returned however, some of which is presented in the earlier sections on those disciplines.

Of the 45 papers returned for analysis:

- 5 were excluded on review of title and publishing venue as engineering discipline literature not situated in an educational context.
- 23 were excluded on review of abstract as not relating to laboratories in engineering education, or not involving any form of practical inquiry methodology.
- 2 were excluded as they were conference proceeding entries with no available proceedings.
- 15 were reviewed as full papers:
    - 9 were excluded as not reporting a practical inquiry methodology in engineering education – the papers may have been reporting steps within a practical inquiry process, but no practical inquiry framework, iteration, or other context was provided to indicate this.
    - 6 papers were selected for analysis, where 2 were from the same author reporting on the same project.

Table 1: Selected papers on practical inquiry in laboratories from systematic review

1	Insightful learning in the laboratory: Some experiences from 10 years of designing and using conceptual labs [122]
2	What matters for learning in labs? - Experiences from designing for insightful learning in labs based on a symbiosis of American and European thinking [123]
3	Designing hybrid physics labs: Combining simulation and experiment for teaching computational thinking in first-year engineering [124]
4	Combining hands-on and virtual experiments for enhancing fluid mechanics teaching: A design-based research study [125]
5	Development of an innovative learning environment for engineering education [126]

6	Remote laboratory system for technology-enhanced science learning: The design and pilot implementation in undergraduate courses [127]

Papers 1 and 2 are both by Bernhard reporting on conceptual laboratories developed for teaching undergraduate engineering physics [122], [123]. The use of design-based research in this project leads to demonstrably enhanced outcomes, fully presents the theoretical and concrete contexts for the study, and is communicated effectively; furthermore, the papers are published in reputable venues and relatively highly read and cited. An interesting observation is that the author, while an academic variably in or adjacent to engineering education, has a strong connection to science education via physics, and it is perhaps not a coincidence that they demonstrate effective use of a methodology uncommon in laboratories in engineering education.

The specific methodology employed is design-based research, however this process began as early as 1994, and so is founded on the far earlier work by Brown [128] and the Design-Based Research Collective [129], a now-defunct research group founded to develop and promote the then-nascent design-based research methodology [101], [129]. The author is also known to have used the term design science (from Collins [130]) in [131].

A significant deviation appears relevant to explaining the lack of significant adoption of similar methodologies within the laboratories in engineering education context. The papers focus heavily on the laboratories as artifacts and the evaluation work conducted on validating their effectiveness, and do little to extract generalisable learnings that might be applied in other studies and other contexts. There is however significant discussion of the design process, and it may be speculated that researchers already more familiar with a design research process may have been better-situated to make use of this as a research output as opposed to focusing on the end-of-study outputs equivalent to more traditional engineering education research.

Paper 3, Fennell et al [124], is another engineering output with a strong influence from physics education, similarly reporting on an engineering physics laboratory. As with the above, the study is shown to have enhanced the laboratory artifact and effectively communicates the design, the theoretical context, and the observations. The methodological traditions followed were Barab & Squire [89] and Sandoval [117], both largely aligned with the Design-Based Research Collective [129].

Distinct from Papers 1 and 2, Fennell et al give significant attention to the generalisability of their analysis, identifying lessons learned [124]:

- Gradually Remove Scaffolding as Students Progress
- Avoid Compound Questions with Multiple Concepts
- Show Learners How to Leave Meaningful Comments
- Facilitate Reflection Through Open-Ended Prompts

The lack of significant adoption in this case may be due to time, with this work being far more recent. It may also have to do with venue, with the identified conference having been downgraded due to reduced quality soon after publication [132].

Paper 4, Gutierrez et al [125] has used a foundation of truncating Kelly's [133] statement:

> "Design-based research can contribute to engineering education research because it also draws on (1) a tradition of studies in mathematics and science education and (2) frameworks

> from diffusion of innovations and more recently (3) from educational data mining." [133], internal citations omitted.

As

> "Design-based research can contribute to engineering education research because it also draws on...from educational data mining." [125, p. 1]

And applied an entirely quantitative approach without applying practical inquiry structures to the study, leading to a separated design and research without the cycles of systematic reflection through an iterative process that are critical to these methodologies. There is some influence of Sandoval's conjecture mapping [117], but again these are presented in isolation rather than included as part of an iterative process. This shows an example of the nuance of design-based research that can be lost due to vague or loose terminology, especially in less rigidly defined versions compared to the previously described tradition of EDR.

Paper 5, Rumler et al [126], presents an in-progress study which switched to design-based research in the tradition of Plomp [90], who notably uses the term educational design research and is largely aligned with van den Akker, however this paper instead terms their methodology design-based research. The paper presents some foundational work motivating the use of a design-based methodology after a pure quantitative experimental design with surveying failed to answer the necessary research questions. However, the design-based component is reported only in the concept stage and has not yet been implemented in its teaching context. This is valid reporting for iterative studies, as observed in [134], however in this instance no further reporting on this project can be found in the literature.

Paper 6, Tho & Yeung [127], reports an educational technology-focused design-based research study, largely following Reeves [59] though with their own modifications. The paper situates the context of the study clearly, and in particular provides details on the inputs from prior work informing the initial design. However, its reporting of the reflection processes has a distinct separation of evidence and decision-making – the data and intermediate analysis is presented as it normally would be in a quantitative paper, then the decisions made about the design through the iterative process are presented separately, with only broad callback to the overall data. This prevents an external practitioner from seeing the specific observations that led to specific decisions, limiting the paper's generalisability.

The above narrowly scoped review provides confidence that there is no significant research community within engineering education publishing in a separate citation structure to the identified structure drawn on by this thesis. The selected papers show a lack of research maturity in the field in which quality work is not having the impact it appears to have in disciplines with a more established practical inquiry research community such as science or nursing education, as well as fundamental research design and/or reporting errors that are not being addressed.

It is near-certain that the above review is not exhaustive. The practical inquiry methodologies field as a whole suffers heavily from the observed 'Google problem' [101] (originally termed a 'Yahoo problem'), in that terms with far more common meanings are used to indicate specific avenues of inquiry, which in turn limit their ability to generate and grow a body of literature. Merely adding the identified "engineering research" term grew 45 papers to over 600, and other variations such as "practical inquiry" or "iterative process" were worse by several orders of magnitude. The limitation of search fields also excludes studies which apply a design-based research methodology, but only mention this in body text rather than a searchable field. Nevertheless, it has shown that there is some appearance of practical inquiry methodologies applied to laboratories in engineering education – i.e. there is no inherent barrier to their use – yet without sparking the significant adoption and growth of these techniques described in other fields.

2.7 Factors inhibiting the adoption of Practical Inquiry Methodologies in the Design and Evaluation of Laboratories in Engineering Education

The research environments described above show a stark contrast between the rich traditions of complex problems misaligned with quantitative experimental designs being investigated and enhanced through practical inquiry methodologies in educational domains such as mathematics, nursing, and to some extent science – compared to the far more limited literature available in engineering education. The reasons behind this difference are not readily apparent, given the seemingly well-aligned reflection of engineering design practice to the processes of design-based engineering and similar traditions, and the significant issues exhibited by more commonly used quantitative methodologies.

Investigating the literature for characterisations and investigations of this difference has identified proposed contextual drivers that may help to inform work with practical inquiry methodologies in engineering education research moving forward. It should be noted that none of these are specifically targeted towards laboratories, which leaves unanswered the question of laboratories in science education – an area observed above to have greater adoption of practical inquiry methodologies over engineering laboratories, but less so than science education overall.

Arguably the seminal work in this space is Borrego's "Conceptual Difficulties Experienced by Trained Engineers Learning Educational Research Methods" [65], referred as 'Conceptual Difficulties' and Borrego et al's "Quantitative, Qualitative, and Mixed Research Methods in Engineering Education" [135], referred as 'Research Methods in EE'. These papers investigate the journey into engineering education research by engineering education practitioners, and the methodological state of research in engineering education as it was in 2009.

Conceptual Difficulties [65] presents findings from a workshop which trained participants from either engineering or social science backgrounds in engineering education research – with the focus on difficulties experienced by those of an engineering background. These participants were archetypically engineering educators – people trained in the practice and study of engineering, who then became engineering academics expected to teach, as opposed to people trained in educating who then had to learn engineering. This background is proposed as a context for particular ways of thinking that were examined as part of the workshop observations.

The trigger for these differences is proposed as one of consensus – with engineering being a discipline exhibiting high consensus of though compared to a discipline such as psychology [136] via [65]. This also maps to the prevalence of different epistemologies, with positivism being heavily used for engineering with high consensus, and interpretivism being prevalent in disciplines with low consensus [137].

The observed conceptual difficulties from the workshop then included [65]:

- Transferability of results. This aspect is derived directly from the positivist paradigm of both science and engineering, and the universal generalisability of natural phenomena. Engineering academics have conceptual difficulties with appreciating the need for careful management of context for external validity.

- Theoretical frameworks. Participants experienced conceptual difficulties with the existence of multiple perspectives or lenses to interpret and analyse observed phenomena. Frameworks are given detailed treatment in a subsequent section 2.6.
- Operationalisation and Measurement of Constructs. This conceptual difficulty is tied to data collection methods and instruments, and participants' gravitation towards familiar methods of measurement and communication of results from engineering practice.
- Qualitative and Mixed-Methods Approaches. Again, tied to positivist tendencies, alternative epistemologies and their associated methodologies, and particularly the ways in which rigour can be derived in their application proved challenging.
- Interdisciplinary Collaboration. Participants were initially reluctant to view engineering education as an interdisciplinary domain that should take inputs from other educational domains, psychology, etc., as opposed to a subset or specialisation of engineering.

These conceptual difficulties are each in direct opposition to the effective implementation of practical inquiry methodologies – a successful study of this kind will require a strong understanding of the requirements for generalisability, will fundamentally rely on and be expected to contribute to theoretical frameworks, use more than simple quantitative instruments for data collection, will likely involve qualitative or mixed methods, and will benefit greatly from collaboration with other disciplines.

Research Methods in EE [135] presents a review of papers in the Journal of Engineering Education analysing the methods used. A significant portion of the paper describes the application of different methodologies much as the earlier literature review does here and so will be omitted – critically for this section however it then analyses the observed state that engineering education research was predominantly quantitative studies conducted as classroom experiments with an eye towards comparison between groups or longitudinally. The primary explanation presented from the literature being that engineering educators following a positivist or post-positivist epistemology and with greater training and experience generating and understanding quantitative data will gravitate towards quantitative approaches.

Notably neither of these publications deal with practical inquiry methodologies, except as a potential member of the qualitative or mixed-methods categories.

While these papers are indeed seminal towards the investigation of this space, they are increasingly aged, and particularly situated within the US engineering educational research context which is observably more insular than other areas of the world [138]. In support of this, Valentine [139] shows similar origins of engineering education practitioners from a more recent and global perspective, and Malmi et al [140] present a more recent methodological review of EJE, which identifies a similar weight towards quantitative approaches and low adoption of mixed-method studies. Again however, no explicit treatment of practical inquiry methodologies is given.

The adoption of practical inquiry methodologies is not completely unstudied. Kelly [133] presents an examination of design-based research in engineering education research, identifying some example DBR projects in educational technology applied to engineering, and alongside these examples describing opportunities for adoption, or ways to deal with barriers. These opportunities are only loosely related to design-based research, however, and are primarily concerned with the creation of a DBR practitioner online social network with associated big data and statistical methods activities rather than any practical inquiry methodologies.

Minichiello and Caldwell [102] present a literature review on DBR in engineering education, examining empirical studies which followed that broad approach. N = 24 sources were selected within the range of 2005 to 2019, with 15 applied to or including tertiary engineering study rather than k-12 – this broadly mapped to publishing venue also, with 11 published in engineering education specific venues.

Similarly to the laboratory-scoped review performed above, this review encountered significant limitations in the implementation of DBR in engineering education research [102]:

- Inconsistent reporting of key DBR elements – with studies variously failing to include research questions, descriptions of iterations, contexts, and other aspects without which a DBR study is of drastically reduced usefulness to an external reader, or in egregious cases made entirely invalid.
- Inconsistent reporting of participant demographics / Limited research with and for underrepresented groups – these are arguably contained within the first element, but is extracted to provide a spotlight on the opportunities DBR presents for work on inclusion and diversity that are not being undertaken.

Challenges to conducting DBR are also synthesised [102]:

- Managing the development of an interdisciplinary DBR team / Need for Expertise across Multiple Educational Research Approaches and Methods
- Funding
- Disseminating DBR Results, similarly to [92]

A consideration which is drastically understudied in the literature is the differences between engineering and science education. Earlier investigations of the literature on laboratories between these two disciplines by the author in this thesis and earlier cited work suggests that science education has a greater connection between tertiary and earlier (largely secondary and primary) education when compared to engineering education. This suggestion may be confirmed through literature on concerted efforts to increase the connection between tertiary and k-12 engineering as it is so much less prevalent than science [5], [141], [142]. As yet however no literature has been identified which examines the effects of this relationship in the other direction, i.e. the hypothesis that due to this difference in k-12 penetration of topics, tertiary science education research is more strongly influenced by k-12 education research practices (for example, mathematics education with its strong EDR traditions) than engineering. It is further hypothesised that this is the reason for the increased adoption of practical inquiry methodologies stated above, but as yet no investigation of this has occurred. Certainly no inherent aspects of the disciplines themselves that differ between engineering and science that would explain this observation are apparent.

2.8 Frameworks

A final component of background literature that is necessary to provide the foundation for the work in this thesis are frameworks. Frameworks in this thesis is an umbrella term that captures the usual formal term of 'conceptual framework', but has also been encountered as 'theoretical framework' or even simply 'model' or 'theory'. While these terms each have their own specific (unfortunately often contended) definitions that may only be pinned down in specific scopes [143], differentiating between them semantically is not critical to the work of this thesis and a local definition will be applied.

Frameworks are mechanisms for sense-making of complex interactions, attributes, relationships, systems, etc. They provide a way to accelerate understanding when applied to multiple different situations, by transferring structural understanding, and allowing the 'reuse' of this knowledge while 'filling in the blanks' with the specific circumstances under study at any given time [144]. This thesis places particular emphasis on frameworks which are represented graphically rather than merely described textually, however this is a result of the eventual trajectory of the artifacts involved rather than a limitation to the scope of what a framework might be.

This transfer of structural understanding may have several impacts on the sense-making process [144]:

- Directing thought to elements a framework has identified as important, relevant, or even merely existing, which may otherwise have been treated differently or missed entirely.
- Embedding concrete knowledge about some relationship or directionality between elements that would otherwise have needed to be established independently.
- Providing a method of fast and effective communication between practitioners who are both familiar with the framework being used.

Frameworks do come with risks – the 'concrete knowledge' described above is not accurate simply because it is embedded in a framework, nor do the relationships described necessarily exist. Frameworks are challengeable research findings in themselves that must be empirically established and validated, and may be appropriately or inappropriately applied [144].

Frameworks are typically applied to research questions which may not be answered from a purely positivist epistemology – more often following interpretivist, post-positivist, or other epistemologies which embed the idea of multiple valid descriptions of reality. This (i.e. with what perspective) combines with the extent to which a framework captures reality (i.e. how much) and the established validity (i.e. how accurate) a given framework may possess to yield multiple competing frameworks attempting to make sense of the same phenomena or circumstances. The relevant consequence of this is that a given framework need not, and depending on the driving epistemology most likely *can*not, be a pure, perfect, and singular model of reality – there is no ground truth framework that is being discovered, only communication tricks that are being developed to help one human understand something, and then effectively convey this to another human.

As such, frameworks require more than simply being true representations of reality; they are not end-result findings of fact that hold immediate value, they are indirect tools that must then be applied to arrive at direct and actionable knowledge. Thus, developing a quality framework is about developing something that is *useful* to the questions it intends to support answering. A framework which is empirically validated as correct for some phenomenon, but which is confusing to use or to understand outputs from, is a poor-quality framework irrespective of its empirical validity [144].

2.8.1 How frameworks are developed

Frameworks can be research outputs via several processes. They are often the direct output of qualitative research methodologies such as grounded theory. Work is done to empirically observe some phenomenon, system, etc., and a framework may be generated out of those observations, with the research methodology providing mechanisms to establish the validity of that generation [145].

Frameworks may also be generated at a distance from direct empirical observation, via literature synthesis. This process being considered separately is perhaps an artifact of modern research culture and practice, as the final framework generated from existing empirical literature is based on empirical data just as one in which both tasks were performed by a common research team, with the difference being simply administrative. It is highlighted to ensure the inclusion of these synthesis frameworks as empirically based.

Frameworks may also be developed through expert development – itself also a form of empirical observation, but highlighted independently due to its extensive use in education, and the relevance to the approach used in this thesis. In the following sections frameworks related to education are presented and discussed in terms of the way they were developed, and from where they derive their empirical validation.

2.8.1.1 Constructivism

Constructivism is a framework which describes the mechanisms by which knowledge and learning is achieved by students. It was and continues to be developed as an ongoing synthesis founded on the work of Piaget [146], and later Vygotsky [147] via Bruner, e.g. [148]], among many others working on refining and expanding the theory, and continues to be influenced and inherited from by further empirical research.

The essence of constructivism is that learners are not 'given' knowledge the way one might be handed an object. Rather, they 'construct' knowledge and integrate it into their minds or worldview through a process that can only occur within their own heads, and following their own context – albeit one which may be influenced through external stimuli. This is in opposition to the behavioural framework of Skinner [149] and is a widely used framework in pedagogy.

The development of constructivism began as the outcome of pedagogic research observation of child learners - indeed the term pedagogy having its etymological roots in 'child' education, many theories in this space have their origins in research on early childhood cognitive psychology, whereas the term 'andragogy' referring to the study and practice of adult education is sometimes proposed as a refinement for applications such as engineering education.

This outcome was Piaget's 'theory of cognitive development' [150], a series of defined sequences of cognitive maturation in children, up to a level where abstract concepts might be conceived. Notably Piaget's original theory is itself a framework that can be applied to supporting early childhood development. This theory described the changes in individual learners as they matured, through the construction of knowledge through play providing a foundation for future, more complex and/or abstract thought.

The development of this initial theory involved observational studies on child learners, deriving a theory for what was occurring, and validating that theory with new observations. While it did not explicitly follow a grounded theory approach (having significantly pre-dated this approach), it might be considered an early form of this method of deriving frameworks for understanding, following the disciplines in cognitive psychology of the time.

Adding to this theory of cognitive development were the theories of Vygotsky [147] – originally in Russian and contemporary with Piaget – later promoted more widely to an English-speaking research community by other translating researchers including Bruner [148]. The most relevant of these theories comprise 'social constructivism', which differs from Piaget's theories in that it examines the development of the knowledge itself, rather than the impact of the knowledge on the learner. According to social constructivism, learners construct knowledge through their construction and interaction with others socially and culturally. Knowledge is not isolated or independent to a single learner, but rather is jointly constructed by multiple interacting people – albeit with a unique and contextual 'version' being socially created in the minds of each individual.

Coming out of the same child cognitive psychology context, Vygotsky's theories were developed through observation and reflection much as with Piaget. The work of Bruner and the many others involved in interpreting and refining constructivism variously involved pure literature synthesis (e.g. translating Vygotsky's work in his native Russian language and incorporating them into Western research), as well as empirical validation, such as applying these theories to new, concrete contexts [151].

Constructivism in the modern era continues to see widespread use, particularly through derivative frameworks that apply it to more tightly defined scopes, such as constructive alignment [152] which will be detailed as part of this thesis' artifact creation. Unlike the other two frameworks presented here it is not inherently a visual framework, instead being a textual description of the way learners develop knowledge and understanding. It represents an exemplar of 'observationally derived' frameworks that are drawn directly from immediate empirical observation and reflection.

2.8.1.2 Bloom's Taxonomy

Bloom's taxonomy - Figure 5 - has become well established as a staple of educational theory and practice. It describes a taxonomy for learning objectives, fitting them into categories within a domain and hierarchy. Contrary to what may be assumed, however, it was not developed by Bloom following methods such as Piaget's and Vygotsky's above, but was rather the combined output of a committee of education experts.

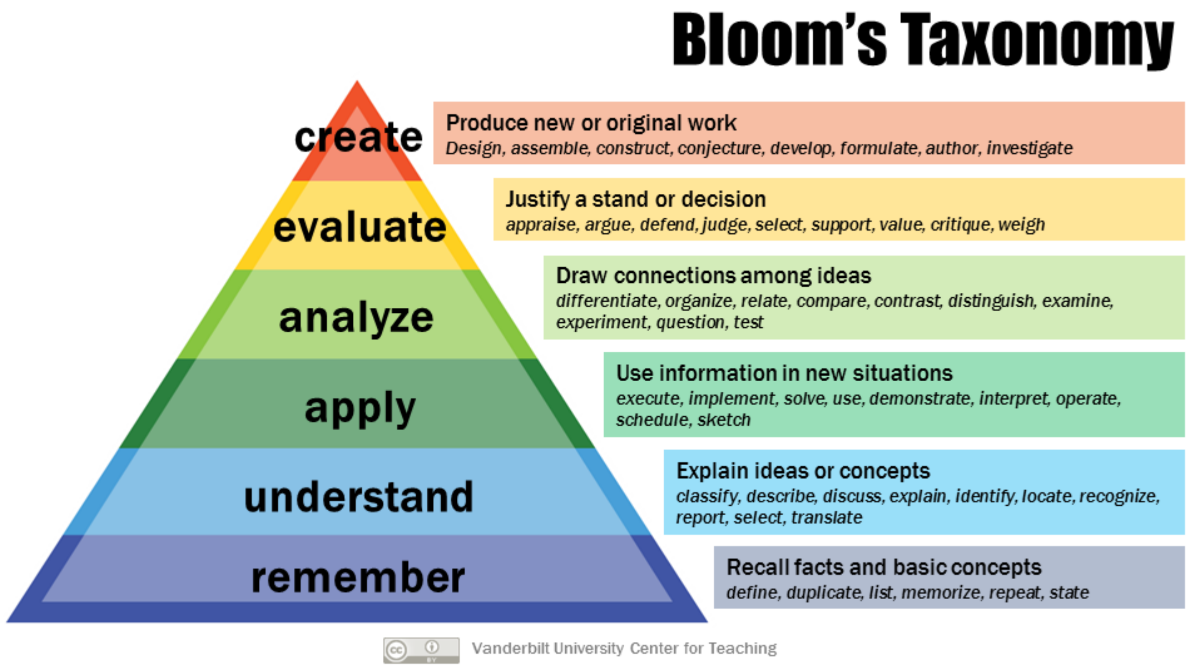


Figure 5: Bloom's Taxonomy, Cognitive Domain [153]

Bloom's taxonomy defines three domains: Cognitive, Affective, and Psychomotor. Learning objectives in the cognitive domain involve thought, knowledge, understanding. Affective objectives involve emotional responses, attitudes. Psychomotor objectives involve physical activity, skills, coordination.

Within each domain are different levels, in the above example in Figure 5 the Cognitive domain has 6 levels from remember to create. Note that several different versions of the taxonomy exist, the above image is from the 2001 revised version [154]. The levels represent categories of learning objectives, with each (atomic and well-constructed) learning objective matching a specific level of a specific domain. For example, a learning objective might be stated as “By the end of this lesson, students will be able to recall the dates of key transitions in Australian statehood history”. This learning objective refers to facts that students should ‘remember’, and thus aligns to the remember level in the cognitive domain.

Pairing learning objectives with their appropriate level is achieved through the use of particular verb keywords which become categorised into different levels. ‘Recall’ in the previous example pairs with the ‘remember’ level. ‘Explain’ as a keyword would likely map to the ‘understand’ level. This is only a loose mapping however, with different constructions of learning objectives leading to different levels, particularly where multiple verbs are used. For example, “Explain the significance of insulation in household construction using the theory of k- and r- values”. In this case ‘using’ can be replaced with ‘applying’, and thus this learning objective actually suits the ‘apply’ level.

The primary utility of Bloom’s Taxonomy in the first instance has been to establish learning objectives along the hierarchy expressed by the levels they represent. In doing so, a pattern of deeper, more impactful learning can be expressed, following a learning journey up through the different levels as increased mastery in a topic is gained. Being able to map learning objectives in this way allows educators to make judgements on relative foundation-intermediate-advanced degrees of instruction, interrogate the appropriateness of a given level of learning objective to a given situation, and communicate the different degrees of mastery attained by students who meet different learning objectives.

Bloom’s taxonomy in use can be divided across a broad spectrum of depth and expertise. It is common to see ‘use’ of Bloom’s taxonomy that might be stated as simply “write learning objectives using a verb and teach those things”. Moving deeper, being able to recall the different levels of the cognitive domain as English verbs rather than drawing any deeper on the literature on each. Continuing to deepen as the affective and psychomotor domains are included, and rich understanding and leveraging of the contents of each level of mastery are applied [154], [155], [156]. In this way the usefulness of a framework in slowly changing behaviours and practices is highlighted – unlike a mathematical algorithm following quantitative methods in which a slight deviation might see wildly different and incorrect results, doing ‘a little bit’ of a framework and slowly learning more and doing more with it can be an effective strategy.

While the idea of a hierarchy such that Bloom’s presents has been challenged in the literature [157], [158], this reflects the perspectives aspect of frameworks described previously, and Bloom’s Taxonomy and its revisions remain heavily used and studied educational frameworks. It represents an exemplar of ‘expert development’ frameworks that are derived through the iterative discussion and reflection of domain matter experts. Several competing and/or complementary frameworks to Bloom exist in the form of taxonomies such as SOLO [159], or Fink’s [160].

2.8.1.3 Kolb’s Learning Cycle

Kolb’s Learning Cycle describes a mechanism for learning to occur within an established, experienced person – that is to say it generally applies to *adult* learning, or at least a cognitively mature adolescent, rather than being derived from child cognitive psychology as in Piaget and Vygotsky [161]. This framework is a sub-framework which contributes to the wider framework of Kolb’s Experiential Learning Theory. Kolb’s Experiential Learning Theory includes Kolb’s Learning Styles Inventory, which unlike the Learning Cycle has seen significant challenges in the later years and is widely considered a teaching myth rather than something supported by the bulk of empirical observation [162], [163]. Kolb's Learning Cycle, however still sees significant use and study.

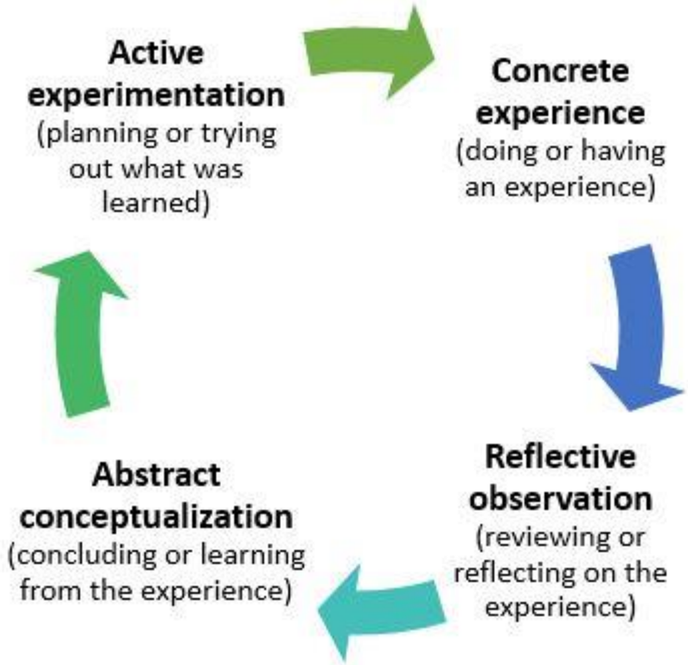


Figure 6: Kolb's Learning Cycle [164]

Kolb's Learning Cycle - Figure 6 - presents an individual undergoing some concrete experience in the real world. This could be anything from being told something by a teacher, to observing some scientific experiment, to hearing a piece of music. This experience is 'digested' through reflective observation, during which conflicts or gaps may be encountered when making connections between the experience and the individual's existing knowledge and worldview. Conceptualising those cognitive dissonances into some sort of eventual outcome, or perhaps lingering uncertainty, leads to active experimentation or investigation to test out new ideas or resolve uncertainty [161].

Kolb's Learning Cycle was developed as a synthesis of extant literature, some of which had previously been written by Kolb and his collaborators, studying adult learners in business environments. This is a good example of a perspective influencing the scope and nature of a framework – Kolb's Learning Cycles explicitly excludes the visceral behavioural responses that formed the basis of far earlier learning theories such as Pavlov's Conditioning [165] or Skinner's Behaviourism [149]. By focusing on white collar worker research subjects learning complex and abstract knowledge and skills, complex and abstract knowledge and skills are thus embedded into the framework itself.

Deliberately excluding the wider Experiential Learning Theory and Learning Styles Inventory, Kolb's Learning Cycle has been subject to significant empirical validation. Both from Kolb's original work and follow-up studies [161], [166], as well as validation by other practitioners [167]. It represents an exemplar of 'literature synthesis' frameworks that are derived through the combination of extant literature describing a phenomenon, albeit one subsequently validated through empirical observations.

2.8.1.4 Constructive Alignment

The final framework to be introduced into this discussion is Biggs' Constructive Alignment [152], Figure 7. This framework is a remarkably successful way to support academics designing teaching and learning structures from large scale degrees/courses down through individual assessments and specific activities, in a way that leverages constructivism theories to maximise the effectiveness of the teaching and learning effort.

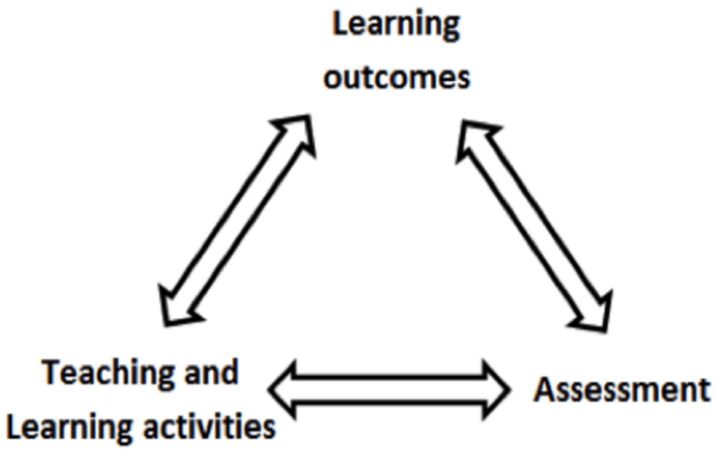


Figure 7: Constructive Alignment [168] *via* [169]

CA presents a view of the educational process that as an educator, one is designing a 'flow' that one's students are intended to flow through, from participating in some teaching and learning activity, being assessed on the learning achievement from that activity, and through this process being validated as achieving the 'Intended Learning Outcomes' or objectives [152]. This presents the classic triple element diagram depicted in Figure 7.

Pinning down the full details of a framework that has seen as widespread adoption, use, variation, and differing purposes as CA is difficult. For a given general purpose there are different elements which comprise CA. CA has been used as a design and implementation tool for teaching and learning activities, but also as a lens for conducting research into education, e.g. [170], and in addition as an input to refined, adapted, complemented, and otherwise derived alternative frameworks [169]. A key distinction lies in the depth of knowledge that must be included about the foundational *inputs* to CA, that is theories such as constructivism, while another is the depth of knowledge about the *domain of application*, that is the goals, contexts, etc. which are present in the discipline of teaching to which the framework is being applied.

A brief outline of the framework is presented below, with the acknowledgement that there is significant omitted detail within many of the stated elements – derived from [152]:

- Teaching students should be done for some purpose, target, or a vision of what the student will be capable of after successfully learning what is being taught.
  - The nature of this is a key point of variance for CA in practice, with competencies, attributes, skills, and other terms being used to define and characterise these outcomes, while remaining within the overall CA framework. Bloom's Taxonomy, SOLO, and similar are often used [152].
  - These Intended Learning Outcomes (ILOs) define the 'value' proposition of teaching, i.e. they define success or failure of the learning. If a student achieves the ILOs, then the learning was successful, and vice versa.

  - ILOs must have an active component – that is, they must describe the student 'doing' something in order to be valid. This is drawn from the constructivism and social constructivism theories and maps closely to the verbs used in Bloom's Taxonomy and similar frameworks. For example, "At the end of this class the student should have learned about 1920s architecture" is not a valid ILO, whereas "At the end of this class the student should be able to characterise 1920s architecture using classic examples" might be.
- The achievement of ILOs must be demonstrated using Assessment Tasks (ATs).
  - These ATs are only valid to the extent their achievement is a valid proxy for demonstrating the ILOs. For example, an exam which might be passed simply via rote memorisation of textbook statements cannot be evidence of achievement of an ILO at the "Apply" cognitive level in Bloom's.
  - Being a valid proxy for achievement is the 'alignment' between ATs and ILOs, with a fully aligned design assessing mastery of all ILOs.
  - An extension to this applies to assessments requiring mastery of *other* skills or knowledge not part of the ILOs, and the treatment of assumed or pre-requisite learning, however this is not covered in the initial CA framework and has different implementations in practice.
- Learning occurs as a result of teaching and learning activities (TLAs). Note 'as a result of' and not necessarily 'in', with self-reflection and self-regulated learning triggered by TLAs also being captured.
  - TLAs should prepare students to succeed at the tasks they will be expected to perform in the ATs.
  - Alignment between TLAs and ATs thus requires that no task may be assessed prior to the student undertaking TLA(s) which provide them the opportunity to develop the mastery sufficient to perform them.
- A form of transitive property applies here – if ATs are well-aligned to ILOs, and TLAs are well-aligned to ATs, then TLAs are well-aligned to ATs and thus alignment has been achieved.

The above defines the framework itself, and what attributes a valid, aligned design of some teaching and learning component must possess. From this, several applications of the framework can be derived in practice. When designing a new component of teaching and learning, it is necessary to 'go backwards'. Rather than designing things in chronological order of their performance, one should start from the purpose of the teaching and learning, state ILOs, design ATs which will demonstrate mastery of these, and finally design TLAs which will develop those masteries.

When critiquing a design to seek improvements, one should follow this backwards process to detect misalignments flowing from ILO to AT to TLA. Some extensions to this further seek alignment between ILOs and the authentic purpose of a particular context, i.e. curriculum alignment matching degree components within a wider stream, to global alignment e.g. 'are we teaching engineers the correct skills for their future lives'. Alignment is not the only component of CA however, and ensuring constructive, active engagement in the ILOs, ATs, and TLAs is also a necessary part of this process.

There is more detailed evidence available in the literature on the specific mechanisms for generating the CA framework than the older frameworks described above. The initial 'trigger' for the development of the overall theory was an experimental portfolio-based assessment used in an educational psychology class. The introduction of a portfolio submission with compiled examples of learning generated by the student contrasted with traditional exam and essay-based assessment to shift the nature of what was being learned from (paraphrased) ‘hear the positions and opinions of the teacher and repeat them in the appropriate format’ with no tangible impact to the students’ practices and worldview, to ‘reflect on impacts to our own practices from what was being learned’ which led to refinement and integration of new knowledge into that practice [171] via [152].

This trigger, identified as a form of objectives or outcomes-based assessment, was generalised from the specific experiment to be a framework for overall teaching and learning. This generalisation was largely a combination of literature synthesis and expert development on the part of the authors of the papers and books related to CA, e.g. [152]. However, it also included significant empirical development, as the framework was refined for usability, approachability, communicability, and continues through this iterative adaptation and refinement, e.g. [152].

CA, similarly to Bloom’s Taxonomy, is often used at different points along a spectrum of understanding and expertise. The most prominent artifact of the framework is of course the triple element diagram with the circular arrows – coupled with the English definitions of the words ‘constructive’ and ‘alignment’ this can form a basic understanding of what CA involves. An understanding of the psychological theories of constructivism, or an understanding of the different processes for validating that alignment has been maintained through different design and implementation processes is not a barrier to *beginning* with CA, which allows novices to gradually upskill their use of the framework over time [172].

An important caveat to CA is that is a framework for understanding and providing guidance on decisions, and is not a dogma to follow uncritically. This can lead to problematic adoption by managerial or policy-setting elements of education systems which has perception ‘blow back’ on the CA framework as a whole, yet is utterly unsupported by the framework and source materials themselves. This is perhaps a risk to frameworks which provide that element of minimally expert adoption [172], [173].

This framework represents an exemplar of a framework built on top of other frameworks, packaging largely existing but more abstract knowledge or frameworks (e.g. constructivism) into a more concrete and actionable form that does not require such a deep and nuanced understanding of the source disciplines such as cognitive psychology which underpin them. Interestingly this is a framework which has to outward appearances at least been developed independently, with Biggs’ constructive alignment coming out of a largely Australasian and Asian context, with Wiggins & McTighe’s Backwards Design [174] being an almost perfect duplicate with similar foundational frameworks largely out of a North American context.

3 CRITICAL REVIEW

3.1 VIGNETTES

The mismatch between the nature of laboratories as teaching and learning activities and the application of evaluation designs that are not appropriate fits for the constraints they present is a limitation to the internal and external validity of such designs. This limitation however is not simply a reduction in the validity of the conclusions, to be resolved using meta-analysis methods and other aggregations to identify the signal in the noise. This limitation not only leaves the evaluation research that is published with invalid conclusions, but has 2 other significant impacts:

- The constraints leave the laboratory activities under study artificially hampered, meaning that there is a further mismatch between the laboratory that is being studied, and the laboratory that would have existed had it not been studied.
- The publishing habits associated with the methodologies do not lend themselves to the natural interpretation required for integration in new contexts, or literature surveying.

To explore these effects, this section explores 4 research studies as vignettes to highlight the ways in which they occur and their impact. These vignettes all fall into a pattern of being focused on non-traditional laboratories – rather than a deliberate choice for this thesis, this is reflective of the pattern observed in Feisel & Rosa that continues to characterise laboratory research, that one of the key triggers to conducting such evaluations is the introduction of new ways of doing things, rather than evaluating established practice. 3 are studies following established methodologies and representative of the misaligned approach under discussion, and the final is an alternative approach which follows a different methodology to the established quantitative one.

3.1.1 Vignette 1: Comparative Study

J. E. Corter, S. K. Esche, C. Chassapis, J. Ma, and J. V Nickerson, "Process and learning outcomes from remotely-operated, simulated, and hands-on student laboratories," *Comput. Educ.*, vol. 57, no. 3, pp. 2054–2067, 2011.

The first vignette is a follow-up to one of the key papers used as the foundation for this work, the identified impacts being one of the drivers for the initial exploration of alternative approaches. This paper is the most relevant in a series of publications by the same group developing both remote laboratories and the associated model in [37]. It is highly influential in the research space, ranking among the top papers by citations on laboratory evaluation.

It should be clearly stated here that this analysis will be quite critical of the study design and interpretation, for this and for subsequent vignettes. This should be tempered with an acknowledgement that this was unchallenged as best practice research design at the time of publication (and arguably today, hence the motivation for this thesis) – whether due to habit, existing knowledge, or publishing venue methodological preferences/requirements, criticism is reserved solely for the methodologies and not the researchers making the best use of the tools available.

In this research the design was a 3-way comparison between a traditional hands-on, proximal laboratory, and equivalent implementations via simulation and via remote operation. These 3 versions were assigned to different lab instructors with discrete sections of students – students also varied as to whether they performed the 'data collection' portion of the laboratory activity alone, or in groups (the overall laboratory activity was done in groups regardless).

There were 2 key outcomes under study – learning outcomes, and student satisfaction. Learning outcomes were measured through comparing student prior GPA, compared against a post-laboratory knowledge test, using a between-subjects ANOVA approach. This knowledge test also gathered data about processes during the laboratory activity, such as time taken and group work behaviours. Student satisfaction was measured using a single post-laboratory survey. All reported empirical data was quantitative.

This study must be interpreted as being pseudo-experimental in nature – there is no control group. At best if one considers the 'traditional' laboratory to be a non-intervention with remote and simulation being interventions to be compared against, the study would at best be quasi-experimental as there is no true random assignment (students do not get randomly placed in laboratory timetable slots, and the effect of different instructors for different groups is not random).

There are several observations to be made about the study design and findings. The first is the identification of a process anomaly via the knowledge test – software installation required by the simulation version left a distinct spike in time required to setup, partly due to requiring installation from CD-ROM in-class, which left the remaining time for subsequent required tasks reduced. This was due in large part to the artificial constraint in the nature of the laboratory activity between modes – a 2-hour runtime requirement was fixed for all modes, the initial instruction time was kept the same, and the requirements for collecting data were kept the same.

The next is various observations regarding interesting and/or unexpected results from groupwork process results in the knowledge test. A sizeable minority of students instructed to work individually in the hands-on mode responded that they worked together. This was interpreted as:

> "Psychologically, this suggests that students in the hands-on sections were more likely to judge that they were "working together", even when they were individually collecting readings with different parameter settings. Of course, this may have arisen because in the hands-on condition, all group members must set up and share one apparatus, and tend to remain co-present during data collection even under individual data collection instructions, while in the computer-mediated formats three separate computers were usually available." [175]

This is an interesting line of inquiry, but was not able to be pursued under the study conditions a) due to only discovering it in the post-test phase, and b) needing to maintain experimental conditions and not modify the laboratories under test partway through the study. It should be noted that there is no reported empirical evidence in the study for the interpretation given – this is an example of the inherently *qualitative* interpretation of quantitative results required when dealing with such contextual, stochastic, and complex entities as teaching and learning activities. The lack of local empirical evidence for the interpretation of quantitative results is an inherent feature of the purely (pseudo) experimental design being used.

An alternative and (given the data reported in the study alone) equally supported speculation might include that students did not follow the instructions they were given to the letter, or that students misinterpreted the definition of 'data collection' in the question being asked and conflated it with other activities where they were expected to work together. Without more detailed (and preferably contemporary) observations these are impossible to differentiate. This is somewhat supported by a comment elsewhere in the paper:

> "apparently because lab team members tended to gather around one lab equipment setup and watch each other collect data, even in the individual condition" [175]

In a similar question, students working on the group collection hands-on mode were far more likely to all participate in the collection rather than designate one group member, with the interpretation being:

> "It may be that in the hands-on format, interacting with the lab apparatus was seen as especially beneficial educationally, or especially enjoyable, thus lab team members were careful to share this role. These differences in behavior among the lab format conditions should be taken into account in interpreting the superior learning outcome for the group hands-on format." [175]

Another line of inquiry that appears critical to understanding what is occurring during the laboratory activities, but which is not investigated as part of the experimental design. One alternative interpretation is that the hands-on laboratory version lent itself to requiring multiple hands doing simultaneous work, or that students recording dictation from someone operating equipment saw that as doing data collection (whether or not the study designers would agree).

The primary results of the paper centre around statistical analyses of learning outcomes and student satisfaction. Both measures are compared between groups (group and individual cross hands-on, simulated, and remotely-operated). In the case of learning outcomes, group hands-on and individual remote were found to achieve higher results, with hands-on in general being better than remote, which was better than simulation. These are arrived at using the comparisons of means and ANOVA methods named earlier. These conclusions are however deeply flawed, with the root cause of this flaw being approached by the paper itself in the comment:

> "Thus, it is not completely clear whether the advantage for this condition [hands on over remote and simulated] is due to educational superiority or to simple familiarity, which could have enabled the students to concentrate more on the concepts and procedures of the lab and less on understanding work-process instructions and negotiating social roles." [175]

Again we see an interesting qualitative speculation that lacks further investigation and is unsupported by qualitative methods, however beyond this there is the issue of the mathematics being used to establish this 'superiority' relies on conditions that the study design and collected data (being pseudo-experimental) does not meet.

Even if this were the case, the constraints around the laboratory design – like-for-like emulation rather than taking into account different modes' different affordances, and holding the laboratories constant in the face of observed activities, mean that if they were mathematically valid, they would be giving insight into laboratory designs that would never exist outside the study constraints that generated them.

Consider some relevant and fundamental inherent affordance arguments in favour of remote and simulated laboratories (note these conform to the conflated archetype definitions in use in this paper):

- Remote laboratories
  - Can be accessed anywhere an internet connection exists, not just the classroom.
  - Can be accessed outside of class times without requiring set-up, tear-down, human operator safety concerns, etc.
  - Can display computer-mediated data in new pedagogically inspired formats.
- Simulated laboratories

- Can do everything a remote laboratory can do, with the added bonus of not being constrained to reality when conducting experiments.

These affordances were artificially eliminated when conducting this study – thus reframing the question from a comparison of "remotely-operated, simulated, and hands-on student laboratories" to a comparison of the hands-on design contorted into remotely-operated and simulated forms.

Further to this, and circled around by the previous 'familiarity' excerpt, is the issue of design maturity. While the remote laboratory has been reported on in previous publications (the model [37] paper contains a pilot study for this paper), this was still a very immature laboratory design being contrasted with a mature traditional laboratory. Ignoring familiarity by the students, the knowledge of how the laboratory should best be designed and run by the instructors lies firmly in the traditional mode. This is most exemplified by the increased set-up time for the simulated laboratory (which would be unlikely to occur in a modern simulation, CD-ROM installation having been supplanted by far more convenient methods).

Critically, this study already identifies several of the points raised. In particular the single lab format issue and simulation installation time were indeed recognised as limitations, as well as the contorted affordances. A key statement here is:

"But if a study were designed to take advantage of all the natural affordances of each medium, it would be difficult to ascribe causality to the medium itself, as opposed to differences in how each medium could be applied or how it tends to be used by students" [175]

It is the position of this thesis that the affordances of a medium ARE the medium itself, and trying to separate the medium from what that medium allows is nonsensical. If one compares a horse to a car by requiring the car to be fuelled with precisely the same hay as the horse, one cannot be said to be validly comparing the two regardless of the adherence to protocols.

Nonetheless, several of these limitations are identified and stated (though not all, particularly the pseudo-experimental nature and lack of mathematical validity). As appropriate as this is, representative of best practices in research, this recognition is insufficient to bring the best possible value to a reader, because there is insufficient data collection and reporting for a reader to appropriately incorporate the findings into some alternative context. The interesting lines of inquiry mentioned earlier are representative of this issue – there is no rich context, no follow-up data collection, no tweaks to investigate causation or to dive into what processes are occurring. In the pursuit of validity, which was mathematically impossible under the study design conditions, this valuable knowledge has been left on the table, so to speak.

3.1.2 Vignette 2: Literature Aggregation

J. R. Brinson, "Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research," *Comput. Educ.*, vol. 87, pp. 218–237, Sep. 2015.

One lens by which to view research on laboratory evaluation involves directionality - research presented in order to support future evaluations, and research which collects knowledge from existing evaluations to derive some aggregate or abstract knowledge. While all research will have aspects of both, in the context of this thesis the previous vignette sits primarily in the 'before' stage, and this vignette looks at a paper at the 'after' stage. This correlates strongly with the aspects of laboratories which these papers are interested in - Corter et al [175] (and the Nickerson et al [37] paper which it) looks most strongly at the 'input' side of laboratories, and the work presented is aimed at supporting future studies (with the proposed framework). The partner paper which is seminal to the work of this thesis is Brinson [21] which looks most strongly at the *outputs* of evaluative and particularly comparative studies of laboratories, and while it provides insights for future studies, focuses most heavily on collating the findings of many existing studies.

Brinson [21] is a synthesis paper looking at:

> "empirical studies that focus on directly comparing learning outcome achievement using traditional lab (TL; hands-on) and non-traditional lab (NTL; virtual and remote) participants as experimental groups" [21].

Leaving aside the conflation of categories that per Nickerson et al [37] are not the atomic categories they are described as [46], this comparison serves as an excellent 'surveying of the land' regarding the types of studies being performed, but is hampered greatly by the nature of the available data. This paper exemplifies one of the core motivations for this thesis - current research practices, particularly around publication and what is reported, does not support the meta-analyses that are necessary to gain knowledge between contexts. This manifests itself in Brinson [21], weakening the methods available to be used, and in large part preventing such meta-analyses from being valid from the outset.

Brinson looks at the learning outcomes measured in contributing studies, categorised into a KIPPAS framework: Knowledge & understanding, Inquiry skills, Practical skills, student Perception, instructor Perception, Analytical skills, and Social & scientific communication. This is done by coding contributing studies as measuring these learning outcomes, or not. It also identifies the channels used to collect data - Lab Reports/Written Assignments, Quiz/Exam, Survey/Questionnaire, Interview/Observation, Lab Practical, or generic Grade. Finally, it codes each study as finding 'higher outcome achievement' for the traditional laboratory, the non-traditional laboratory, equal achievement or no significant difference, and differing achievements between the 2 types.

The most obvious and critical issue with the synthesis is the superficial aggregation between studies of highly varied contexts. To reduce it to its most absurd conclusion, if a particular type of non-traditional laboratory was greatly beneficial in teaching a chemical engineering phenomenon in some context, but woeful in teaching some electrical engineering phenomenon, or chemical engineering in a different context, this would appear as '50%' in the analysis, but would entirely lack the nuance of under what contexts the lab design choice was beneficial or not. The idea that the discussion of non-traditional laboratories compared with traditional laboratories can arrive at a single yes or no answer across multiple contexts stretches credibility.

As problematic as the aggregation of comparative findings is, the paper is nonetheless incredibly useful for its aggregations of what actually occurs in research on laboratories. Aside from the simple literature collation, identifying the methods in use - and in particular the reliance on existing assessments as evaluative tools, supplemented by affective surveying, highlights the limitations in the practice of evaluating laboratories.

Brinson [176] is a follow-up to the original review, which does not add new papers to the review, but digs further into the nature of the original body of work, identifying further aspects. These aspects draw very close to identifying the issues presented in this thesis, but come to an alternative conclusion. In particular the variability of context is highlighted, though only at the macro scale: k-12 vs tertiary education, engineering vs natural sciences, US vs international. The recommendations are primarily to increase the depth and breadth of quantitative experimental design studies on laboratories. This alternative conclusion was the focus of investigation in the earlier described Truss study [39].

Were the contributing studies to have reported more to do with their contexts, and in particular to provide more details in their findings than resulting numbers that cannot be questioned deeper after the fact, meta-analyses such as these would be far more empowered to come to actionable conclusions for specific contexts. Questions such as "under these circumstances, which we observed by these methods, yielded these outcomes", can only be useful to a practitioner if there is sufficient disclosure about each of these 3 main components for the practitioner to identify or convert them between contexts.

3.1.3 Vignette 3: Value in Quantitative

D. Basu and V. K. Lohani, "Learning and engagement with an online laboratory for environmental monitoring education," Eur. J. Eng. Educ., 2023.

The previous paper vignettes establish the nature of evaluation of laboratories in engineering education – either being a review in the case of Brinson [21], [176], or being derived from one in the case of Corter et al (Ma [20]). Laboratories are largely examined through quantitative experimental designs, with significant issues around complex and confounding contexts, bringing about the lack of internal validity and generalisability discussed previously. Basu & Lohani [177] is a more recent publication (published in February 2023), and is included to establish that these practices and study designs are ongoing in the field, and is present in top publication venues such as the European Journal of Engineering Education (EJEE).

Basu & Lohani [177] does not include the description of laboratory implementations normally seen in papers of this nature, due to this being established in prior publications on this same implementation, i.e. [178]. This has left significant scope in the paper to spend time establishing the basis for the study, and going into depth regarding the design and methodology.

The study is particularly well-founded in the literature, bringing in perspectives from more general education and cognitive psychology to identify the areas for investigation, as well as to provide insights and anticipated results as predicted by this literature prior to conducting the study. The theoretical framework used was the situated perspective of learning, as described in e.g. [179]. It also incorporates theory on student engagement, drawing from the fields of broader education research as well as Human-Computer Interaction (HCI). An important side observation is that this foundational theory was not derived from quantitative experimental methods, following the practices more commonly followed in the disciplines in which they were established, such as ethnography and case studies.

The study also correctly establishes the nature of the design it conducts – identifying it as a pseudo-experimental design as per [48]. This design lacks a control group, and does not have true random sampling – while there was a randomness element in the study in Basu & Lohani [177], it was taken from a non-random student pool, being the class the study was conducted within. As discussed previously, this design is almost certainly the design used in any true-to-life quantitative experimental design for laboratory evaluation, though it is uncommon in the extreme for publications to recognise this limitation and discuss it. The most relevant characteristics of a pseudo-experimental design are a lack of causality (only correlations may be identified), and a lack of generalisability.

Student engagement with the system was collected from various computer-mediated sources, and situated in conjunction with a pre-survey on demographics and an HCI-derived post-survey on perceptions. While some qualitative written survey responses were collected, these were coded into quantitative data for analysis rather than being analysed using any established qualitative methodology. The analysis employs several quantitative statistical approaches, establishing relationships through statistical analysis and arriving at variations of means with p-values. These variations are highlighted as the findings of the study, with the p-values being used to support findings as statistically significant. The findings are split into discussions on learning outcome achievement, engagement with the system, and the relationship between these.

The learning outcome achievement discussion states

> "a detailed analysis of the content questions indicated a significant improvement in students' performances from the before to after the OWLS-based intervention. Although general water quality concepts were covered in the course, this improvement in students' performance could only be attributed to the OWLS-based task as the content questions were tied to the five LOs." [177]

This correctly identifies that students scored better on the aspects the laboratory taught, but does not examine the quality of the laboratory itself – whether it was an improvement vs teaching via some simpler alternative.

The discussion on engagement centres around

> "As there was no control group, we cannot provide inference about the effectiveness of the OWLS in terms of engagement" [177],

This is a correct observation given the pseudo-experimental design, and one which speaks to the need for alternative methodologies for evaluating laboratories. Some correlation discussion is present, but in essence the related research questions cannot be answered. Further, there is no established generalisability for the correlations identified.

The final discussion point looking at the relationship between learning outcome achievement and engagement is thus hampered by the limitations in the contributing halves – given the nature of the study any conclusions can only be correlative rather than causative, and cannot establish external validity. Notably however, these findings do speak to the specific nature of the laboratory being investigated – the HCI surveys identified aesthetic requirements the system was not meeting, and highlighted areas of interest that correlate to larger effect sizes in outcomes. This speaks to a core idea in this thesis, that while quantitative *methodologies* cannot be effectively used to answer the sorts of research questions the literature is currently applying them to, quantitative *methods* have a valuable place in a broader approach.

3.1.4 Vignette 4: Design Research

S. Verstege, J.-P. Vincken, and J. Diederen, "Blueprint to design virtual experiment environments," Comput. Educ. Open, vol. 2, p. 100039, Dec. 2021.

This paper presents a challenge in constructing the narrative of this thesis – it is part of the background informing the thesis overall, but also skips ahead to the core of the selected methodology (particularly having been published midway through the data collection phase in Dec 2021). Within the context of the thesis, it may be characterised as recognising the same issues with current practice in research on laboratories in engineering education, identifying the same broad solution, but not applying it fully, nor for the same reasons as this thesis.

The paper describes the application of educational design research (EDR) to the design of simulated laboratories (termed Virtual Experiment Environments or VEEs). As discussed, this approach to research involves the iterative investigation of an artifact, generating 2 categories of findings – a 'local instruction theory', which captures the iterative improvements to the artifact generated by the research, and design guidelines, which represent the generalisation of the local knowledge which may be interpreted by others, and incorporated into other contexts where it is deemed appropriate.

One of the speculations made by the author of this thesis when investigating why educational design research was not commonly employed to resolve the issues with laboratory research in engineering education was "educational design research is difficult to apply well, and requires understanding that is outside the current practice of engineering education research". This paper appears to validate this speculation, as the implementation of educational design research being used is incomplete, and in particular includes structural issues – whether misinterpretation, misreporting, or misalignment – that limit the effectiveness of the study.

It should be noted however that the nature of educational design research is resilient enough, and so closely aligned with the practice of engineering as a discipline (in contrast to the practice of engineering education research), that even with these issues being present, the value of the paper in informing the design of laboratories is significant. Unlike a study which compiles down to one or a set of quantitative values, limitations, or errors in which can completely negate their validity in other contexts, the rich description inherent in EDR means that these limitations can be navigated around by an informed reader, and the partial knowledge interpreted in meaningful ways.

The study presented by this paper examines design requirements and contributing design principles for the development of VEEs, as generated alongside the design of 4 VEEs themselves. These VEEs are then assessed against the design requirements and principles to evaluate them, which in turn is used as evidence for both the validity of the design requirements and principles, and the quality of the VEEs so developed. This process is structured with an EDR approach including analysis and exploration, design and construction, and summative evaluation.

The first issue in the study is the unclear identification of the artifact under study, and the local instruction theory. The development origin of EDR as it is appreciated in this thesis is via Van Den Akker [109], and an approximate model is used by Verstege et al [180] via Bakker [101] who heavily leverages Van Den Akker and others in the Netherlands tradition of EDR in their work. This grew out of mathematics education, and is able to very precisely articulate what the local instruction theory is, and what the artifact was. By a purist interpretation, the artifact in this paper would be the VEE or VEEs, and the local instruction theory would be the cognitive processes students would undergo during these teaching and learning activities.

In this paper however, the local instruction theory, while not explicitly termed as such, is instead the design requirements and principles. The artifact however is reported in a way that is conflated – being simultaneously the VEEs under development, but critically also the design requirements and principles. This conflation manifests as a perceived error in the study design for the paper, and has led to the iterations described in the study (critical for the effective application of EDR) being misaligned.

There are 2 possible well-formed study designs that could be drawn from this paper. In one, being the purist approach described earlier, the artifact would be the VEEs, and the local instruction theory would be the students' cognitive processes that occur within them. Under this approach, the iterations would need to be changes to the VEEs, with the VEEs being evaluated and those evaluations being fed back into the design.

The other approach, which appears more closely aligned with the stated aims of the paper, would be to have the VEEs be the artifacts under study, and the iteration be progressive elaboration of the design requirements and principles. It is possible this is in fact what was done, but the manner in which it was reported did not identify the cycles of iteration that took place, the findings at each step in these iterations which motivated changes, and thus the evidence for validity of the design requirements and principles themselves. Instead, what was reported was the post hoc application of the design requirements and principles to the same VEEs which generated them, to evaluate whether the VEEs met these.

The net effect of this on the paper, irrespective of whether the error was one of study design or study reporting, is that the local instruction theory (being the design requirements and principles), is unevidenced. There is significant evidence presented in the subsequent portion of the paper investigating whether the VEEs developed met the requirements and principles, but no empirical evidence that the requirements and principles are justified is reported, besides the description that they were iteratively developed.

This misalignment of evidence also manifests in the evaluation of the VEEs themselves – seemingly falling out of the EDR approach, the VEEs are evaluated in the post hoc application mentioned previously in the same style of pre- or pseudo-experimental, quantitative methodology described in previous papers. This goes directly against the essence of EDR, which mandates an iterative approach which recognises the inherent challenges in experimental approaches to educational design and evaluation.

The limitations section of the paper highlights the impact of this 'big bang' style after-the-fact evaluation. The inability to quantify certain aspects of interest, or in cases the ability to measure them at all – which may have more appropriately been assessed by local, qualitative, small-scale evaluation which was then fed back into an iterative process – means that a large portion of the final conclusions were not properly supported by empirical data. Rather than a discussion of the validity of the local instruction theory, being the design requirements and principles, and the likelihood of them being applicable in other contexts, this is instead left to final conclusions rather than treated with any significant depth. The essence of this impact is captured within the final discussion paragraph:

> "In the evaluation we have shown that the VEEs were successful, as they satisfied the design requirements. However, we have not yet proven that our general output classes (i.e., design requirements, design principles, design architecture) will function in other levels and other fields of education." [180]

This flipping of subject-object (VEE and design requirements/principles/architecture) is the conflation that is the core misapplication of EDR in this paper.

While these issues are significant, and are held in support of the efforts in this thesis to create a more approachable way to embed EDR in laboratory design and evaluation, the nature of EDR is such that value may still be drawn from this paper. Unlike a quantitative study, the bulk of the value in EDR research lies in the process and discussion that occurs in the various stages prior to any final summative evaluation. In particular, this paper gives the source external motivation for the design requirements and principles – sadly lacking the local, empirical motivations for refining them as should be disclosed in EDR studies, but nevertheless providing a high-quality base of relevant research backing the inclusion of each in future work. Identifying which of those design requirements and principles *can* be successfully investigated through more straightforward quantitative surveying and which cannot is similarly a valuable contribution that may be incorporated into future work.

3.1.5 Vignette Conclusions

There are a multitude of papers that arguably should be given a similar deep dive treatment in this thesis – and indeed a significant portion of the exploratory design phase for the artifacts this thesis developed was doing exactly that. For reasons of appropriate length and scope however they are omitted in favour of these most critical representatives. The core findings that these yield are articulated here as inputs to the design process for the artifacts under study:

- Quantitative research in educational evaluation, particularly for laboratories in engineering education, are largely if not inherently restricted to pre- or pseudo-experimental designs.
- Pre- or pseudo-experimental designs are unable to appropriately answer the questions posed by evaluative practice and research on laboratories in engineering education.
- Overcoming limitations in study designs through aggregation and meta-analysis is not possible with the reporting generally made available in current research practices on these laboratories.
- Quantitative methods have a place in answering evaluative questions on these laboratories, but must be situated within a more effective and comprehensive overall approach.
- Educational design research has the potential to be this overall approach, but risks exist in its effective application, particularly as it is not yet well-established in the engineering education research discipline.
- Post hoc data collection alone is insufficient when attempting to conduct EDR, or when answering the overall questions research typically poses when studying these laboratories.

3.2 Research Contortion Revisited

The observation of research contortion as it applies to laboratories has some key consequences for research design, both in that context of laboratory design and evaluation, but which also apply to any comparative evaluations in engineering (and likely wider) education research and practice.

3.2.1 Definition

Research contortion is here defined as:
The manipulation of objects under study in comparative evaluation, which has the effect of modifying them to align with requirements of a comparative evaluation methodology, which do not apply to the intended applied context or use case of those objects.

Objects under study could be laboratories as observed in this critical review, but could conceivably be any other artifact, design, element, etc. in education research which is sought to be compared with some alternative.

Vignette 1: Comparative Study above provides an excellent illustration of one form of research contortion in comparative research, related to affordances. The laboratories being compared as part of that study, to which the conclusions were then applied to the different modalities of these, were forced into 'parity' with one another, to enable the comparative methodology being used. Achieving this parity required the blocking of the affordances of the remote and simulated versions of the laboratory would have been able to provide were they to be used outside of this research context –

affordances which usually appear in the preamble of studies like this one – and even *including* this one. To wit:

> "For students, new technologies can be convenient, eliminating the need to physically attend a class." [175]

In the introduction for Vignette 1, the study design of which required physical attendance in class to use laboratories involving those new technologies.

It is important to call out here that the authors of this paper explicitly recognise this limitation with their study design, identifying the limitation to affordances which limit the generalisability of their findings. Vignette 1 was used in the critical review in part due to it being identified as the highest quality comparative experimental study on laboratories in engineering education in terms of reporting rigour. This paper is a rare example of this limitation being reported and engaged with, however, and it does not receive consideration in the aggregated literature review papers which informed the earlier literature review, e.g. [1], [15].

3.2.2 Illustrations

This affordances contortion can be further illustrated with a thought experiment considering different comparative study designs in a similar form to vignette 1. Consider a study comparing laboratory versions. Version A is a hands-on archetype laboratory where students add acids to beakers containing different metal objects and observe corrosion effects over an hour. Version B is the same experiment, but conducted using a robotically controlled dispensing and camera observation setup controlled over the internet. Version C is a simulation in which the predicted appearance of the metal objects is displayed if given the acid and duration as an input.

The comparative study is attempting to follow a randomised controlled trial (limits to the ability to maintain a control group notwithstanding). As the independent variable has been identified as 'lab modality', and the dependent variable is 'student learning outcomes', all other variables must be controlled lest they invalidate the findings. Thus, the location and time students undertake the laboratory, teacher and peer presence in the room, and all other potential nuisance or confounding variables must be controlled to achieve parity between versions, so that the requirements inherent to the statistical methods of quantitative research are upheld. So students complete all laboratories in a laboratory room, during scheduled class time, together.

It should be clear that this is not a true comparison of versions A, B, and C. The requirements inherent to A have been artificially applied to B and C, when there would otherwise be no reason inherent to those designs which required them. As will be described later throughout this thesis, laboratories can be described as the sum total of their characteristics – and thus this study is **not** comparing versions A, B, and C – it is comparing version A with 'B contorted to A's requirements' and 'C contorted to A's requirements', and there is no evidence at all from this study to 'translate' these findings to the comparison between the original versions.

As a side note, it may well be that 'B contorted to A's requirements' is a better version of the laboratory than 'B' – research contortion does not claim that the objects under study are being made worse. Rather, it is the observation that the objects under study are being made different, and this difference prevents conclusions from applying to the original, non-contorted version.

Affordances are not the only way that laboratories and other objects under study may be contorted. Seeking a control group, or random assignment, or other requirements inherent to chosen methodologies also contort these objects under study into different objects under study. Another hypothetical comparison study illustrates this.

Two versions of a laboratory are being compared as part of an RCT – one in which students work individually, and one in which students work in groups of 3. Everything else about the laboratories is the same, and no interface or equipment affordances are being suppressed. However, to ensure random assignment of students, the laboratories are run by randomly splitting each class in 2, and having these 2 halves given different instructions (individual or group). The first 10 minutes of the class involves assigning these groups, and ensuring the students move to the individual or group halves of the room. To ensure everything is controlled, the same lab demonstrator gives instructions and guidance to both halves, and their time is divided equally between each side. The group version in this hypothetical study does much better than the individual version – but this is actually due to the demonstrator only having to correct errors in experimental setup once per group in the group version, while they need to correct errors individually in the individual version, and this time disparity causes poorer learning outcomes in the individual cohort.

Adding research controls contorts the laboratories under study into identifiably different laboratories, and the research conclusions can be applied only to those contorted versions, not the real context versions which might be implemented due to the comparison results. Another potential research control contortion might involve changing timetabling of classes to ensure random assignment in an AB-BA study design – such as having some students completing laboratory A in mode X and laboratory B in mode Y and some the opposite, where this meant students were assigned to specific class times to match their mode assignment rather than allowing them to select preferred class slots and times as would normally be the case. Many other changes may be motivated to control different confounds and suit the requirements of true or quasi-experimental designs.

3.2.3 Claim to novelty and similar work

Research Contortion is not an entirely unique idea in the wider field of research design. In HCI the terms 'in the wild' and 'laboratory studies' [181] target the confound and nuisance controls elements in particular, by identifying this control as a limitation of laboratory studies. Moving to 'in the wild' studies and giving up the 'clean white room' laboratory-style investigation avoids this contortion, at the cost of bringing in uncontrollable contextual influences. The novelty of Research Contortion is in its application to a context that has yet to consider this as a concern.

There is another similar concept in research design, the Hawthorne Effect, or Observer Effect [182]. This effect can be summarised as the observation of research participants changes their behaviour and thus their observed outcomes. The Hawthorne Effect is embedded in the Research Contortion concept in that it is one of the (but not the only) ways in which an object under study might be contorted.

Research Contortion is to a best effort of term and synonym search not a concept which has been considered in evaluation research in engineering education, and it is the application to this scope and comparative research in this discipline in particular that novelty of this concept is claimed. No use of this term or any related concepts have been identified in any other educational field, though a claim to novelty in this broader concept is not as well supported by the earlier literature review which focused heavily on engineering and science education literature as a foundation.

3.2.4 Impact

Identifying limitations to research does not require the broader recognition of research contortion as a concept, as seen in Vignette 1 which does just this. The impact of having this term defined is to specifically call out these limitations in research design – both to promote this recognition which is absent in much of the literature even if it is able to be independently identified, and to promote the investigation and adoption of alternative, contortion-avoiding or -mitigating approaches to research, which is a major theme to this thesis work overall.

Research Contortion is a significant impediment to effective evaluation research, particularly comparative research. The establishment of this concept as a concern allows for a better-targeted effort to address these limitations to improve research quality and the internal and external validity of research in engineering education. This remains an underlying design consideration throughout the subsequent work described in this thesis.

4 METHODOLOGY

This thesis undertook to explore the broad scope of the question "How may laboratories in engineering education be effectively designed to meet the purposes their stakeholders require, and evaluated against these purposes?". This exploration generated a candidate artifact intended as a potential answer to this question – the Purpose-First Approach and accompanying PIPOV framework – and thus settled on the specific research sub-questions of

1. How do the research methodologies used in evaluative research on laboratories in engineering education support or inhibit answering the questions this research poses?
2. What methodological alternative provides the appropriate characteristics to resolve the identified issues with experimental approaches in the design and evaluation of laboratories in engineering education?
3. Can Educational Design Research be successfully applied to the design and evaluation of laboratories in engineering education?
4. How can practitioners be supported in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?

4.1. Do the Purpose-First Approach and PIPOV Framework support practitioners in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?

Answering these research questions were jointly inspired by and required the use of a particular methodology – which indeed was chosen as part of the process of articulating the specific research questions, given the inherent links between methodologies and the types of questions they are able to answer. While there were several potential methodologies considered for use in this thesis where the project would have looked markedly different, the overall type of methodology that was best suited was quickly identified as being the practical inquiry family of methodologies described previously.

4.1 PRACTICAL INQUIRY & EDUCATIONAL DESIGN RESEARCH

The nature of laboratory design and evaluation inherently requires a design process. Not all design must be research, but in order to be effective design should include external inputs to support conjecture, validated by testing in practice. The fundamental conclusion from reviewing existing practice in research is that these research outputs were ineffective at supporting this ideation based on external inputs stage of design. It is possible that purely qualitative methods might resolve these issues, but the review also observed:

- Practitioners have background issues that inhibit the adoption of these methods.
- In particular issues of generalisability and transferability are a challenge.

Qualitative methods tend to require significant understanding of backing theoretical frameworks and following a rigorous process where the derivations of that rigour are likely outside the normal training of an engineering practitioner. A failure to properly account for this is likely to lead to equally flawed and ineffective research. The potential value of quantitative methods applied not as controlled, attempts at true experimental designs, but as inputs to a triangulation process involving multiple sources of information was also established.

Further to this, the processes employed by engineering educators already map to iterative, sequential design, reflection, and refinement – design being a distinguishing characteristic of engineering [183]. Aligning with these practices is thus a potential avenue for change from existing practices while staying within familiar territory. As such, the most appropriate methodology is likely to be found within these traditions.

The practical inquiry family of methodologies includes several candidates that may have been applied to this thesis:

- Action Research
- Design Experiments
- Educational Design Research

Of these, Educational Design Research emerged as the most appropriate in order to properly tackle the wide variance in contexts between laboratories and those developing and evaluating them, and to ensure authenticity – for both points seeking maximum generalisability. A project undertaking Action Research would have necessarily focused more narrowly on a given laboratory to achieve specific goals. This would have more rigorously sought improved outcomes for a specific case, but would not have as fully explored the generalisability of the interventions developed. Depending on the specific traditions and contextual theories applied, a project undertaking Design Experiments would have focused more on the earlier stages of candidate development, exploring multiple untested designs rather than a single design through multiple contexts and multiple iterative revisions.

Educational Design Research (EDR) as the previously identified tradition within the design-based research etc. family of methodologies is more commonly applied to specific teaching and learning activities and their wider curricula. Indeed, EDR forms a core component not only of the overall approach to answering this thesis' research questions, but is also a feature of the generated artifact itself. EDR has some key features that support the intentions of this research:

- The development of an artifact, which is intended to deliver particular learning outcomes.
- Iterative design, evaluation, and redesign phases, which take in observations of the artifact in authentic contexts, and feed this back to resolve observed issues and improve the artifact to better achieve the learning outcomes.
- The development of a "local instruction theory," that posits how the activity occurs in practice, and how this leads to students achieving the intended learning outcomes.
- The articulation of this local instruction theory for the purpose of generalisability, through the disclosure of context.

This methodology is sufficient in itself to be applied to laboratories in engineering education directly, with significant confidence based on its success in other educational disciplines that it would be effective. However, the observations of current research practice when evaluating laboratories indicate that there are barriers to its adoption that must be addressed in order for it to achieve significant penetration and thus have the desired impact.

- Many if not most practitioners of laboratory evaluations in engineering education come from engineering discipline backgrounds, rather than engineering education.
- There are significant scope, time, and design limitations on the evaluation of laboratories in engineering education.
- Performing educational design research well requires a good understanding of the methodology and its constituent methods.

- There is no 'minimum viable research approach' currently explored in the literature such as has been observed for e.g. Bloom's Taxonomy or Constructive Alignment.

As a result, some artifact was considered likely to be necessary in order to ameliorate or at least mitigate the impacts of these barriers, to promote the adoption of EDR for the purposes of laboratory design and evaluation. As stated in [109]:

> "Design research is not done to test theories, even though its results can sometimes suggest weaknesses in theory. Rather, design research discovers ways to build systems based on theories and determine the effectiveness of these systems in practice." [109, p. 11] – where system is interpreted to include artifacts being designed.

Educational Design Research is well established in the literature, particularly in mathematics education – however it is primarily intended for the specific application to teaching and learning activities. The significant parallels between the evaluation of a teaching and learning activity intended for students, and what is essentially a teaching and learning activity for practitioners, means that this is likely to be transferrable in this altered context. Thus, the original EDR methodology underwent a slight conversion to apply to the development of the Purpose-First Approach and accompanying PIPOV framework.

Under this conversion, rather than intended learning outcomes as would be the case in a teaching and learning activity, the desired outcomes are instead the practical outputs of the research questions posed – namely the effective design and evaluation of laboratories by practitioners. It is important to reiterate here that EDR is a synergistic research methodology – it combines both practical outcomes and scientific, or rather, it obtains the latter by way of targeting the former. The artifact under development, normally a teaching and learning activity and supporting material, is instead the Purpose-First Approach itself, with the accompanying PIPOV framework.

4.2 EDR And Meta-EDR in This Thesis

The concurrent use of EDR at 2 levels of abstraction in this thesis has the potential to cause significant confusion. A converted form of EDR is the primary methodology by which this thesis' research was conducted. EDR is also the recommended methodology for evaluating laboratories in engineering education, and is embedded in and espoused by the Purpose-First and accompanying PIPOV framework artifact this thesis has developed. The following outlines the application in these two levels to make this distinction clearer, with the details of each to follow as the design is explored:

Table 2: Meta and Direct application of Educational Design Research in this thesis

Meta	Direct
The research methodology for this thesis	The research methodology embedded in the Purpose-First Approach
Applied by the researcher with practitioners as collaborative research subjects.	Applied by practitioners with students as research subjects.
Artifact: The Purpose-First Approach and accompanying PIPOV framework	Artifact: The laboratory being designed/under evaluation

Local instruction theory -> Local evaluation theory: how the Purpose-First Approach and accompanying PIPOV framework are used by practitioners, and what types of insights it can yield	Local instruction theory: what students learn in the laboratory and how they learn it.
Generalisation: applying the local evaluation theory in new contexts, designing and evaluating laboratories using the generated artifacts, using the reported insights and surrounding context to determine the appropriateness in the new contexts	Generalisation: applying the local instruction theory in new contexts, designing laboratories to match previously designed laboratories if and to the extent to which the reported context indicates a likelihood of replication. Developing a domain-specific instruction theory for laboratories in engineering education.

4.3 EDR Design Goals

As previously mentioned, the design was intended to resolve identified barriers to the simple application of EDR by practitioners seeking to design or evaluate their laboratories. These four identified issues are co-mingled, and can be expressed in a more unified way as:

> Designing and evaluating laboratories using EDR is highly likely to be effective, however it requires skills outside the experience and background of many practitioners. Evaluation is often done under considerable time pressures, with this often leading it to be 'tacked on' after the fact, which is incompatible with EDR. The lack of an easier approach to EDR also represents a hurdle to overcome.

The success of Educational Design Research may be characterised as its ability to generate practical, functional artifacts that have been iterated upon to slowly move in the aggregate towards the best practicable version, coupled with its ability to communicate the findings from the generation of such artifacts in a way that can be readily consumed and incorporated by practitioners in other contexts, with both increased chances of comparable benefits being obtained, and the ability to observe in place and intervene should the contexts not align as expected.

These successes hinge on several characteristics of what EDR involves and produces:

- EDR acknowledges the inherent uncertainties in educational contexts, and follows the paradigm of satisficing – obtaining reliable data that shows trends within a specific context, and using those observed trends to 'aim' the design at the correct target, moving the outcomes closer to that target over time/iterations.
- EDR acknowledges that holding a teaching and learning activity design firm in the face of new data is entirely inauthentic, and would invalidate the results (if an educator notices something going poorly in a class, they will almost certainly react to this). Teaching and learning activities can and should change in the moment to incorporate new data as and when it is observed.
- EDR acknowledges that no two moments, classes, educators, institutions, etc., in education can ever be exactly alike, and that in order to generalise findings from one single context to another, expert qualitative interpretation must occur:
    - Corollary to the above, EDR recognises that this qualitative interpretation must be supported by the reporting of research findings if it is to be given the highest chance of success.

- Further to this, and in combination with the reactions to observed data, EDR requires an articulation of the *internals* of a teaching and learning activity, not merely a design and a set of outcomes with the activity itself a black box. This allows a practitioner to incorporate findings from another context, and observe the alignment with their own context, without needing to wait until final outcomes to determine the accuracy of the application in their own context.

Each of these characteristics requires the practitioner of EDR to have a good understanding of the nature and practices of EDR, which as observed above represents a barrier to adoption of this methodology in the design and evaluation of laboratories in engineering education. An artifact developed in order to promote the adoption of EDR for the design and evaluation of laboratories in engineering education therefore needs to retain and support those characteristics, without requiring the in-depth knowledge of EDR in order to achieve this.

These therefore are the 'goals' which are the target of the meta EDR applied as the research methodology for this thesis. The artifact shall target them as desirable outcomes, and insights into how the artifact behaves in real contexts (with generalisability to other contexts) are thus the scientific counterparts to these practical outcomes.

In order to validly incorporate the benefits of EDR into laboratory design and evaluation:

- The artifact should lend itself to iterative development, incorporating a general target(s), observations to be made as to its immediate effectiveness in light of those targets, and iterative migration towards them.
- The artifact should acknowledge and support mid-run reactions to observations rather than locking in a quantitative experimental design.
- The artifact should support the qualitative interpretation by readers of design and evaluation research using it, through rich reporting of context and process.

In order to ameliorate or mitigate the identified barriers to EDR adoption:

- The artifact should be approachable and usable by those from an engineering education discipline rather than specifically engineering education research.
- The artifact should support use where time and scope available may be limited, including where these limits arise out of a lack of experience or exposure to specific methods on the part of the practitioner.

This second set of requirements is necessary for the realistic expectation of real impacts of the artifact, a necessary component of EDR as opposed to more theoretical research.

The last point requires deeper explanation, as it perhaps goes against more traditional views that quality research must derive from a proper understanding of the methodology and methods employed, must be well designed, and well executed. This is of course true in the sense that the quality of research can only improve if these are followed, however it is not realistic nor of any real benefit to answer the problem of designing and evaluating laboratories with "the available time for overloaded practitioners must increase, and practitioners whose careers are anchored (and continue) in quantitative engineering disciplines must upskill into sophisticated engineering education research techniques".

4.4 Peripheral Engagement

It is relatively commonplace for practitioners conducting research to leverage methodologies and frameworks of which they either a) do not have the time or did not make the necessary prior preparations to conduct effectively, or b) do not fully understand [184]. This is crippling when it comes to quantitative experimental research designs, as the mathematical conclusions reached rely on assumptions about the data and how it was collected, and cannot be relied upon if these requirements are not met.

One of the perhaps most important characteristics of both Bloom's Taxonomy and Constructive Alignment, however, is that a small scale, minimally informed attempt at integrating these artifacts into design or research is may still yield some benefit. For Bloom's, this might involve applying the cognitive domain without understanding of the affective and psychomotor domains, if not a further subset of the cognitive domain itself. For Constructive Alignment, the understanding may often be of a single diagram, and an understanding that teaching and learning activities, and assessment, should be aligned with learning outcomes.

This benefit from small forays into these frameworks requires what this thesis terms peripheral engagement with the "minimum viable methodology" – the smallest amount of knowledge about a framework or practice that still yield benefits. The minimum viable methodology of a given framework or similar is the extent to which that framework must be understood in order to provide benefits, and most critically to avoid damaging outcomes.

In this area qualitative approaches have an inherent advantage over quantitative, by virtue of their outcomes not being considered absolute. Consider a quantitative experiment that was designed or executed poorly due to poor understanding of the techniques required, or due to limited time available. As observed in the earlier literature review, conclusions based on statistical significance that contain errors cannot be relied upon, and because the design treats the laboratory as a black box, no further quantitative inferences can be made.

Compare the least expert qualitative design – something educators do regularly when they use no experimental design at all, and simply observe the effects of an intervention anecdotally. The rigour involved and the overall validity of such research is negligible, however over a long time period the overall effect is to 'drift' towards higher quality. This can be observed as the difference in quality between an educator with many years of experience and one with few.

The exact nature of the minimum viable methodology for a new framework or similar is inherently unknowable at the outset – the question of what is the minimum required for valid research can be speculated, but what is far less clear is what pieces of the methodology will be adopted most readily, and which ignored. As such, acknowledging and supporting the existence of this peripheral engagement, and designing such that the minimum viable methodology aligns with the concepts and components most readily adopted by practitioners must be a design goal for the artifact this thesis develops.

5 PHASES

With the selection of Educational Design Research as the methodology for this thesis, and the design goals for the artifact to be generated determined, this thesis work then undertook the iterative process for creating and refining the local instruction theory (local evaluation theory in this case), and the artifact for delivering it. EDR follows a multi-phase iterative process as in Figure 8, where each step in the process combines progress in both the practical artifact, as well as the theoretical understanding of the local instruction theory.

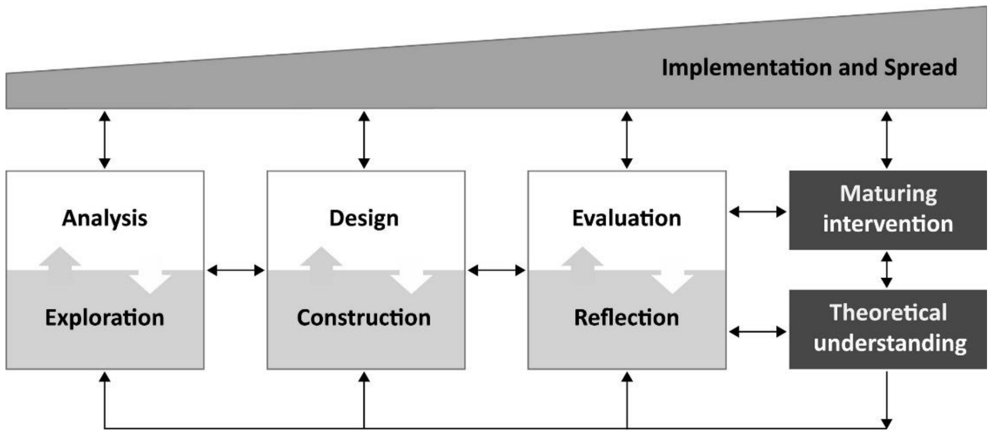


Figure 8: Characterised phases in Educational Design Research [134]

This thesis broadly follows these three phases, with the progressive elaboration and refinement of the Purpose-First Approach and accompanying PIPOV framework being the practical artifact, and the local evaluation theory being the observations of how these ideas and frameworks manifest in practice, as in Figure 9.

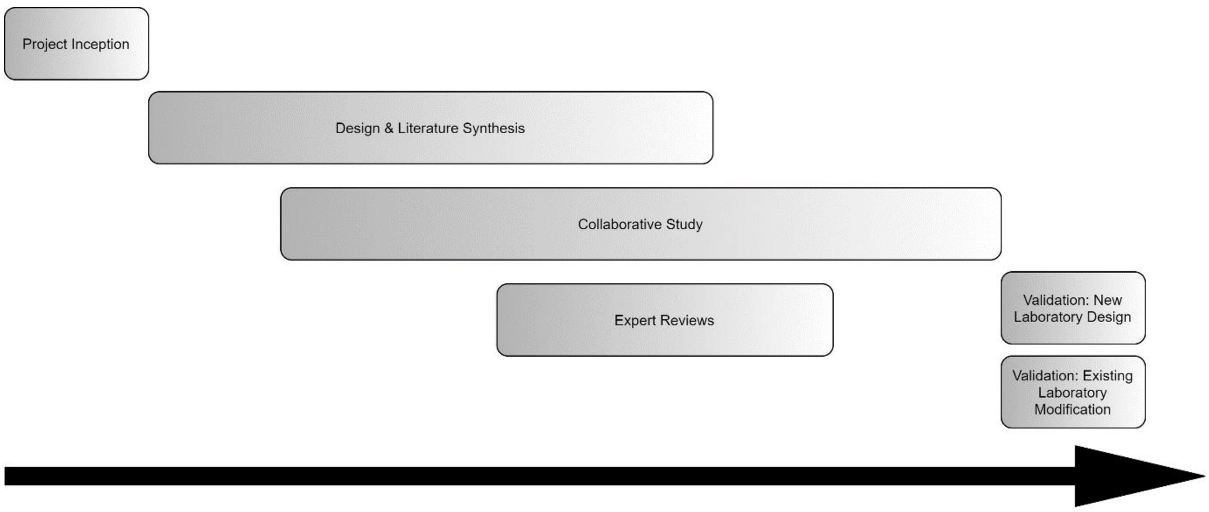


Figure 9: Phases of Educational Design Research conducted in this thesis

Accompanying and intertwined with the concrete development of the artifact was the research work progressively answering components of the overall problem statement for this thesis work. Below each of these phases are introduced in these pairs, with the research sub-questions, concrete development work of the artifacts, and the outcomes of these phases. The requirements for each phase to have successfully contributed to the overall iterative process are also specified. The details and internal methods of each phase are left to the following chapters which address them individually – here only the motivations and outcomes for each are included to provide the overall structure of the work as a whole.

5.1.1 Project Inception

This step, while technically outside of the EDR process itself, was nevertheless critical in its identification of the artifact goals described above. This process was motivated by the results of the researcher's prior work on laboratory evaluation [39] . This step primarily involved literature surveying outside the existing scope of engineering education, to identify approaches that may already exist to solve the identified issues with laboratory design and evaluation. The Critical Review component of this thesis was a part of this step.

The first research question being specifically investigated in this step was "How do the research methodologies used in evaluative research on laboratories in engineering education support or inhibit answering the questions this research poses?". The outcome from the overall literature review and Critical Review for this question has been described in the previous chapter, but for consistency this was that approaches which seek to isolate laboratories for their context, whether experimentally or using snapshot style qualitative investigations, inhibit the design and evaluation of laboratories in engineering education. Quantitative experimental designs lead to contortion, and the failure to meet inherent requirements in statistical methods damages the validity of the results. Qualitative methods may answer fundamental questions, but do not holistically answer design and evaluation questions.

The second research question for this step was: "What methodological alternative provides the appropriate characteristics to resolve the identified issues with experimental approaches in the design and evaluation of laboratories in engineering education?". Practical inquiry, and EDR in particular, addresses these limitations through the iterative triangulation using contributions from both qualitative and quantitative methods.

5.1.2 Collaboration Study

Purpose: to gain further insights through a deep dive into the application of the artifact in a real context with authentic motivations.

Occurring both in parallel with the informal consultation step as well as continuing afterwards, the researcher was involved in the design process for a new laboratory. This design process involved regular consultation among the design team, with a particular focus on potential methods for evaluation. This codesign process followed the Purpose-First Approach explicitly, with the researcher participating in the articulation of the laboratory purposes, identification of intended inputs, processes, and outcomes, and the eventual observations of these in context.

The relevant outputs of this step for this thesis were primarily theoretical in nature – that is to say the design of the artifact did not require further revision, but the understanding gained regarding the local evaluation theory was substantial.

The research question under study in this step was "Can Educational Design Research be successfully applied to the design and evaluation of laboratories in engineering education?". In this case 'success' has several meanings, within the laboratory project itself, and the research project for this thesis. Success for the laboratory project can be summed up as "meeting the requirements each stakeholder has in the laboratory" – in this case this turned out to be largely the case, with some limitations identified through the implementation which led to future improvements, particularly regarding the user interface.

Success in the context of the research project has two distinct scopes. There is the question of whether EDR was validated as applicable to the design and evaluation of laboratories in engineering education. Following proof by example, this study shows that the expected applicability of the process and effectiveness at answering the design and evaluation requirements of the laboratory stakeholders was achieved. It does not globally prove that in all possible cases this would be the outcome, but it validates that the expectations proposed by the existing literature in other disciplines held true for this example in engineering education.

There is also the question of whether the answers to the questions to this point sum to an answer to the overall research for this thesis. In this case the answer is no – but the phase did successfully provide sufficient information and insight to yield the next research question to motivate the next phase. This insight was that while EDR was successful as predicted by the literature analysis, it required a practitioner versed in EDR as an approach (the author) to be directly involved in the implementation of this approach, which is unlikely to sustain real methodological change in the discipline and thus does not fully answer the overall thesis question, “How may laboratories in engineering education be effectively designed to meet the purposes their stakeholders require, and evaluated against these purposes?”. And so the next phase investigated methods for detaching or replacing this requirement while retaining the benefits EDR provides.

5.1.3 Synthesis

Purpose: to develop the initial artifact through research synthesis.

The explorative development of an EDR artifact typically involves the integration of multiple participant stakeholders and perspectives. This thesis broke this phase into two components – the first being a synthesis of existing research in engineering education – this step – and a stakeholder integration step involving consultation and feedback.

This research synthesis step involved literature surveying within the discipline of engineering education to identify existing knowledge regarding the design and evaluation of laboratories in engineering education, chief among these being the previously identified [37] and [21], synthesising these into a holistic understanding, identifying gaps, and seeking to fill those gaps with further synthesis involving research from other disciplines. The external disciplines involved were substantial in number – science education (particularly laboratories), higher education research more generally, pedagogic theory, system and process design, HCI research methodologies, mathematics education, and nursing simulation providing the most material.

This synthesis generated the first versions of the Purpose-First Approach and PIPOV artifacts, as the primary inputs to the next steps – the visual component of which is Figure 10.

The research question this phase was focused on was “How can practitioners be supported in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?”. Frameworks embedding design principles having been identified in the literature as effective, this informed the development of the Purpose-First Approach, and the accompanying PIPOV Framework. This question would be the last question to be answered by the thesis, but doing so required candidate answers to be designed and tested as a second level into these sub research questions. Success in this case was focused on whether these candidate answers were able to promote the EDS practices previously seen to be successful in the collaborative study, without requiring this collaboration.

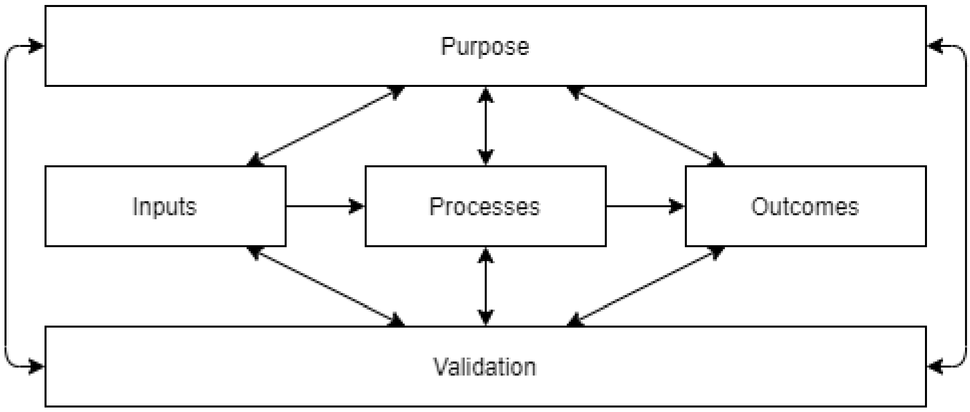


Figure 10: The initial version of the PIPOV Framework visualisation

5.1.4 Expert Reviews

Purpose: to explore the compatibility of the generated artifact with practitioners, and identify unresolved gaps or issues with the initial synthesis.

With the initial design ready for review, the exploration phase then integrated a series of consultations with practitioners, culminating in a workshop detailing the intended use of the Purpose-First Approach and accompanying PIPOV framework. This consultation followed a semi-structured interview format, wherein the existing evaluation techniques used by the practitioner were discussed, then the Purpose-First Approach and accompanying PIPOV framework introduced, with observations sought from the practitioner regarding their reactions to this artifact.

The workshop integrated this initial elicitation of existing methods used, followed by hypothetical case studies aligned with the different components of the PIPOV framework, with participants asked to consider this application in their own contexts and provide feedback. This consultation, workshop, and consideration of the obtained feedback led primarily to the redesign of the PIPOV framework:

This phase was necessary in moving from the prior research question "How can practitioners be supported in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?" into the candidate answer question "Do the Purpose-First Approach and PIPOV Framework support practitioners in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?" – however this step is rather a design process and does not provide research conclusions in isolation.

The learnings gained from this step can be summarised as "what elements and communication are needed to achieve the required performance of these artifacts". EDR does identify this as a learning and a research output – the artifact under study embeds the learnings which create it, even if this is in service to research questions rather than answering them explicitly. That is the case for this step, which does not answer any articulated research questions, but yet embeds research learnings from this iterative review and redesign process. It yielded a final visual component, displayed in Figure 11.

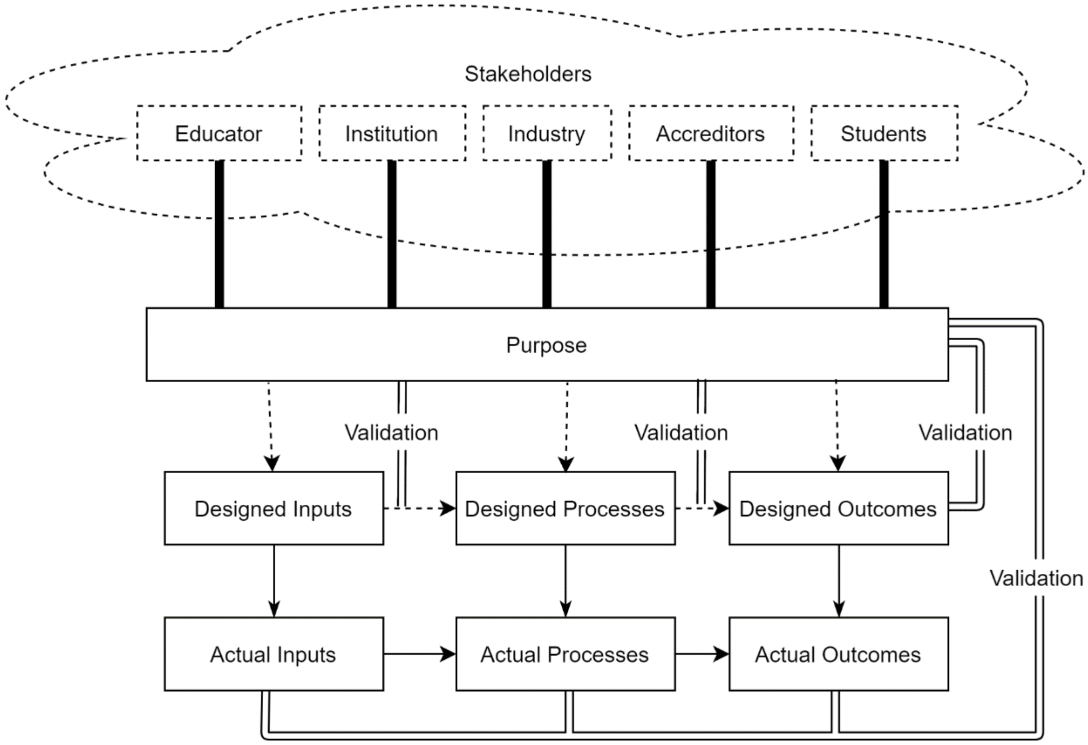


Figure 11: The final version of the PIPOV Framework Visualisation

5.1.5 Validation Studies

Purpose: to investigate the extent to which the current artifact design meets the design goals, and to reinforce the local evaluation theory identified in the previous steps.

These studies each target the two broadest divisions between specific stages of the design and evaluation process for laboratories in engineering education, to answer the question "Do the Purpose-First Approach and PIPOV Framework support practitioners in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?". This has similar types of 'success' as the collaboration study – success for the laboratories themselves being defined by the requirements of each stakeholder. This laboratory success is the overall goal in terms of 'impact' for this thesis work, but is not the research goal – it is unlikely to be the end result, but a real-world impact where engineering education practitioners were fully informed as to how to design and evaluate their laboratories to achieve success but chose not to would be a failure of the real-world impact of this thesis, but would still contain research success and contribution. The research success for this sub-question as it was investigated in the two validation studies is in essence 'did the practitioner academics using the Purpose-First Approach and PIPOV Framework apply EDR techniques to their laboratories without the author needing to either do this application themselves, or train the practitioners in EDR directly". In both cases this was observed – practitioners in each laboratory applied EDR techniques, specifically identification of purpose, identification of non-experimental white-box evaluative opportunities, and validating their intended observations of process and outcomes back into an iterative loop.

5.1.5.1 New Laboratory Designs

Purpose: to examine the integration of the artifact in a new context with other practitioners

The Purpose-First Approach and accompanying PIPOV framework have been integrated into a substantial new laboratory design project. This design project remains ongoing, but has yielded initial outcomes through group interview and consultation research activities. The key outputs of this step have been the elaboration of a potential "minimum viable methodology" as a subset of the Purpose-First Approach, as well as further insights into the communication of the details of the PIPOV framework.

5.1.5.2 Existing Laboratory Modification

Purpose: to examine the integration of the artifact for the purpose of modification, rather than new laboratory designs as in previous/concurrent research activity

The Purpose-First Approach and accompanying PIPOV framework have also been integrated into a laboratory conversion project. As with the other evaluation project this remains ongoing, but as before has yielded initial outcomes through group interview and consultation. The key outputs of this step have been refinement of the local evaluation theory for the markedly different context of redesign rather than initial development, as well as identifying a need for expansion of the input component of the PIPOV framework.

PART 2: EMPIRICAL STUDIES

6 COLLABORATION STUDY

Both this section and the later synthesis portion of this thesis are asking the overall research question "How may laboratories in engineering education be effectively designed to meet the purposes their stakeholders require, and evaluated against these purposes?" – involving deciding on educational design research as part of this answer, and refining this into an artifact with sub-questions "How can practitioners be supported in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?" and "Do the Purpose-First Approach and PIPOV Framework support practitioners in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?". The empirical work of the expert reviews and workshops served to provide evidence that the Purpose-First Approach and PIPOV Framework were effective at promoting the use of educational design research in practitioners untrained in that methodology.

Overlapping with this work was a long-term collaboration study, involving the joint design and evaluation of a laboratory in engineering education by the author, one of the author's thesis supervisors, and researchers in engineering education laboratories out of the University of Edinburgh School of Engineering. In the context of this thesis this collaboration study was in part to develop the artifacts, but also to provide further evidence for the core research question that was previously stated as another sub-question:

> "Can Educational Design Research be successfully applied to the design and evaluation of laboratories in engineering education?"

This question has extensive support from the literature across a wide range of contributing disciplines as explored in the earlier literature review and synthesis, however empirical evidence for this was also sought as part of this collaboration study. If the expert reviews and workshops can be characterised as investigating 'how' to communicate with practitioners to support an enhanced design and evaluation practice, this collaboration study can be characterised as investigating 'what' to communicate, with both questions supported through the literature synthesis.

The technical development of the laboratory in question is not being claimed as a research contribution by this thesis nor the Practable infrastructure, having been conducted by other members of the collaboration to the author, albeit with consultation and feedback with the author. The meta-investigation regarding the design and evaluation approach used however is the work of the author drawing on data collected during the collaboration. The author designed the qualitative analysis for the published work and performed this coding using a consensus model with the first author, and iteratively contributed to the design and analysis of the quantitative component.

6.1 LOGISTICAL CONTEXT OF THE COLLABORATION

The cross-institution, international team of researchers which developed to work on this project arose out of a common interest in research topics and a shared research background. The collaboration was multi-faceted, with the structure of the people working on the development of this laboratory being:

- A teaching academic at the University of Edinburgh with a need for a new laboratory
- University of Edinburgh teaching technicians – a more officially recognised and celebrated role in the UK when contrasted to laboratory technicians in the author's Australian context, e.g. [185], [186]
- The research collaborators at the University of Edinburgh, comprising a senior research partner with extensive experience developing remote laboratories, and a junior researcher who joined the collaboration mid-way.
- The research collaborators at the University of Sydney, similarly comprising a senior research partner with equally extensive remote laboratory experience (the author's thesis supervisor) and a junior researcher (the author)

The central points of contact for the project were the research collaborators at the University of Edinburgh, who liaised with the teaching academic to develop the laboratory requirements, the teaching technicians to physically build the bespoke laboratory equipment, and connected with the University of Sydney research collaborators to discuss the design and evaluation research, as well as overarching development work expanding outside the individual laboratory.

The nature of the collaboration was significantly impacted by the COVID-19 pandemic. Initial connections were made based on a shared research interest discussed below in early 2019, with the author visiting the University of Edinburgh to discuss research collaborations and potential longer-term exchange at the end of 2019. After the onset of COVID-19 preventing international travel, the collaboration shifted to purely online via email and regular video conferencing through 2020-2022.

6.2 Data collection

The collaboration was a deeply-embedded longitudinal study with the author involved in the ideation, planning, and implementation processes throughout the design and evaluation of the laboratory over multiple years (approximately 3, but ongoing) – albeit without physical presence. Data collection through this time included emails, design and discussion meetings held over video conferences, notes/journals taken by the author during these meetings, and artifacts generated by the collaboration such as laboratory web interfaces and evaluation survey instruments. In addition to this, a 'capstone' set of direct semi-structured interviews was conducted by the author with the Edinburgh researchers at a defined point in the ongoing process (the conclusion of the first semester of use of the laboratory and publication of the first research output).

6.3 Theoretical Context of the research study

The initial research trigger for collaboration between the researchers in Sydney and Edinburgh was around the topic of trust and/or veracity with laboratories. All 4 collaborating researchers have previous experience in non-traditional laboratories, and the question of trust, or framed differently veracity, in remote and simulated laboratories has been at the core of the argument for and against them since their inception.

Nickerson et al [37] strongly indicate this via their "laboratory frame" and students' perception of whether a laboratory is real, remote, or simulated. Lindsay et al's "Establishment reality vs. maintenance reality" [187] and related work [38] explicitly calls this out as necessary for there to be appropriate impact to learning by the laboratory activity. Consider Kolb's learning cycle described in the earlier section 2.8.1.3 - in order for a learner to conceptualise what they are experiencing and have that reflection change their knowledge or worldview, they must trust that what they are experiencing is 'correct' rather than dismiss it as in error, or fantasy.

Beyond this level of trust was a consideration that more than simply being believed as accurate, in order to learn from laboratories students must have a visceral response to the activity. This is often presented in various forms as the argument behind studying natural phenomena via laboratories rather than through theory, lecture, etc. – e.g. internal references 21-25 in [6]. This visceral response may thus provide the impact for the clash between existing knowledge and expectations with some difference in observed reality that precipitates learning. This necessary immersion and belief in the laboratory was termed veracity to contrast it with trust.

An important observation is that the collaborating researchers were all involved with non-traditional laboratories well before the onset of the COVID-19 pandemic. The laboratory developed as part of this research was a remote laboratory accessed via the internet, however this was always planned from before the pandemic. Remote laboratories were the primary topic of the author's visit in 2019. Thus, considerations of trust or veracity were consistently at the forefront of the design motivation, rather than being characterised as 'emergency online teaching' focusing on putting out fires and ensuring the minimum requirements are robust against disruption [188].

6.4 The Laboratory and Teaching Context

The more abstract research theme described above was later associated with a laboratory which the Edinburgh researchers were commissioned to develop in support of teaching colleagues at their institution. This laboratory was already in place as a proximal laboratory used in the class with physical equipment students interacted with directly. The motivation for change was due to robustness and scalability issues with the existing equipment – the proximal laboratory equipment involved bespoke modifications to a commercially available laboratory that had been in use for several years, and class sizes were increasing to a point this was difficult to support.

The laboratory was in a control systems class, jointly undertaken by undergraduate mechanical and electrical engineering students in the third year of their degree. The original design used Quanser Qube equipment similar to [189], in the spinning disk configuration which had been modified with custom equipment to increase the available configuration options and complexity of the available experiments.

The new laboratory being commissioned was required to inherit a significant portion of the context the previous laboratory was situated within – serving the same educational purposes in the class, following largely the same experiment guides. While this did not take advantage of any inherent affordances of remote laboratory designs for the design of the equipment, the increased access to the equipment via the internet rather than requiring proximity to equipment in a laboratory room was not constrained to the proximal laboratory circumstances and the study was able to seek the most effective designs in this respect. Further to this, the interface in which students accessed the laboratory was not constrained to the interface used for the physical equipment, and was similarly an available space for design exploration.

To illustrate this, the only change in the stated design and outcomes for the class the laboratory was integrated with was a single line, from

> "The course also has a hands-on laboratory (3 hours in total split into 2 sessions) which allows the students to get practical experience in working with a dynamic system and designing a simple controller."

To

> "The course also has an interactive lab component which allows the students to get practical experience in working with a dynamic system and designing a simple controller."

With the sessions similarly removed from the class schedule as access to the remote laboratories was not tied to a physical session in a timetable.

6.5 PIPOV ANALYSIS

A significant portion of the collaboration study was conducted with the author providing a context of educational design research as it exists in the literature and a nascent version of the Purpose-First Approach, not the PIPOV framework – as this work preceded much of the development of that framework. As they were developed in tandem however, the design will be communicated through this framework – leveraging the clarity this framework brings to communicating laboratories and providing a full example of a laboratory described through this framework.

6.5.1 Stakeholders

The stakeholders in the laboratory comprise the listed collaboration members – educators, local researchers, Sydney researchers, teaching technicians. It also includes:

- The institution of the University of Edinburgh and associated structures i.e. the School of Engineering.
- The Scottish government – accrediting the class containing the laboratory as SCQF Level 9 3rd Year Undergraduate [190], in alignment with the UK Engineering Council as an accrediting body equivalent to ABET or Engineers Australia, and through them industry taking on graduates as employees.
- The students undertaking the laboratory.

6.5.2 Purpose

Through these stakeholders the purposes of the laboratory were identifies. The government/accreditation bodies/industry, and much of the educator and institution purposes were captured in the service the laboratory was intended to provide against the stated Learning Outcomes of the class [191], numbers assigned for reference purposes:

1. Understand how transducers work and interface with a control system and appreciate the theoretical and practical limitations in any measurement;
2. Derive the transfer function of a linear (or linearised) dynamic system;
3. Understand the concept of feedback in control systems;
4. Analyse the behaviour of a linear SISO system in both time and frequency domains and assess its stability;
5. Design appropriate controllers for simple control systems to meet performance specifications;

For each of these learning outcomes the laboratory served as practical reinforcement of knowledge and skills taught in theory-based instruction throughout the class. The methods of e.g. designing appropriate controllers were taught through traditional lecture, tutorial, and self-guided work, with the laboratory being used as a practical testing ground to validate and reinforce these skills, rather than the laboratory being a stand-alone vehicle for teaching some new concepts.

These purposes above were all fully captured in the pre-existing face-to-face version of the laboratory which the remote laboratory replaced. The purposes motivating the design of this new laboratory were drawn from the educator, institution, technicians, and designers (the research collaborators as distinct from the academic directly teaching the class).

The educator, institution, and technicians recognised the increased pressures on the existing laboratory design in the face of increased student numbers and aging equipment:

- In order to support class sizes, already too high for the current equipment numbers and expected to increase, more laboratory equipment, space, and teaching time was necessary in order to continue with the existing design.
- The educator understandably wanted to continue using the laboratory style which was considered to be working successfully rather than redesign a new smaller scoped laboratory or remove it entirely.
- The technicians as well as the educator and designers involved with the existing equipment were aware of its limitations and the high cost and effort required to scale.
- The institution presented the purpose of more effectively managing limited real estate and student timetabling as resources – laboratory classes taking significant room which was under increasing capacity pressures, and the increase in student numbers making timetabling around clashes to see students physically in those rooms challenging.

The designers were in alignment with these purposes, but also had their own purposes for the laboratory that influenced the design:

- Acknowledging the affordances of remote laboratories, leveraging innovations in the user interface to more effectively deliver the original learning outcomes.
- Recognising the institutional purposes behind space and time limitations motivating this remote laboratory, and understanding these constraints being pervasive throughout the industry rather than being institution-specific, providing a platform to promote and support these remote laboratories more effectively.

6.5.3 Designed Inputs

The majority of inputs to the laboratory were not under the control of the design. The class would be staying the same, the material covered and the material the laboratory would need to support would be the same, the broad nature of the equipment, available degrees of freedom, method of interaction, etc. would be the same, and the institutional context would be the same. The overall context would change drastically, with the 2020-2022 timeline obviously being one of incredible flux in education, but similarly not at all under the designers' control.

The primary inputs of significant meaning in terms of design were the interface the students had on the equipment, and the corresponding change in student access to the laboratory dependent on the choice to adopt a remote laboratory, moving from 3 scheduled hours to 24/7 availability.

6.5.4 Designed Processes

The physical processes of the laboratory were modified from the hands-on interaction to web-enabled, however the overall experimentation process of reaching equilibrium though designing control inputs remained constant. Exploratory behaviour, groupwork, etc., were not included in the design of the laboratory or evaluation.

The cognitive processes were an open speculation that formed a significant part of the investigation. These were included implicitly in the previous design, with the laboratory intended to 'reinforce learning' and similar statements, without much explicit treatment of how this would occur. One critical reflection of this however was the appreciation of students moving their control decisions from pen and paper theory to real equipment, and how improper assumptions made on paper (including assumptions such as no measurement error which are valid assumptions on paper but not in reality) would yield unexpected results on real equipment. These unexpected results would then lead to reflection, investigation, and thus learning.

The open speculation regarding these cognitive processes with the remote version were tied into the trust and veracity trigger to the initial collaboration – if students need to see their mistakes on real equipment and believe them in order to learn from them, what happens if they don't trust the remote version the same way they do the version they can physically touch? Leveraging prior work on remote laboratories e.g. [175] it was expected that the remote laboratory had the potential to be effective, but as this thesis observes the lack of generalisability/transferability means this remained an open question prior to the initial run of the laboratory.

6.5.5 Designed Outcomes

The bulk of the intended outcomes were drawn from the learning outcomes described under purpose earlier. However, examining the outcomes from a logistical perspective was also necessary in order to serve the non-cognitive purposes also identified. The expectation was that the laboratory would be developed for a cheaper cost than the Quanser Qube with modifications, that it would take a lower real estate footprint, that students would 'spread out' their access beyond the earlier 3-hour window, and that students would respond positively to the remote laboratory paradigm in support of further adoption of these in the future.

6.5.6 Designed vs Actual

There was less potential for variance from the design inputs due to the inherently loose nature of the laboratory design, with less rigid structure around group interaction, timings, etc., and a significant portion of the inputs unspecified or unmeasurable. Some aspects of the inputs were capable of being measured, but were not – due to there being no identified reason why they might have varied and no observed effects to suggest that they had. For example, the class demographics were assumed to have remained unchanged based on the teaching academic who had also run the class previously not raising this as a concern. Some aspects of the inputs were known to have changed but were not investigated during the run of this laboratory as they were not the focus of the evaluation – for example, students performing all their work in the class remotely due to COVID-19, rather than just the laboratory. Some aspects that were previously assumed (e.g. individual work) were able to be somewhat observed by way of the interface (e.g. individual logins)

Designed processes were the specific focus of this laboratory as influenced by the UI and remote paradigm input changes – these were investigated in detail and will be discussed later. Designed outcomes were tied to the designed processes and will be included in this discussion.

6.5.7 Validation

Validation will be broken down into components matching the specific relationship lines in the PIPOV Framework

6.5.7.1 Validation of Designed Inputs -> Designed Processes

This initial validation step was conducted via a combination of informal observations regarding the existing laboratory and literature synthesis informing the design of the new implementation. While formally unexamined, the existing design was considered to be working by the educator teaching the class. Based on the informal methods of observation regarding a class 'working' common to many educators – submitted student assessment contents and performance, consultations with students, observations within the classroom, regular institutional feedback surveying – the design of the laboratory was not identified as an issue from a perspective of learning outcomes, notwithstanding the identification of logistical issues motivating the change.

The design of the new laboratory was informed by the literature, particularly through existing work on remote laboratories and their interfaces both by the collaborators and by others publishing in the field. Key elements of the interface motivated by this literature were the combination of live camera view of the equipment to establish the reality of the equipment, and the extracted controls and data capture elements presented separately.

6.5.7.2 Validation of Designed Processes -> Designed Outcomes

Validating that the designed cognitive processes would achieve the designed outcomes was similarly embedded in the previous laboratory, with the anticipation that if students trusted in the veracity of the laboratory that it would similarly result in positive learning outcomes.

Validating that the logistical outcomes would be achieved was based on feasibility analyses of the intended laboratory. This was done via a pilot physical design process and continued testing through the run of the laboratory.

6.5.7.3 Validation of Designed Outcomes against the Purpose

Validation of designed outcomes against the purpose was largely through collaborative reflection by the different members of the research team. Identifying the learning outcomes for the class, reviewing back through the design of the laboratory and in particular the design of the interface and any possible innovations that might be applied in support of this, and reviewing the physical traits of the equipment and web infrastructure and attempting to understand the implications for scalability.

6.5.7.4 Validation of the Actual laboratory against the Purpose

This step of validation was largely unmeasured at the time of the capstone interviews conducted to ‘finish’ this ongoing project for inclusion in this thesis. Rather than conducting summative evaluation at the end of the run of the laboratory, the evaluation process this project undertook was mid-point evaluation, focusing heavily on the movement from input to cognitive processes, considering this as a useful proxy to serving the intended purposes via the designed outcomes. By ensuring that the intended processes are occurring, subsequent phases of the iterative design process could be better informed for answering this question.

Consider the order of validation beginning with a traditional summative evaluation. In the event that the outcomes were validated to have been achieved and the purposes being served, a conclusion heavily prevalent in the literature using other methodologies is that this supports the use of the stated design elements – that is, the inputs. But this is not really what has been shown, as the inputs and in particular the processes they foster occurring have not been validated as occurring at all. The equivalent occurs in the event of the outcome not being served, only in this case the aspects of the design that need changing are not apparent, as the variation in designed and actual processes is unknown. This type of study is not invalid, but requires subsequent triangulation in order to arrive at actionable design knowledge.

Instead, by validating that the designed processes are in fact occurring, an evaluation can have confidence regarding the causes of any observed differences in outcomes. Follow-up studies are still needed once the designed processes are in fact occurring and have been validated as occurring, but there can be added confidence that this follow-up will be confirmed due to the literature that should be used to establish what the designed processes should be.

6.6 The Laboratory Design & Implementation

6.6.1 The Practable Infrastructure

The remote laboratory which was developed required the parallel development of a management infrastructure to connect students to the laboratory and manage this access, shown in Figure 12 and Figure 13. This mirrors infrastructure previously worked on by the author such as Sahara [40], but with significantly more support for modern technology and better support for the laboratory requirements. Existing remote laboratory infrastructure has suffered from a model of grant-funded and research student-driven development – leading to spikes in development of this system or that which 'fizzle' after key personnel graduate, retire, change institutions, grants end, etc. [192]. By the time of development of this collaborative project there was no available infrastructure to meet the logistical needs of the wider remote laboratory initiative required – existing out of date and limited platforms may have resolved the immediate learning outcome requirements of the single laboratory and class, but not the ongoing call for remote laboratory development in support of the identified real estate and timetabling purposes.

As such, a federated remote laboratory infrastructure was developed using an agile process which yielded an early functional prototype in time for the first delivery of the laboratory. The ongoing development details of this infrastructure may be found at https://practable.io/ and https://github.com/practable.

6.6.2 The Remote Laboratory

Remote laboratories are conceptually distinct to traditional laboratories in that their development is inherently separated between the 'equipment' and the 'interface'. As noted in [46] this distinction similarly exists in traditional laboratories, but can often and easily be ignored. The equipment for this laboratory exists as a shoebox laboratory, and there is an associated web interface for controlling this.

6.6.2.1 Physical Equipment

The physical equipment was an independently designed parallel to the existing laboratory equipment stored in what may be called a 'shoebox laboratory', an enclosure for a piece of laboratory equipment which is modular and takes up the minimum possible footprint to promote scalability.

Figure 12: Spinning disk control experiment [193]

Figure 13: Practable shoebox laboratories [193]

There are multiple copies of the same laboratory equipment which combine to form 'the laboratory', similar to purchasing multiple kits for a traditional laboratory. Before producing these copies, the equipment went through a pilot design process to ensure it met the requirements of the laboratory. This process was largely a study in the practice of engineering itself rather than being specific to education, but highlighted the value of including skilled technicians in the laboratory development process. The quality of the final laboratory was enhanced through this pilot process identifying likely modes and causes of failure, and ensuring that the non-cognitive purposes of low cost and high reliability for scaling was achieved.

The nature of the shoebox format led to the laboratories being highly mobile. This proved to be of significant benefit during the heaviest impacts of COVID-19, with some of the laboratories able to be located in personal homes during lockdown while the design process was underway. The locations the laboratories are now stored in represents the possibility of unintended purposes such as students potentially visiting the laboratories to increase their veracity.

Observed issues with the laboratory through the design process that had the potential to impact their effectiveness were failure rates, fungibility, and the risk of uncaught error. The laboratories were generally reliable, but were produced from largely consumer-grade parts and easily fabricated modifications to avoid the lock-in experienced with the earlier hands-on equipment. This meant that tolerances needed to be increased to avoid the equipment failing, as the remote nature of the equipment meant that these failures could not be as easily corrected on the spot.

Fungibility of the different laboratories was also identified as a concern, with students potentially expecting that they could develop some control inputs at one time with one version of the laboratory, come back to another version in another session, and continue with the work on the new equipment unchanged. Slight variations in the weight, distribution, friction, alignment etc. of the specific components meant this was not the case. It was observed that this was likely to be a beneficial observations for students to make in regards to the ‘messy’ nature of real world measurements, and so rather than try (certainly futilely) to make the different versions behave precisely the same, or to automatically compensate for variation in the controlling software, it would instead be better to call out to the students that the equipment was *not* in fact fungible, through the use of numeric signs prominently displayed in the camera view on each of the laboratories. In this way the risk of misconceptions by the students was leveraged into a design which was anticipated to yield learning benefits in support of the laboratory’s purpose.

A remaining risk regarding errors in the laboratory equipment was identified through the design process – if a laboratory had an error, but was still mostly able to function, e.g. a sensor reading was out, this would damage the data being collected. This is of course a risk of any laboratory, and is less likely for a remote laboratory where students are not responsible for setup and teardown of the equipment. It is however speculated that the ‘benefit of the doubt’ might cause different interactions with the equipment remotely than proximally. For example, when obtaining an unexpected measurement with proximal equipment the students may assume they assembled something incorrectly and ask for help correcting it, which would lead to identifying the error. There are different and often unrealistic expectations of software [194], which leads to the potential outcome of students being more likely to assume that the issue lies with their inputs and understanding rather than the equipment – and losing trust in the interface once this was proven incorrect. To mitigate for this a combination of the system regularly checking for programmatically identifiable errors (such as a broken video feed), coupled with communication support to the students regarding e.g. what to expect from the first task using the equipment.

6.6.2.2 Web Interface

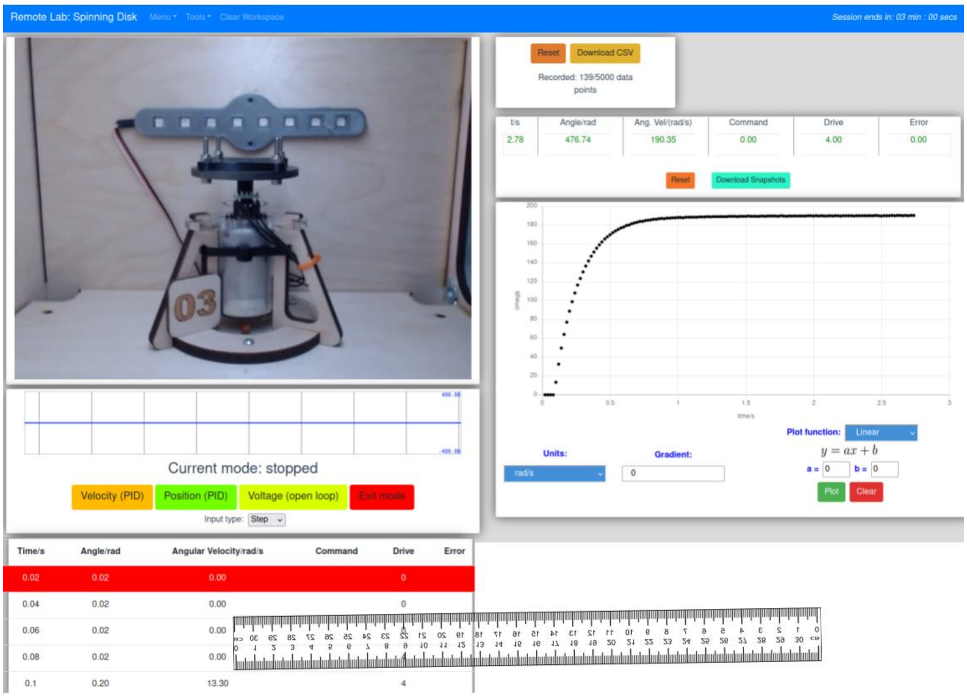


Figure 14: Remote laboratory web interface [192]

The remote laboratory was made available through the Practable infrastructure via the internet. The specific user interface was informed by pre-existing work on remote laboratories as mentioned above in validation section 7.1.5.2.6, following a pattern of separate internal UI elements featuring a video feed, raw data inputs and outputs, and a focused visualisation, as in Figure 14. These elements can also be seen in the prior Truss laboratory work in [39]. The UI was designed as a baseline, following current practice with the intention of undertaking specific evaluations on the UI to inform subsequent designs.

Users were generally expected to interact with the equipment via a PC or similar large format screen device rather than a small format such as a mobile phone, and interaction while not entirely precluding touch input was designed for mouse and keyboard input. This was in part motivated by the standard used for remote labs previously which was similar, by the ease of quick implementation, and by the large number of potential other materials used – the class learning management system with associated learning materials, textbooks, other websites, etc. that would promote the use of a large screen device regardless. This was not explicitly called out to the students in evaluations, but was investigated as part of the UI overall.

6.7 EVALUATION

The laboratory was evaluated via a combination of formal and informal processes, according to the nature of the attribute under study and the level of focus being applied to it. The logistical aspects of the laboratory were able to be observed directly without requiring research effort – costs to produce, real estate requirements, Practable infrastructure requirements, relieving time pressures due to increased access. These were observed to meet the requirements set out by the purpose of the laboratory and so these aspects of the laboratory purpose were considered achieved.

The learning outcomes of the laboratory were not evaluated invasively – the class assessments involved were undertaken as normal throughout the class by the educator, and the remote version of the laboratory was not observed to have caused any significant changes to student performance in these assessments. This is not considered to be evidence confirming that this purpose has been met, but rather that this purpose is maintaining its current status quo and is unlikely to have been significantly damaged by the design changes.

The aspects of most significant interest when developing this laboratory, motivated by the observed inherent differences between proximal and remote laboratories, was the students' trust in the veracity of the laboratory, and the influence of the user interface on this. As a result, these aspects were evaluated directly with a dedicated survey instrument. This survey instrument, reproduced as Appendix G included questions seeking both quantitative and qualitative data.

Ethical review of this study design was obtained via the University of Edinburgh's Research Ethics & Integrity assessment process by the University of Edinburgh researchers who retained access to and control of raw student data.

The quantitative questions were drawn from the User Experience Questionnaire UEQ [195], a pre-validated instrument measuring usability, as well as bespoke survey questions targeting the veracity considerations the study was investigating, as well as direct comparison between students' perceptions of remote and proximal laboratories.

The qualitative questions were paired with the quantitative questions to invite deeper reflection and explanation of earlier responses.

Of the N ≈ 250 students enrolled in the class and sent the survey, N = 24 responses were received, with N = 21 being internally valid and used in analysis. This response rate again reflects the challenges in invasive evaluation techniques that require participant engagement and limits confidence in the eventual conclusions. Triangulating with the educator reflections on student learning and the iterative nature of the overall process however means a lack of confidence in a single snapshot of data may be mitigated.

The detailed observations of the evaluation are presented in [192]. The relevant conclusions in light of the articulated purposes of the laboratory from the quantitative data were that the UEQ results were generally positive, however with a noticeably reduced score for dependability – which correlated with specific statements in the qualitative data regarding bugs and other observed issues in the system. The moment of laboratory outcomes not matching expectations as a trigger for learning was part of the bespoke quantitative questions, and indicated a very wide range of opinions with significant misconceptions identified – flagging this as an element requiring greater support in later designs. Student perceptions on proximal and remote laboratories after engaging with the remote laboratory were not conclusive.

The qualitative responses were coded against the different scopes of issues that may exist in a laboratory system, noting that these are issues as perceived by the students and may not necessarily be flaws – conceptually something that annoys a student may be necessary for learning:

- Technical issue – a bug or error in the software or hardware that was not intended to occur by the designers.
- Design issue – an observed issue in the way the laboratory equipment or interface functions, which while not necessarily explicitly chosen by the designers was not a bug.
- Teaching issue – an observed issue with how the laboratory was integrated into the class. The laboratory guide and what they were asked to do, the surrounding teaching context, etc.

These categories showed significant links to the quantitative data, with students discussing technical issues responding more negatively to other aspects of the survey, in keeping with [194] as previously speculated. Collecting these statements into the listed scoped themes also allowed for identifying common areas with stated recommendations that students described differently. The layout of components and ability to manage screen real estate was a common root cause to some stated observations by students, and within-interface guidance was also identified as a likely positive change.

The low response rate to the survey and in particular the high bounce rate with participants opening but not completing the survey indicated the survey was too long, which was itself an observation to be integrated into the design. Leveraging the ideas in [76] a design motivation was identified to combine passive observation through interface interaction data with local micro-evaluation questions within the interface.

6.8 Developing Engagement with Educational Design Research

This collaboration provided a vehicle for the investigation of Educational Design Research in authentic practice. As part of the author's contribution to the laboratory design and discussions, the process followed was guided by the Purpose-First Approach, and Educational Design Research as a specific methodology – though notably not using the PIPOV framework until the capstone interviews given this did not yet exist.

For the majority of the collaboration, promoting the EDR methodology was more about preventing changed behaviours rather than changing them. The design practices the collaboration team employed already fit many EDR practices without requiring significant change or upskilling. This may be summarised as

- Taking a literature informed view of design, seeking precedents, evaluations, comparative studies for potential design decisions and incorporating those into designs for a teaching and learning activity.
- Iteratively developing and reflecting on observations when undertaking the design.
- Incorporating observed evidence into new versions of the design each iteration.

As has been highlighted in the literature review and the expert review sessions, the EDR process is one that engineering practitioners are already very familiar with, and appear to follow almost as the default. Only perceptions of what research 'should' be that preclude EDR practices appear to drive researchers away from these.

There is more to EDR than simply a design process, however, and this difference is where significant gaps were identified between the informal design process used by the educators and the requirements of EDR. Explicitly identifying the cognitive and physical processes occurring during the laboratory to craft a local instruction theory rather than treating the activity itself as a black box was recognised as a flaw in the 'default' engineering educator skill- and mindset. Appreciating the difference between the intended design and the design in practice (designed and actual in PIPOV) was similarly not automatic.

An observation should be made here regarding the collaboration participants. Both the Edinburgh and Sydney participants were led by engineering educators with decades of experience in engineering education, laboratory design, remote laboratories specifically, and most importantly with extensive backgrounds in engineering education research rather than simply engineering education practice. This may have implications to the observed 'default' practice in that the participants were representing best practice, rather than common practice. The junior researcher in Edinburgh was similarly aligned however, so there is no inherent requirement for this level of experience.

The most significant consequence of the difference between EDR with formal identification of local instruction theories and 'unpacking' of black-boxed activities lies in the ability to disseminate the findings of the design process as a design study. The reflective practice applied to the design was able to identify issues through the early ideation process and incorporate scholarly literature to motivate design choices, but the lens of this process was inwards, not outwards. Thus, a significant portion of the author's development of an EDR process to the direction of the collaborative work was raising these as design elements to discuss, embedding where and how these would be examined empirically, and identifying ways in which they can be reported.

Reporting the laboratory design process and findings was the most significant point of dissent between the established practice of the involved collaborating designers and the practices required by EDR. Several design decisions which came about as a result of the collaborative discussion, reflection on observations made during the design process, and were informed by the literature, were not initially considered as learnings to be included in a research paper.

Discussions around what could be included as a research output began from a space typified by summative evaluation rather than contemporary observation. The general structure of designing and developing the laboratory, deploying it in a teaching context, and *then* the research on the laboratory would begin. This is obviously the more common approach in current laboratory design and evaluation practice as described in the earlier literature review, but a shift in perspective and assumptions was required to engage with EDR as a methodology.

Within the collaborative team this shift did not prove to be a significant challenge. There was a need for evidence behind the methodology to make the argument for its adoption convincing – this evidence being sourced from the same EDR literature reviewed for this thesis. Once the ability of the EDR methodology to establish rigour was made clear this process became the overarching methodology for the project. The elements within PIPOV described for this laboratory above were generated through this methodology with guidance from the author rather than through any direct discussion of the PIPOV Framework itself.

As observed in [92, p. 255] there are significant challenges when reporting educational design research, and few publishing venues in the education discipline, let alone engineering education, commonly publish literature following a design narrative style. Due to the nature of the venues for publishing (work in the main paper cited here was presented first during the Practical Engineering Education Conference 2021 [196]), the initial laboratory design paper was presented capturing that specific iteration of the project rather than the design process. This does not limit the potential effectiveness of the paper itself, but does limit the adoption of Educational Design Research as a methodology due to significantly reduced visibility.

6.9 Collaborator Interviews

The collaboration study had been running for nearly 3 years, beginning with topic discussions around trust and veracity and the development of the Practable infrastructure, leading into the design and implementation of the remote laboratory, the design of the research evaluation survey, the laboratory running through a semester of teaching and finally the analysis of the first results and publication of the initial paper. While the nature of the EDR project was iterative and the laboratory work would ongoing, it became clear that the learnings regarding the practice of EDR was reaching saturation, and so a ‘signing-off’ on the project within the scope of this thesis was needed.

This was implemented as a combination of author reflections on the written notes/journals and generated artifacts from the laboratory of which the above analysis is the output, as well as a pair of dedicated semi-structured video conference interviews with the University of Edinburgh participants. These interviews were a tonal shift in the nature of the discussions that were regularly happening from ‘let’s talk about our remote labs project’ to explicitly taking the researchers through the Purpose-First Approach – which they were already familiar with through use in the remote laboratory study, and the PIPOV Framework – which they were aware of but was not a component of the remote laboratory study and had not been discussed in depth.

These interviews generally followed the same protocol as with the expert reviews, though as may be expected the familiarity between the interviewer and interviewee led to some questions being skipped (e.g. details of how the laboratory worked), and a deeper dive into particular aspects of mutual interest (e.g. UI design and evaluation in computer-mediated laboratories).

These interviews/discussions resulted in the following insights that were not already encountered during the collaboration:

- The lack of direct student contact for research into their cognitive processes (observations, focus groups etc.) was a significant gap, which was not pursued at the time due to COVID restrictions.
- The lack of evaluation prior to the implementation of the remote laboratory was also a gap, causing the lack of a baseline to compare to and thus limiting the understanding of the current implementation.
- Discovering that the students felt like they were ‘in control’ of the laboratory, even though it was computer-mediated and remote and prior research had speculated that this was a limitation of remote laboratories, raised interesting questions about what else might be established for a specific laboratory in its own context rather than trying to generalise too much from the literature.
- Leveraging the interaction data in ways explored in [192] (i.e. classification learning analytics), is a paradigm-defining advantage for remote laboratories that requires further exploration.
- The collaborative nature of the study, including the split between the educator, teaching technicians, and laboratory designers/researchers in Edinburgh as well as the split between the researchers in Edinburgh and Sydney, led to significant parts of the PIPOV Framework being developed without the framework being used as a motivator, e.g. having the educator explicitly articulate their purpose for the laboratory.

6.10 Discussion

The collaboration study provided confidence that the selection of Educational Design Research as the answer to the question of designing and evaluating laboratories in engineering education was appropriate. The alignment between the practices involved and the standard practices of the practitioners meant that issues of contortion of the design to the research rather than the design to the best outcomes identified in the literature were not a concern. The study overall has had a strong influence on the design of the PIPOV Framework, and provides empirical evidence for the structural decisions it incorporates.

A key observation made during the collaboration and reinforced as part of the capstone interviews was the impact of simply being a collaboration to begin with. As observed above, this meant that aspects of the laboratory such as the purpose and especially the stakeholders were inherently made explicit as a necessary part of designing the laboratory at all. The motivation for a new laboratory could not begin without identification of the necessary stakeholders. The designers couldn't design anything without the educator articulating the purpose.

One of the identified purposes behind the use of frameworks was to aid in communication. By proving a common structure or language with which to discuss a complex topic, ideas can be conveyed between people with a shared understanding of that framework more easily and effectively. Collaboration between people requiring the population of a framework's elements even when that framework was not being applied strongly suggests that a shared understanding of that framework would make similar communication easier and more effective. Consider the lack of explicit purpose hindering the usability of research on laboratories in engineering education described in the literature review and especially in the vignettes. If collaboration of the sort used in this study were used, those purposes would have to have been made explicit, and so improve the research – or, alternatively, if the PIPOV Framework was applied and did the same.

Another key finding was the nature of the laboratory research. One interpretation of Educational Design Research resolving the issues encountered in the telescoping bias study might be that it must have some way to indeed measure all of those confounding variables, just not quantitatively. This is not the case – the way EDR resolves the encountered issues is by *not needing to measure them at all*. This is not to say that the aspects contained in those variables are not important, even critical for certain questions, but what EDR allows is for focus to be applied to the specific aspects of the laboratory which are of concern. What can be observed should be reported, what is under study should be observed, but what is not observed due to some constraint does not mean that there can be no design work. There will always be attributes of interest in any laboratory design that are simply impossible to observe, as a direct consequence of the telescoping bias and factorial sample size discussions – but an EDR project can continue to progress without needing to wait for their value to be known.

Take for example the students' cognitive processes during the collaboration study laboratory. These were not studied directly, but rather a baseline or general diagnostic was used in the form of the academic performance of the students, and the informal observations by the academic. So long as these diagnostics give researchers no reason to believe that an issue is occurring out of view, and so long as this is reported as a diagnostic rather than certain knowledge, the study remains valid. This occurs with an indefinite number of variables in any study in education: so long as there is no reason to believe that the colour of the whiteboard marker used influenced student learning, for example, there is no need to track the colours of whiteboard markers used.

An argument against this might be "what if we discover that there might have been an influence later?". This is absolutely a reasonable argument pertaining to validity for summative, 'one-shot' methodologies which take a snapshot and generalise or transfer these outcomes to other contexts. But it does not apply to iterative, contextualised, satisficing methodologies like EDR. Consider the elephant in the room for the collaboration study laboratory, COVID-19.

It could easily be reasoned that this invalidates the findings for any laboratory not being conducted through the COVID-19 pandemic. Participants, educators, researchers were undergoing all sorts of new stresses and circumstances. The outcomes e.g. regarding laboratory preference between proximal and remote may easily have been influenced by students desperately wanting to get back to campus, or just as desperately wanting to stay away from crowds. The findings are therefore not generalisable because there was no control group of students undertaking the laboratory in some COVID-19-free context.

Except that those outcomes in EDR are outcomes for the laboratory under study, and must be fed back into that laboratory in order to improve that laboratory. They are always invalid for any laboratory that is not that laboratory, at that time, under that context. What is generalisable are the design processes leading to those design decisions, what was observed to trigger a belief that a given cognitive process was occurring, how some attribute of interest was measured and how its value changed under different circumstances. These findings are equally as transferrable as they are able to be contextualised when reading the findings and the explained context, and they are what is useful for informing the design and evaluation of other laboratories. Educational Design Research is about holistic reporting of what is known when practicing design, not exhaustive data collection.

7 ARTIFACT DEVELOPMENT

7.1 THE PURPOSE-FIRST APPROACH & PIPOV FRAMEWORK

7.1.1 Context

There is fundamentally a chicken-and-egg issue with the narrative of this thesis. Rather than a traditional experimental design involving literature review, study design and execution, then analysis, the Educational Design Research-based meta-methodology for this thesis involved these 3 tasks occurring in parallel throughout, following iterative micro- and macro-cycles of design, execution, and reflection, with inputs to these cycles including literature synthesis and empirical studies. The artifacts in use through this process were fluid, with as many versions as there were cycles. As a result, what is presented here is the final design that was generated by the end of the empirical studies described in the subsequent chapters. It encapsulates and takes advantage of the learnings that will be described.

The thesis work that provided design contributions to these artifacts were:

- Literature synthesis- including the empirical truss laboratory study conducted prior to this thesis.
- Expert Reviews & Workshops
- Collaboration Study
- Validation Studies

7.1.2 Critical Synthesis Sources

There are several levels of adoption of contributing material which has influenced the creation of these artifacts. Several sources have had a major individual impact and can be readily identified in the final artifacts/

- The first of these was the recognition of a need for contextualised research on laboratories to provide an anchor for analysis, as a part of [39]. This is the origin of the term "purpose-first", though at this early stage this phrase held quite a more limited meaning.
- Also originally encountered in [39] but applied very differently was the combination of research by Nickerson et al [37] and Brinson [21].
- The Input-Process-Output from systems theory [197] was also critical in connecting this research.
- The artifacts are opinionated about and embed the cognitive psychology and educational theories of Piaget and Vygotsky among others via Biggs' Constructive Alignment [171].
- The final and likely most significant impact on the design was Educational Design Research as a research process or methodology [109] – not just as applied as the methodology for investigating the artifacts, but embedded in their intended use.

The artifacts further incorporate knowledge from the literature and empirical work. This has largely been incremental in nature rather than structural, informing the manner of presentation, the contents of different elements, and how practitioners might best be supported in their use.

7.1.3 Design Discussion

7.1.3.1 Purpose, Assessment, Evaluation

The fundamental claim behind the Purpose-First Approach is that the processes of design require evaluation, and the process of evaluation requires some way to assign value. As previously explored, design is a process that involves iteration around feedback – incorporating feedback into how something is planned or constructed in order to make it better. This process of acquiring and incorporating feedback is an application of the process called ‘evaluation’. A designer must somehow assess the current state of a design, find problems with it or opportunities to improve it, and take action to address these.

Evaluation is a sometimes loosely defined term which is used in many different contexts and with many different specific interpretations [198]. In education, it can often be conflated with ‘assessment’ given the English definitions of these words, which the author encountered extensively through conducting the earlier literature review. For example, from the Cambridge English Dictionary:

- Evaluation: the process of judging or calculating the quality, importance, amount, or value of something [199].
- Assessment: [200]
  - the act of judging or deciding the amount, value, quality, or importance of something, or the judgment or decision that is made.
  - the process of testing, and making a judgment about, someone's knowledge, ability, skills, etc., or the judgment that is made.

This thesis takes the definition of ‘assessment’ from Biggs’ Assessment Tasks, to mean tasks given to students in order to identify their level of mastery of learning outcomes, provide grades, control whether a degree may be awarded, etc. ‘Evaluation’ is thus considered only in the context of ‘evaluation of a teaching and learning activity’, and is defined from the dictionary version above ‘the process of judging or calculating the quality, importance, amount, or value of a teaching and learning activity’.

Based on this definition of evaluation, it is clear that there can *be* no evaluation if there is no context for the judgement or calculation. There is no way to judge or calculate quality without a way of identifying what quality is, no way to judge or calculate value without some lens by which to interpret some outcome as having value or not. Thus, there can be no evaluation without this context being present.

It has been commonly observed in the earlier literature review and earlier work, however, that much laboratory research is conducted *without* a stated purpose – and yet appear to evaluate the laboratories under study, to assign value. The reason for this is that there is an *implicit* purpose behind evaluation whether one is articulated or not. If something is stated as being valuable and this position is argued with evidence, there is a path backwards from that statement and evidence to identify what it is that is being claimed as value. In the case of a laboratory evaluated through the use of an affective survey, the students’ affective response to the laboratory (liked it, did not like it) is implicitly being upheld as the ‘purpose’ of the laboratory, whether the evaluator would agree with that being the purpose or not. There may be arguments to be made regarding some flow-on purpose where the students’ affective response leading to improved learning, but where this is not stated and evidenced it is not appropriate to assume.

This was largely the position which was arrived at in [39] regarding the *reporting* of laboratory evaluations, from a perspective of sense-making of those reports by external readers deciding whether their own laboratory was intended for a similar purpose. There are further-reaching consequences, however, which lead to the Purpose-First Approach.

7.1.3.2 Design as a Process; Satisficing

The literature exploration that began as part of this thesis work involved an investigation of ontologies, epistemologies, methodologies, and methods. This work was initially focused on developing a methodology for the thesis as a research project, and Educational Design Research was selected from practical inquiry methodologies out of that process.

After this selection it quickly became apparent that Educational Design Research was relevant to the design and evaluation of laboratories itself, and this led to the expansion of the original idea of purpose-first from a statement regarding research reporting to a recommended process and theoretical framework. Design is iterative, involving that iterative cycling with feedback and reflective evaluation stated earlier. Evaluation research on laboratories encountered as part of the literature review, however, was overwhelmingly *not*. Summative evaluation of a single snapshot of a laboratory using contextless affective responses, or associated student assessment scores as evidence dominates the research landscape, while the empirical work has made it clear that this is not the way laboratories are actually designed in practice.

Further to this, the type of evidence being required was also mismatched. Literature on laboratories were attempting and universally failing to provide externally valid experimental designs proving causation and effect size of design decisions on laboratory outcomes. As an example, if this approach was working as intended it would be possible to state ‘if you need to students to master learning objective x, then you need to use laboratory design y which has been proven to yield this result’ – with certainty that this would then be what you as an educator observed in your own context.

As this is not the case, with laboratory research findings of no statistical significance, or without controlling for confounds, or with unestablished external validity, designers of laboratories do not have this knowledge bank of facts like the designers of a bridge choosing a type of steel would. Thus, given that laboratories are nevertheless being designed and implemented in engineering education, it must be the case that laboratory design does not follow the quantitative experimental process that is being embodied in the evaluation research. There is a fundamental and crippling difference between the way laboratories are designed, and the way they are evaluated – when these processes should be inseparably linked.

In order for design and evaluation of laboratories to be effective, it is claimed that embedding an effective design process into the evaluation process is required. This embedded design process is part of the Purpose-First Approach.

7.1.3.3 Frameworks as Vehicles

Educational Design Research is not the creation of the author. Neither have they been the first to recognise that purpose is important in design, or that iteration is necessary as part of a design process. Most of these are practices that engineering practitioners are intimately familiar with as a part of being a good engineer. For some reason, this has not led to these practices being the default for laboratory design and evaluation publications in engineering education literature.

It is hypothesised as part of this thesis work that there are no inherent limitations to this adoption by engineering education practitioners, and that it *is* a drastic improvement over current practice, but rather that the barriers are related to barriers to adoption seen in other studies of new 'things' being adopted – perhaps the most heavily researched being new technology. The Technology Acceptance Model [201] is an early example of several proposed models for the adoption of new technologies, e.g. VCR players. Design features yield potential adopters' perceptions of the usefulness and ease of use, which leads to their attitude towards use, and subsequent adoption.

These concerns translate readily to a model of adoption of EDR and similar iterative processes – engineering educators have been observed through the literature review to be trained in positivist epistemologies, used to quantitative methods and quantitative certainty in their discipline, and are used to seeing quantitative methods in the existing literature on laboratories. This leads them to employ these same techniques in improper contexts, because they are either unaware of more appropriate alternatives, or do not perceive them as being easy to use, or effective.

This gap has led to the development of the PIPOV Framework as a way for engineering educators to be exposed to Educational Design Research as a process for laboratory design and evaluation without requiring the years of literature review and synthesis that the author undertook on this. The framework is intended to allow practitioners to apply EDR techniques without needing to be familiar with the specific details and requirements of EDR, the cognitive psychology which underpins it, or the contributing frameworks that it contains.

7.1.3.4 Frameworks as Lenses; Frameworks as Communication Media

The recognition that there needed to be some artifact to support practitioners in applying the Purpose-First Approach for evaluation provided an opportunity to support the other 2 aspects of that task as well – design, and reporting.

In [39] the author used two contributing components to the framework – Nickerson et al [37] and Brinson [21] – for the purposes of evaluation of a laboratory and attempting to capture the sum total of contributing inputs for given outcomes. However, this set of potential inputs had the potential to contribute to the design process as well. By providing a way to compile and make sense of the different available design decisions in a laboratory, these decisions could be supported through associated material – particularly the experiences of others applying the same framework.

On the reporting side, by leveraging a framework the shared understanding discussed in the literature review section 2.6 would allow for more effective communication about the laboratories being studied. This would improve the process of design and evaluation by taking the current, less effective literature on laboratories and enabling a reader to recontextualise them using their own classes and purposes. It would also mean that literature synthesis as observed in the vignettes would be far better supported and so more generalised knowledge could be derived by the collective efforts of the wider discipline.

7.1.3.5 Two Hard Things

> *"There are only two hard things in Computer Science: cache invalidation and naming things."*
> -- Phil Karlton via [202]

The development of the Purpose-First Approach began with its name being readily provided by the conclusions from which it was derived – it communicates exactly what it needs to, is provocative in what it demands from a practitioner, and as later found in the empirical studies it 'just makes sense' to engineering educators.

The name for the proposed framework was not as simple. While this decision was potentially less impactful than the other influences on the design, in keeping with EDR practices as it was a design decision it is included in the discussion.

The primary motivation for considering the name was to avoid the 'Google Problem' that plagues practical inquiry methodologies. It is nearly impossible to seek out material on design experiments, design-based research, educational design research, etc., unless you are already familiar with the traditions of researchers involved, can follow their citation chains backwards and forwards in research history, and know the different publishing venues that are associated with these. Attempts at systematic literature reviews are significantly hindered, communicating one's work in the space is less effective, and misinterpretation of what is being practiced due to mismatches in that shared understanding can occur. Having developed a decades-long background in research and out of that process having multitudes of potential names associated with it, it is unlikely at this point to be corrected to some single, distinguishable term [203].

Thus, the term 'purpose-first approach' was first examined in different search tools to determine its suitability – Google, Bing, Google Scholar, Scopus, Web of Science. The term 'purpose-first' exists in other contexts with an aligned meaning, however Purpose-First Approach with or without association with laboratories yields no results.

With this as a benchmark, the acronym PIPOV was generated out of the synthesis of Nickerson et al [37], Brinson [21] and IPO from systems analysis [197], which also created the initial visualisation (section 7.2.2.3.1). Benchmarking this term yielded a rare spelling of a Russian surname, and an outdated and no longer in use ray-tracing software package [204], and so this was selected moving forward.

7.1.4 The Purpose-First Approach

The Purpose-First Approach to Laboratory Design and Evaluation in Engineering Education can be expressed as a series of defining statements:

- Laboratory quality and effectiveness is defined by a laboratory's ability to meet the purposes to which it is applied.
- Effective laboratory design requires iteration.
- Iteration involves phases of ideation, planning, execution, observation, and reflection.
- Ideation involves the acknowledgement of some trigger for what is occurring. There is some reason behind a new or changed laboratory project that motivates it.
- Planning involves the development of a design that is intended to meet the needs of or revealed by the previously observed motivation – the laboratory's purpose(s).
- Execution involves running the laboratory in an authentic context.
- Observation runs in parallel with and subsequent to execution. Both it and the Reflection phases are done in the context of the laboratory purpose(s) which define value.
- Observations are considered as part of the laboratory design process rather than added on separately, and must not require the laboratory to be controlled or otherwise modified solely to meet data analysis requirements. Avoiding modification from data *collection* requirements would also be nice, but is not always possible.
- Reflection takes the findings from those observations, feeds them back into the ideation and planning phases and begins the process again.
- Reporting the results of a laboratory design includes all iterations and all phases of the design, not just the observations and reflections of a single run.

- Perfect knowledge or a perfect design are fundamentally unattainable – the results of the process will be an iteratively improved laboratory, not a perfect one.

While the above defines the approach, it is likely not an approachable definition, and so a more digestible description/prose discussion useful when describing the approach to practitioners has been:

> The Purpose-First Approach says that when you are designing and evaluating a laboratory, you need to start with the purpose the laboratory is being used for. Why is the laboratory there? They take time, and money, and effort to run, so what is it doing? Do students learn more? Are they happier about doing it and more willing to spend time doing boring calculations? And once you've got a reason to use a laboratory, what sort of laboratory can you use? Do you need to try and get 1000 students through a lab at a minimum cost and time, or are you trying to meet some sort of accreditation requirements – what constraints are there?
>
> Once you know what the laboratory is for – or possibly what you *think* the laboratory is for, you need to design it. You can go and have a look in the literature, or your own experience with labs, or some vendor selling pre-built kits, but unfortunately you probably aren't going to find the perfect answers talking about your class with your students there. You'll need to think about what is best, and a lot of the time take an educated guess. Remember to write down what you decided and why, because you might need to change things later.
>
> Once you have a lab designed and built, you need to run it. You can pilot it if the class sizes or risks involved are high, but you need to get it in front of real students doing real work to really understand it. Watch what the students are doing, use some research methods/instruments like a survey, an interview, an in-person observation, etc., to see if your guesses were correct, and most importantly whether there is evidence that the purpose(s) you said the lab was for, are actually being met. Are they actually learning that topic? Are they happy with the lab like they needed to be? If you see something you can improve at this stage you can, you don't need to wait for the end of an iteration.
>
> So now the students have gone through the lab, you have some observations recorded, and you know a little bit more about what is going on. Now you can have another look at the purpose(s) and your design. Were any of your assumptions wrong? Were there purposes or constraints you hadn't thought of? Make any changes you need to make, head off and read any supporting literature you might need, and get ready to do it all again.
>
> Eventually you will want to tell people about your lab design and what you've found. You need to be able to tell them what the purpose(s) of the laboratory were, what the design was, why the decisions in that design were made, what changes were made, and why you made those changes. The 'why's come from the literature, your expert analysis, and your observations.

There is also a single line, 'if you remember nothing else, remember this':

> Designing, refining, evaluating, and reporting on laboratories in engineering education must all explicitly derive from the purpose(s) to which that laboratory will be applied.

7.1.4.1 Purpose-First as a Theoretical Framework & Methodology

The Purpose-First Approach follows the ontological and epistemological paradigm of post-positivism. This paradigm is approximately a combination of positivism (largely associated with pure quantitative experimentation) and interpretivism (largely associated with qualitative research), and holds that there is simultaneously an objective truth that may be investigated and knowledge may be gained, but that there are inherent limitations to any understanding of that truth and so subjective interpretation is an inevitable part of any claim to knowledge [205]. It is most closely associated with mixed methods research.

The Purpose-First Approach is an *opinionated* approach. It embeds this post-positivism in its fundamental nature, and mandates certain research behaviours. For example, it explicitly rejects the attempted use of purely quantitative experimental designs such as RCTs. This does not mean it rejects the use of quantitative data collection or analysis, but it holds that this collection should not impose constraints on the teaching and learning activity which are not part of the design, such as needing random assignment, control groups, etc. The analysis of this quantitative data does not give perfect knowledge of an objective truth, but rather yields a subjective and incomplete perspective on that truth based on the context of the laboratory within which it was observed – and generalising that perspective requires recontextualisation for any new context.

Another definitions and semantics issue exists with constructivism – this exists partly in competition with post-positivism as a lens on the establishment of social knowledge, related to interpretivism. The Purpose-First Approach embeds constructivism via Constructive Alignment, but does so as a pedagogical theory rather than a research philosophy.

The Purpose-First Approach is also opinionated in terms of methodology, but runs into the previously observed issue surrounding the difference in nature regarding e.g. Action Research and Ethnography, which are both ‘methodologies’ but which exist at entirely different levels in the research philosophy and execution hierarchy. The ‘process’ level which contains practical inquiry processes such as Action Research is mandated by the Purpose-First Approach, explicitly embedding Educational Design Research into its application. The level of methodology containing ethnography, phenomenology, etc. is left agnostic, and should be selected based on the requirements set out by the laboratory’s purpose(s) – similarly the approach is agnostic to specific methods and instruments.

As a result of this structure, there is significant scope for alternative theoretical frameworks to be applied in parallel with a Purpose-First Approach research study. So long as the process is inviolate and there is no contortion to the laboratory that is unrelated to the purpose, some other ontology, epistemology, and methodology can be intersected as ‘the way the laboratory is observed’, with these observations out of this self-contained research study being fed into the EDR process. Thus, if it was determined that students’ experience of some phenomena in the laboratory was of interest, a phenomenographic interview study could be run with the students in the laboratory to inform the later iterations – though it is anticipated that the methods of observation selected would more commonly be informal.

Of course, research studies conducted entirely outside the laboratory context can be incorporated by way of literature, even experimental designs. They cannot be assumed, however – if incorporated into a design their effect should be a part of the observation and reflection process.

7.1.5 The PIPOV Framework

A study may have existed where the Purpose-First Approach, in isolation and without supporting artifacts, may have been introduced to practitioners directly, and how they made use of this approach studied. This thesis did not take that approach, largely due to the observation that there were established methodologies in other disciplines in education research that already leveraged this, resolving the issues this thesis is exploring, and yet these were (and remain) exceptionally rare in laboratory research in engineering education.

A different study may have existed where these approaches could be introduced as they are in these other disciplines – exactly as the author has encountered them, with the body of knowledge from research publications and supporting explanatory or competency development material. This thesis did not take this approach either, based on the observation that the natures of these traditions are already incredibly well-established in engineering as a discipline – to the point where one might make the claim that these design approaches have originally been derived from engineering themselves. In addition to the permeation of constructive alignment into the engineering education academic psyche it is concluded that engineering education practitioners in the aggregate already know about design, and already know about the critical nature of purposes when considering education (at least those purposes limited to intended learning outcomes).

The above material for the Purpose-First Approach is dense, abstract, philosophical, and only the tip of the iceberg in terms of digging into the mechanics of Educational Design Research and the post-positivist ontology and epistemology described in the associated literature. It is plausible that this approach is valid and would be of benefit to the practice of laboratory design and evaluation, but it is less plausible that practitioners would adopt it based on the description and sources. To aid in adoption, the PIPOV Framework was developed, serving as a framework for understanding and communication, and a vehicle for peripheral or casual engagement – allowing practitioners to 'dabble' with the concepts in order to see some benefit, before diving into the details of the Approach.

7.1.5.1 The Visual Framework

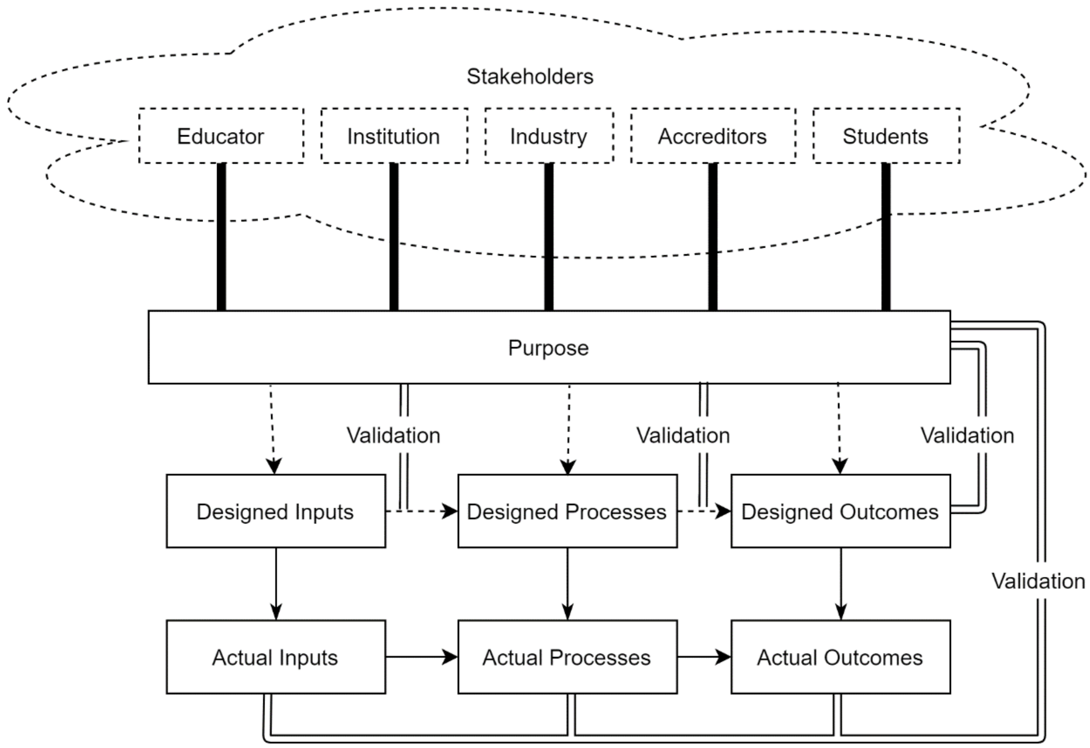


Figure 15: The final version of the PIPOV Framework Visualisation

The Purpose-First Approach and the associated source literature and epistemology were identified in the early stages of this thesis work – the definitions and prose description underwent some refinement, primarily being delivered verbally through conversation rather than defined in writing. The PIPOV Framework on the other hand saw significant revision and refinement over the empirical studies run in support of this development. The details of this design process are documented through the empirical studies chapters and particularly section 7.2.2.3, but the final version for this thesis is presented as Figure 15.

The PIPOV Framework describes the elements to consider when designing and/or evaluating a laboratory in engineering education. It presents an iterative view of the design, showing how these elements interact with each other, both in terms of how they 'do' interact, and how they 'should' interact. The nuance of the framework does have meaning, particularly the types of lines used – this is intended to trigger points of curiosity, and motivate practitioners to dig deeper once a surface level understanding of the framework has been established.

7.1.5.2 Supporting Material

Along with developing the visualisation for the framework, supporting material was developed to fill in the different elements with the different attributes that may be varied in a laboratory design, descriptions of what different relationships mean in detail, and validation techniques for specific points of interest. This supporting material was eventually discarded and not used directly when presenting the framework to practitioners, however this was in the context of the participants being able to directly have a conversation with the author about the framework and obtain the same information. The supporting material would instead be required for adoption of the framework without the author's direct involvement, much like the supporting material provided for frameworks as Bloom's Taxonomy, supported by [156], or Constructive Alignment, supported by [152].

This supporting material does not present an exhaustive taxonomy. It is informed by literature which purports to present an all-encompassing view of this aspect or that, but it is the experience of the author that these are limited in perspective and usually tied to an aspect of interest of the work being done – for example the differences between hands-on, remote, and simulated laboratories in [37]. The design space for laboratories (and as will later be shown what laboratories are at all) is sufficiently broad and complex that a complete list of all the ways they might vary is impossible. Nevertheless, what is *already* known to vary is a good reference point from which to think about new designs.

7.1.5.2.1 Purpose

The purpose of the laboratory is the anchor of the design and evaluation. The purpose is the core to the answer of 'is this laboratory good/working' – if and only if the purpose of the laboratory is being fulfilled, the laboratory is good/working. Improving the laboratory necessarily means meeting the purpose more fully. The purpose(s) of a laboratory include the constraints on that laboratory – space, time, cost, effort etc. are intertwined with the desired impacts from the laboratory and similarly motivate the design, and specify whether it is working or not and to what extent.

The best way this thesis work has found to extract and discuss purposes has been to reframe these as "motivations for change". A laboratory purpose is something that, were the practitioner to observe evidence that it was not being achieved, would motivate a change in the design of that laboratory. If students were performing poorly on an assessment the laboratory should have been supporting them in, this is something that would motivate some design change. If the laboratory budget blew out, this is similarly something that would motivate some design change.

There are several categories of laboratory purposes that have been identified in the work of this thesis, though this is likely to be an incomplete list:

- Stated learning objectives: Likely due to the increasing adoption of Constructive Alignment, these are the 'default' purposes of a laboratory often stated in handbooks and linked to accreditation frameworks. Bloom's Taxonomy, SOLO, etc. are examples of these.
- Other student learning: There are often purposes held similarly to the above which are not stated as these broader learning objectives, such as students learning how to use specific pieces of equipment that other aspects of teaching might later rely on, knowledge of spaces, etc. Often these are more difficult to pin down outcomes that cannot easily be assessed for marks.
- Student affective outcomes: This includes changes in the students not typically captured as 'learning' – whether they enjoyed the laboratory or not, whether they were motivated or engaged and how that carries through into other work of a class or beyond, etc.
- Project-style constraints: These are things like time to develop and run, cost, available effort, real estate required, maintaining safety, student timetabling issues etc. which combine to place a hold on what is possible in a laboratory. If a purpose was to teach some outcome, one-on-one mentoring for each student might be the 'best' way to achieve this, but might not be viable – these purposes capture these.
- External or ephemeral outcomes: This is a loose category including outcomes such as the institution wanting an image as a 'practical' university with pictures of laboratory use on its website, or wanting to be able to show industry visitors a group of students in a laboratory room rather than a lecture theatre.

7.1.5.2.2 Stakeholders

The stakeholders in a laboratory are the sources of the laboratory's purpose. In essence, they are those people or organisations who are impacted by the purposes being met or not. Identifying the laboratory stakeholders is important in capturing all of the laboratory's purposes.

The visualised list of stakeholders includes:

- The educator
- The institution
- Industry
- Accreditors
- Students

There are other groups that may have an interest, such as government, overall society, etc. – however these are largely captured in the specifically called-out group, with government and society's needs in theory captured by accreditation and the requirements from industry.

7.1.5.2.3 Inputs

The inputs to a laboratory are the combination of controllable and uncontrollable aspects of the laboratory which together form the laboratory design. This includes all of the teaching material, the equipment, the space, and the educators – but it also includes the students, the institutional and social context, and any number of other variables. The intense, uncontrollable complexity here is the primary reason quantitative experimental studies fail to achieve success in laboratory evaluation.

The first step in considering inputs is to examine what is explicitly chosen. The topic or phenomena the students will be investigating. the laboratory equipment, the instructions on what to do, whether the students are told to work in groups, what associated assessment are they asked to do, whether it is a remote or simulated lab, etc.

Next are inputs that are known to at least some extent but not explicitly chosen. Where in a degree some laboratory occurs, the type of students that are enrolled, their prior knowledge, the institutional culture around laboratories, etc.

Finally are the inputs that are unknown, but which may have an influence. It should be assumed that there will be inputs in here that are ‘unknown unknowns’, having an influence without having been identified during the design process – part of validation can involve discovering these and adding them to the design considerations. It may be possible to observe unknown inputs to some extent once they are identified. An example of one of these inputs could be the influence of other classes being studied in parallel – an educator giving one model for a phenomenon and another educator teaching an alternative model at the same time could cause problems in the laboratory.

Some lists of inputs are described in the above literature review section 2.1.3 laboratory attributes, however as stated these are largely incomplete, and focused on the hands-on/remote/simulated variance in terms of categorisation. Depending on the goals of the laboratory and any specific innovations involved designers will do better to examine the literature around the specific aspects of interest to generate the list of inputs most relevant to that laboratory. The context of the class the laboratory occurs within is often left out of these lists, and so starting from that foundation would also be useful.

7.1.5.2.3.1 Designed vs Actual

One of the fundamental aspects of the PIPOV Framework is the recognition that what the designer *thought* would be the case, is not necessarily what was *actually* the case. Some of these will almost certainly be the same across designed and actual and are trivial to confirm, e.g. did the laboratory happen in the room it was supposed to, were the students given the correct lab guide. However, some are highly variable. This can often be due to an input not being fully understood or easily observable, such as incoming students’ prior knowledge, or the necessary time to run a class of students through a particular set of tasks.

Designed inputs should lead to the designed processes – however this is not inherently guaranteed by reality, and is a requirement for the designer to fulfil. Actual inputs on the other hand do inherently lead to actual processes – what occurs in the laboratory is a direct result of the actual inputs that make up that laboratory, whether controlled or uncontrolled. Thus, the appropriate selection of inputs is necessary to ensure a quality laboratory design, and further the active investigation of what variance between designed and actual inputs occur is necessary for quality laboratory evaluation. In particular laboratory evaluation studies reporting that ‘the laboratory involved xyz’ without demonstrating evidence that xyz was in fact the case are potentially limited.

7.1.5.2.4 Processes

Processes are those things that occur within the scope of a laboratory teaching and learning activity. This is the aspect of the PIPOV Framework that most differs from the source material, e.g. [21], [37]. Rather than being a black box, the laboratory processes are critical to dive into, theorise about, and observe in order to understand what the laboratory is doing and how effective the design is likely to be. Processes is split into 4 categories along 2 axes. One is the designed vs actual split also present in inputs and outcomes, and one is a distinction between physical processes and cognitive processes. Physical processes are defined as an input in the source material – and in many ways this is correct, physical processes are indeed the input to cognitive processes. But they are an intermediate step:

inputs as they are termed in PIPOV lead to physical processes occurring which lead to cognitive processes occurring. In order to reduce the complexity of the model, and in response to participant reactions to previous versions which explicitly called out physical and cognitive processes, these are included together.

7.1.5.2.4.1 Physical

Physical processes are the things that are said, manipulated, clicked, written, or otherwise done during or associated with a laboratory. They might be paired with a corresponding input, e.g. a laboratory guide telling students to work in groups is an input, those students working in groups is a physical process. They might also be emergent, occurring as a result of the inputs (as with all processes regardless of intention), but without being specifically linked to some instruction.

As previously mentioned, physical processes lead to cognitive processes. However, they can also lead to outcomes, where those outcomes are not cognitive in nature – for example, running costs.

7.1.5.2.4.2 Cognitive

Cognitive processes are those processes which occur in the mind of the student and can only be observed by proxy rather than something that might be directly observed in reality. This is where the constructivism pedagogic theory from Piaget/Vygotsky via Constructive alignment applies to aid in understanding these processes. It should be noted that this is not limited to the cognitive domain of Bloom's taxonomy – all learning is covered under constructivism, including affective and psychomotor outcomes.

It is entirely supported within the PIPOV Framework to apply other compatible learning theories to the cognitive processes aspect of a laboratory. For example, Papert's Constructionism [206] which derives from constructivism, is common in the author's background in software education and could be applied in this framework element as well, as could behaviourism, cognitivism, critical theory, etc. However, this was not generally observed during the empirical work with the framework with practitioners – the embodiment of Constructive Alignment with mandated learning objectives appeared to predispose discussions towards constructivist approaches to pedagogy.

7.1.5.2.4.3 Designed vs Actual

The difference between designed and actual processes has been one of the most impactful aspects of the PIPOV Framework and the Purpose-First Approach generally. Rather than observing the outcomes of a laboratory and treating the activity as a black box, the processes should be investigated to see where variation from the designed laboratory might occur.

This is different for the two categories of processes. Physical processes can be observed directly and reacted to. If, for example, a laboratory instruction is not being followed, or expected group behaviours are not occurring, this can be observed and included in the reflections on the laboratory design. Do the students need the instructions explained differently? More demonstrator support? Do the students require further education in team skills? Before these decisions can be made however, the impact to the purpose must be considered – it may be that the group behaviour not matching expectations has no tangible impact on the desired learning outcomes and so can be left alone.

Cognitive processes are significantly more difficult, as they cannot be observed directly and must be investigated by proxy. The designed component involves a reflection on the pedagogic theory being applied to the laboratory (Educational Design Research's local instruction theory), to determine what is expected to be happening in a student's mind when they undertake the laboratory. What conflicts are they likely to experience between their understanding of theory and their own observations of reality? How might they reflect on their experiences to incorporate new knowledge?

Identifying variance is the critical challenge, and the goal of all educational assessment – how does one know what a student has learned? This can be observed as part of an outcome, e.g. whether they could answer a related exam question effectively, but it can also be observed in the other direction via behavioural indicators, as was done in [76]. Education research has indicated behaviours that benefit learning – and so identifying those behaviours that can be directly observed can be used as a proxy for a laboratory improving cognitive processes. Consider a laboratory design where low exploratory behaviour is observed where inquiry skills were a purpose – a change in students understanding of inquiry skills might be impossible to observe, but a change in the amount of exploratory vs didactic behaviour might not be.

7.1.5.2.5 Outcomes

In earlier iterations of the framework the 'Purpose' element could be described as the 'Designed Outcomes', where Outcomes are 'Actual Outcomes'. In the revised version the Purpose are those outcomes which motivate change in the laboratory if they are not met. The difference between outcomes and the purpose is both that:

- Outcomes can be specific outcomes that contribute to larger-scale purposes, e.g. a purpose might be increased student learning about laboratory equipment, and an outcome might be students gaining experience with oscilloscopes.
- Outcomes can include things that resulted from the laboratory that were not part of the laboratory purpose. This might be the result of a difference between designed and actual outcomes and be a concern motivating change (i.e. a purpose was not met), but it may also be due to things that the laboratory causes, which would not be a motivator for change if they were not met. This is relevant in later validation of the purpose. These outcomes can still be important to observe, as their changing indicates some change in the inputs and processes that used to yield them.

7.1.5.2.5.1 Designed vs Actual

The difference between designed and actual outcomes is where the majority of evaluation research on laboratories in engineering education is focused. An intended improvement in some learning objective, or an increased affective response, etc. are the usual drivers for laboratory interventions. The Purpose-First Approach and PIPOV Framework does not preclude that part of evaluation – it adds to it to provide the appropriate context and measure things that otherwise could not be.

7.1.5.2.6 Validation

Validation is the part of PIPOV that turns a design architecture into an iterative evaluation framework. Having identified a design with a purpose, inputs, processes, and outcomes, it is then the task of the designer to observe whether their design has worked, and change it to improve any misalignments afterwards. All observations, instruments, and other data collection occurring for the purpose of evaluating the design are part of validation. Not all validation is active observation of a laboratory in authentic use however – a significant portion occurs as a process of reflection by the designer during the design.

There is implicit validation required between designed and actual elements in the design – finding out whether the inputs, processes, and outputs happened in reality is necessary to determining whether the design is working – these relationships are described above. Some aspects of validation are called out explicitly in the diagram, as they relate to critical potential points of variance in a laboratory design.

7.1.5.2.6.1 Validation of Designed Inputs -> Designed Processes; Designed Processes -> Designed Outcomes; Designed Outcomes -> Purpose

The steps between the designed inputs, processes, and outputs must be reflected upon as part of the design process to consider:

- What evidence is there that the given inputs will lead to those processes, and those processes to lead to those outcomes – noting that prior experience and expert judgement are both valid sources of evidence.
- To what extent is the designer likely to be able to observe and know the states of these elements in the actual laboratory?
- Given the potential for variance, what impacts on that transition might occur if, e.g., some intended group process does not occur?
- Based on the above, what observations should be in place to check on elements identified as being risks to the fulfilment of the laboratory's purpose if they do not actually happen?

These reflections must all be conducted with the purposes of the laboratory at the forefront of decision-making – for example, if an outcome might vary but that variance would still achieve the purposes, then this is not something that needs controlling.

7.1.5.2.6.2 Validation of the Actual laboratory against the Purpose

In contrast to the previous design reflection validation, this validation step requires active observation with empirical data. Acknowledging the incomplete knowledge which characterises these processes, it is insufficient to empirically validate only the alignment between designed and actual at each step – the actual laboratory must be validated against the purpose itself in order to be properly evaluated, using whichever methods or component methodologies have been identified in the literature as able to validate specific claims, e.g. classroom observation, surveying, a contained quant/qual/mixed methods study.

Notably the validation with the purpose is *bi-directional*. The articulated purpose is not considered to be absolute objective truth, because its derivation is a human process. There may be purposes that are missed off this process that are identified when the actual outcomes are observed – or purposes that when their achievement in reality is observed are then rejected. This is a process that should of course involve the stakeholders of the laboratory to iteratively refine the purposes along with the laboratory that attempts to meet them.

7.1.5.3 How to Apply the Framework

The empirical studies have provided significant insight into how the framework can be applied in different ways. There are 3 ways that have been identified as effective for a given context, though there are likely to be many more.

7.1.5.3.1 As a Formal Structure

The originally intended use of the framework was to follow the sequence of elements and use these to design the laboratory:

- Identify stakeholders.
- Articulate the purpose(s) of the laboratory in consultation with stakeholders as needed.

- Determine the outcomes that should be met for the laboratory to be considered successful at those purposes.
- Speculate on the necessary cognitive and physical processes necessary to achieve those outcomes.
- Create an initial laboratory design which should encourage those processes occurring.
- Run through the validation steps as an expert review.
- Based on the results of this review adjust the design as needed.
- Determine observations and instruments necessary to conduct the validation steps in a live run.
- Implement and run the laboratory & observations for the first time (or pilot it).
- Reflect on the observations made, feed this back into the design, and continue iterating.
- Report on the design decisions, contributing literature, observations made, and analysis.

This ‘fully-engaged’ version of the framework embeds the Educational Design Research methodology in the design, evaluation, and reporting of the laboratory and thus should carry the associated benefits identified in the literature.

7.1.5.3.2 As an Input to Other Design Material

Some practitioners had their own perspectives on the design or institutional design documents governing their laboratory design processes. This included the new laboratory validation design, as well as being raised in one of the expert review sessions. One key variation was the overall perspective on the ‘laboratory’. The PIPOV Frameworks centres on the teaching and learning activity, with the equipment being a contributor for that. The alternative perspective encountered was with the equipment at the centre, and teaching and learning activities which could make effective use of it being the derived aspect rather than the other way around.

On its face this appears to go against the EDR and PIPOV process, however when interpreting the use of the equipment as a motivator (e.g. to ensure cost-effectiveness of the equipment involved), this should be considered a purpose for the laboratory – these are not just cognitive learning outcomes.

The elements involved in PIPOV and in particular its iterative nature is retained, but the order of the questions must necessarily change to reflect the central equipment as an anchor.

7.1.5.3.3 As an Input to Reflection

Rather than going in on a full use of the framework to design or redesign an entire laboratory, the framework can be an input to design discussions reflecting on the elements it contains. The critical aspects that must be retained in order for the use of the framework to retain its benefits are:

- The design aspects it motivates must be tied to the purpose of the laboratory.
- The validation measures undertaken to investigate the changes must be iterative, contextual, and not require contortion of the laboratory to run.

This was experienced during the existing laboratory modification field test, with individual aspects of interest identified through ‘thinking through’ the laboratory using the PIPOV Framework as a lens.

7.1.6 Proposed Benefits

The application of Educational Design Research to the design and evaluation of laboratories in engineering education is speculated to resolve the issues observed in the earlier literature review examining the standard practice of research in this space – inconclusive evidence, lack of generalisability or transferability of results, and the same questions asked decades earlier still being asked today seemingly without an overall progression of knowledge in the discipline, e.g. “.

Outside a broad statement that ‘design and evaluation of laboratories will improve’, there are several proposed benefits to the use of the Purpose-First Approach and the PIPOV Framework. Some of these are of a scope to examine within this thesis and appear in the empirical work – some however are on the order of decades and are proposed based on indications from the literature synthesis.

7.1.6.1 Avoidance of Research Contortion

Research on laboratories in engineering education that follows a purely quantitative model introduces motivations for changing the design that are separate from, and most likely at odds with, the purposes of that laboratory. For example, splitting students up into random assignments, not reacting to observed issues due to the need to keep the experiment protocol valid, contorting laboratory designs based on matching some other laboratory for comparison rather than seeking the most effective design.

The Purpose-First Approach explicitly rejects these contortions as inherently damaging what is being studied – if you study e.g. a hands-on laboratory and a remote laboratory which has been designed to exactly match the hands-on laboratory with no use of the remote laboratory affordances (e.g. making a remote laboratory only available during class times as in [175], [207]), not only are students undertaking a laboratory which is most likely not as effective as it could be, but the value of the research is questionable. The study is not a study comparing hands-on and remote laboratories, but rather one comparing one hands-on laboratory with one version of a remote which would normally never have been designed like that.

7.1.6.2 Transferability to New Contexts

As previously observed in the literature review, purely quantitative, experimental research designs examining laboratories in engineering education are not able to fulfil the assumptions required for their mathematical conclusions to be valid. Further, the lack of reporting of context means that practitioners in other contexts may struggle to make effective use of these research outputs in their own designs. The Purpose-First Approach provides a research methodology which does not carry these same issues, particularly through the reporting of design decisions and their motivations rather than relying solely on summative data; and the PIPOV Framework provides for a reporting structure which will capture the available context the evaluation occurs within.

7.1.6.3 Support of Literature Synthesis & Qualitative Meta-analysis

Directly related to the above, extensive use of the Purpose-First Approach and PIPOV Framework would support literature synthesis of a more effective type than the collection, coding, and counting that is prevalent in the reviews discussed in the literature review.

Consider collecting 10 studies on laboratory design and evaluation processes, selected because they each encountered a particular type of design motivation, e.g. ‘students passively engaged with the laboratory following the instructions without reflecting on the natural laws being investigated and showing poor learning outcomes’, with 10 different changes to the laboratory attempting to correct for this issue as part of their design process, and 10 different observations of whether this was effective.

In the case that these were quantitative experimental designs, reported as is observed in the literature review with limited contextual details, sense-making around the collective knowledge they contained would be extremely limited. Quantitative meta-analysis is not likely to be possible, as the confounding factor of different laboratories could not be controlled for even if the data measured met a standard which could be mathematically aligned.

Instead, these 10 essentially case studies could be combined to support decision making. First, by allowing a reader to identify which of the 10 laboratories best fit with their own context or not. Then identifying the changes and observations made and considering whether any variations in context would make the same outcome more or less likely. Having selected a candidate intervention, this then feeds into the reader's own design iterations, they observe the results, and potentially iterate again if the findings differ. This identification of context is already necessary with quantitative studies given the non-fungibility of laboratory contexts (different institutions, purposes, instructors, students, etc.) – only without appropriate reporting of that context and with little ability to look inside the quantitative experiments.

7.1.6.4 White-Boxing

On this theme of looking inside, white-boxing the laboratories rather than treating them as black-box experiments that can only be observed externally opens up significant benefits to laboratory evaluation when using the Purpose-First Approach and PIPOV Framework. As was encountered in much of the empirical work detailed below, 'mid-way' observation and intervention was a recurring theme for laboratory evaluation, primarily leveraging learning analytics.

7.1.6.5 Framework Communication

As observed in the literature review section 2.6, frameworks aid in effective communication through a shared understanding. Reporting on a laboratory that was described in alignment with the PIPOV Framework would be easier to understand for a reader also familiar with that Framework. This is an observable benefit to e.g. Bloom's Taxonomy, and Constructive Alignment.

7.2 Laboratory Scoping and Artifact Ideation

7.2.1 What is a Laboratory Study and Derived Definition

A critique of the foundational literature of this thesis yields a surprising observation – nowhere is the term 'laboratory' defined. There is work on the educational purposes to which they might be employed, there is an examination of new types motivated by innovative technologies, there are comparisons to laboratories between disciplines. But nowhere is there any statement or description that would enable someone to state 'this is a laboratory' or 'this is not a laboratory'.

This is not an oversight in the literature review, but rather a condition of the existing literature in common with the issue of existing laboratories not being evaluated or researched. By far the body of work on laboratories in engineering education and elsewhere lies in investigating *new* types of laboratories, a symptom perhaps of the overall research culture in higher education more generally.

This poses a significant challenge, however, when attempting to address discipline-wide questions about laboratories such as is posed in this thesis. Given a proposal, tool, conclusion – to what does it or can it be applied? What scope might be needed to state that it answers questions 'about laboratories' rather than merely 'about those laboratories'? Resolving this required a definition to be determined, and no definition was forthcoming from the literature. An empirical study was proposed to investigate the definition of 'what is a laboratory in engineering education' with engineering education practitioners.

The theoretical framework used to frame this study was a synthesised combination of attributes described in the earlier literature review – in particular drawing on Feisel & Rosa [6], Nickerson et al [37] and Brinson [21] to identify what sorts of things might be attributed to a teaching and learning activity such that it might be called a laboratory – intended learning outcomes, activities, equipment, location, etc.

The design of this study underwent significant redesign work due to the impacts of the COVID-19 pandemic. The initial design called for a series of face-to-face semi-structured interviews with engineering education practitioners where exemplar laboratory candidates could be presented, and determined as being considered a laboratory or not, as well as discussing why this was the case. Due to the inability to meet face to face, this design was changed to incorporate a combination of remote video conference interviews coupled with an online web platform for presenting the laboratories in a survey format. The web platform could also be deployed independently for participants not meeting synchronously.

Ethical review of the study design was approved under protocol number 2019/1040 by the University of Sydney.

The candidate laboratories were developed to highlight teaching and learning activities with differences across various identified attributes. These attributes and in particular their combination could be in no way exhaustive – the combinatorial nature of even the most basic 10 binary choices would yield over 1000 examples to review, necessitating a qualitative investigation into the discussions about these examples rather than attempting a quantitative analysis.

The examined attributes were:

- Data source (real or simulated [and known or unknown])
- Proximity to equipment (local or remote)
- Interface to equipment (computer or physical)
- Group differences in experience
- Data timing (synchronous vs asynchronous)
- Data/interface control (student-controlled vs student-requested vs provided)
- Activity type – experimental vs purely observational
- Experiment type – didactic vs inquiry vs research
- Activity context – part of coursework theory or a standalone project
- Degrees of freedom – from zero (purely observational) to many

Again, this leaves off many – for example, a proximal haptic interface on remote or simulated equipment, VR interfaces, interactions with industry placements, etc.

The examples were presented as a combination of written description with accompanying image, as shown in Figure 16.

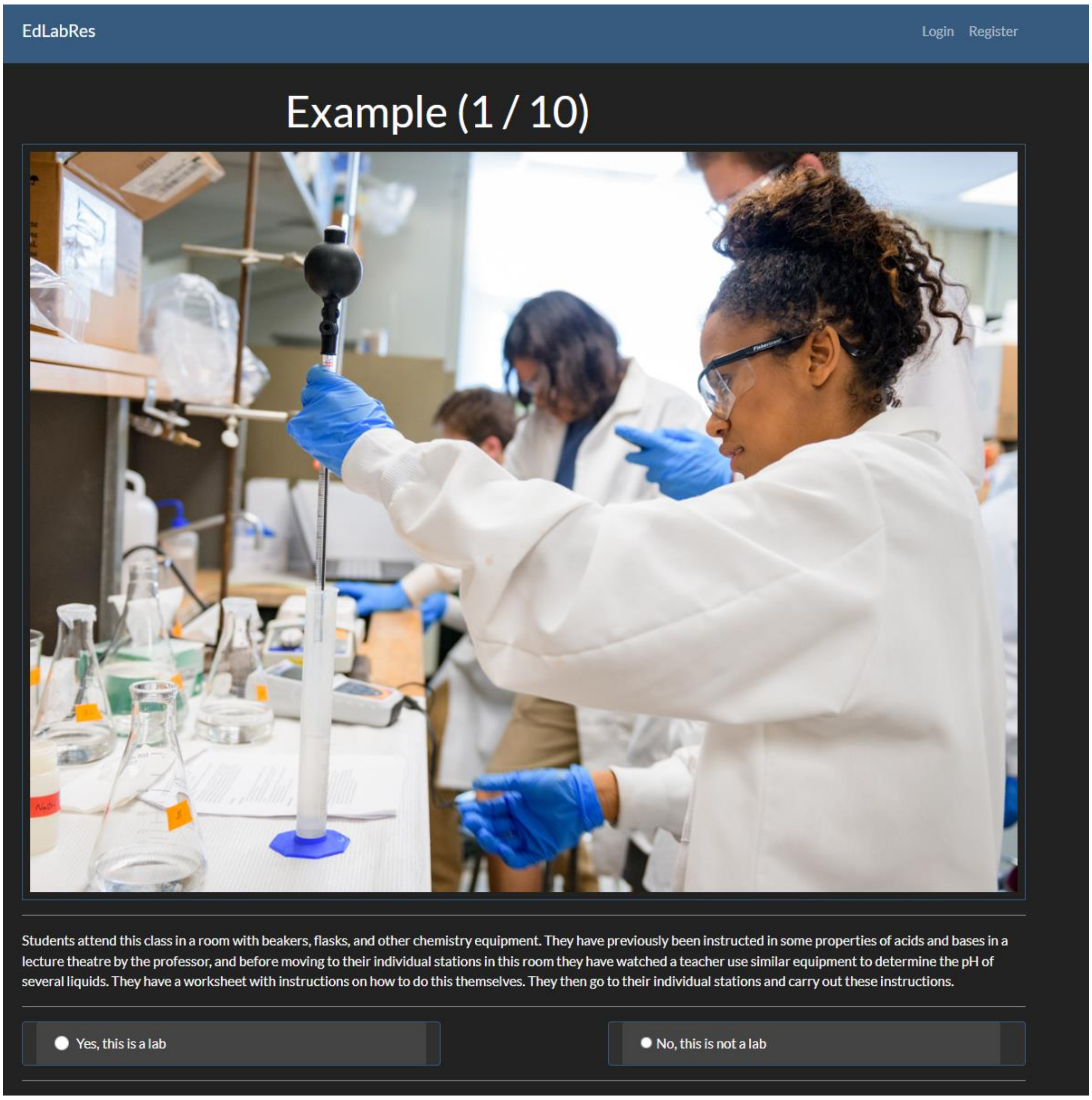


Figure 16: An example candidate laboratory used in the What Is a Lab survey. Image credit Evan Krape, University of Delaware

Of course, this example is intended as a baseline, meeting all of the most stereotypical attributes of a traditional laboratory to give context to the variations that follow. The following is text for an example varying in the proximity to equipment, e.g. a remote laboratory:

> Students begin with instruction in deflection in civil structures in a lecture theatre by the professor. Students then attend this class in a room with computers, which are connected to the internet - there is a website they access that allows them to remotely control robotic equipment that adds real weights to a full-sized bridge, and displays the deflection this causes. They work individually by using the website to add weights on the bridge, and recording the deflection of various members/components of that bridge. They then analyse this data to determine attributes of the material used in the bridge's construction.

And text for an example seeking the least likely to be determined to be a laboratory:

> Students attend a lecture theatre where the professor delivers a lecture on electrical circuitry. During this lecture the professor has various piece of equipment such as a breadboard and oscilloscope that he passes out for student inspection during the lecture.

The design of the survey included an initial demographic and baseline definition set of questions, followed by 10x basic examples, with a follow-up set of 10x examples made available if requested by the participant, and finishing with a set of debriefing questions, listed in full in Appendix E.

Recruitment was conducted through direct email contact with identified engineering education practitioners. These practitioners were drawn from identifying engineering educators in Australian universities who had published literature in journals or conference proceedings describing the development or evaluation of laboratories in engineering education – laboratory being defined by their own use of the word ‘laboratory’, ‘laboratories’, ‘lab’, ‘labs’ in a database searchable field. Note that the limitation to practitioners in Australia was due to the original design which called for physically meeting these practitioners to interview.

From this direct email recruitment N = 9 practitioners were identified and agreed to participate. The participants were broadly balanced against many of the demographic questions – with a mix of English and other languages as a first language, a combination of academic and industry backgrounds as well as experience with international institutions, and with some having been solely a lecture with others having performed multiple roles. In all cases the interviewer/facilitator was the author, and participants attended sessions tentatively scheduled for 60 minutes. In practice sessions went significantly longer, with the majority running between 90-120 minutes and 1 session running for approximately 3 hours. With one exception involving participation in other empirical studies conducted subsequently, participants attended a single session only. Due to the immediately re-identifiable nature of their recruitment and conversations given the publications their laboratories appear in list them as the author, their laboratory descriptions, and identifying details in their statements and about their demographics are withheld or aggregated.

Likely as a result of recruiting based on already published research, years of work was weighted towards the higher ends, with no participants less than 4 years and most having higher than 7. While it was not a question captured in the demographic questions, as observed during the interviews all participants were male-presenting. Academic level was also identified for these participants through their public listings, with practitioners a mix of level C, D, and E, with some participants in school and faculty leadership roles.

The remote video conference interviews were conducted as dual-purpose sessions involving the discussion on this study, as well as a discussion involving the early designs for the PIPOV framework. These discussions were often fluid after an initial structured portion of the session, and moved back and forth between the presented PIPOV framework diagram and the questions around ‘what is a lab’. The analysis of the PIPOV diagram discussions will be presented separately below.

The interview protocol described the following phases:

- Introductions: greetings, context information and consent, a short discussion of the identified paper that led to their recruitment and the laboratory it describes
    - In particular seeking motivations behind the study, and motivations for any changes to the laboratories being studied.
- Stepping through the laboratory web survey with examples
- Switching over to a discussion introducing the PIPOV framework

Where the example questions followed a sequence, shown in Figure 17.

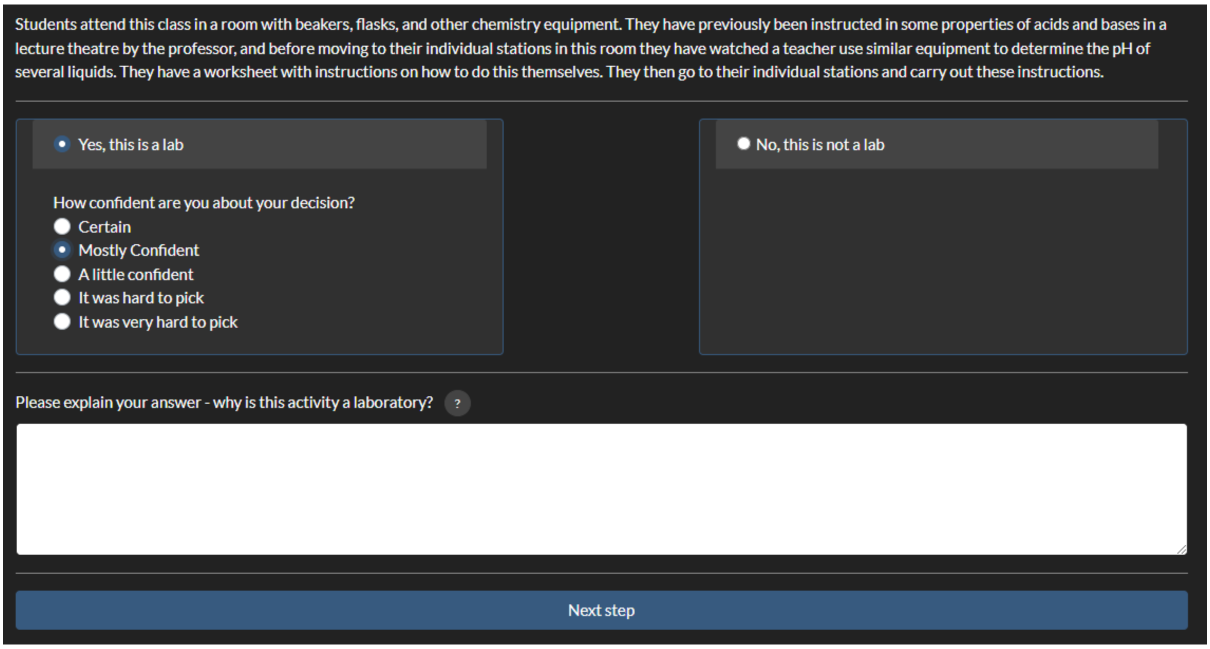


Figure 17: Expanded questions in the What Is a Lab survey

Where each part to the question became visible upon selecting the component above. In practice however, the interviews and surveys were near-immediately and consistently derailed through all participants.

The survey design was piloted using interested academics and associates at the University of Sydney. These pilot participants found the survey to be clear and were able to complete it without significant guidance – some clarifying descriptions were added to the '?' icons variously present in the survey. When run in the wild however, each of the interview/survey participants showed very different behaviours to the pilot. Where the pilot participants simply answered the questions, the real participants began to discuss the underlying theories and concepts that the survey was targeting in the first place.

They very quickly identified some of the attributes of variation under study after seeing a few examples, and began discussing their own lists of attributes, expressing a lack of certainty as to what the qualifying components of a laboratory were, and critically were changing their own definition as they considered the questions.

The initial question in the demographic page 'How would you answer the question "What is an educational laboratory?"' yielded surface level statements, e.g.

> "a laboratory is a class where students investigate natural phenomena"

> "practical work with a guide for some experiment to conduct"

However, upon considering the examples participants quickly began to clarify and expand their position verbally. Some statements typifying this were:

> "Yeah, I suppose it can be a lab if they are on computers. I mean, we call some things programming labs and they aren't really engineering labs, but if they were connecting to equipment then that's a real lab too"

> [in response to a verbal question comparing the student group division of labour vs the lecturer-student division] "I've never really thought about it that way, but yes, that student

is doing exactly the same thing in both classes. I don't know whether that means the lab isn't a lab or if the lecture is a lab"

"I don't think it can be a lab if there isn't any real data. I don't know. Maybe? It feels like there's a difference when it's a computer simulation."

It became clear during this process that attempts to pin down decisions on specific attributes that might vary and gaining any clear border between laboratories and not laboratories was unlikely to be successful – either the knowledge in the discipline is not yet established and/or mature enough to have clarity around where that border lies, or the border is itself ephemeral and subject to change depending on contexts.

The web survey was promoted digitally to a wider audience and saw several initial participants; however, these participants did not complete the survey. It is considered the observed issues with the live video conference participants are the likely cause.

The working theory that was developed in support of this thesis work was one of dimensionality and proximity. Consider a 3-dimensional space which captures all possible variations of laboratories in terms of their proximity to the equipment (remote vs proximal), source of the data (real vs simulated), and synchronicity (run in batch mode or live). Further assuming that each lab might have a different level of each rather than simply being a binary choice, we might plot a series of laboratories against these 3 variables as in Figure 18 below (not real data).

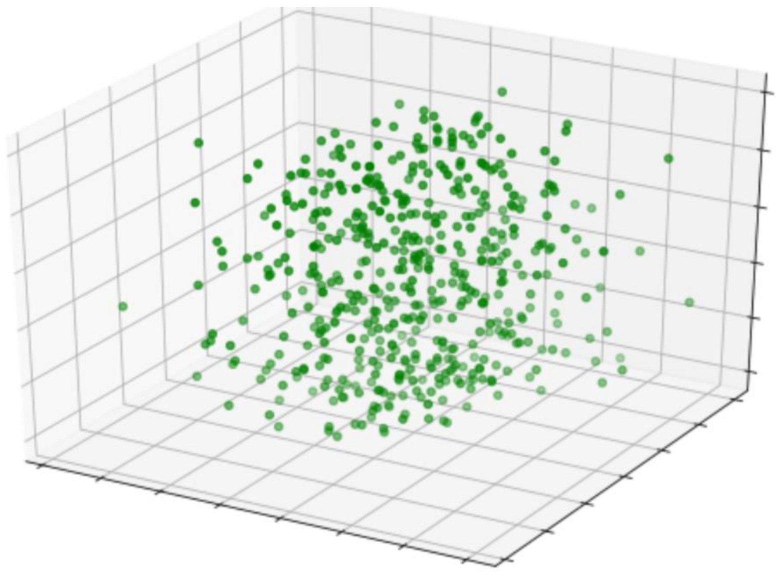


Figure 18: example 3d scatterplot, generated using matplotlib

The laboratories each have a distance to each other, and exist within spaces that might be defined in terms of boundaries for each contributing attribute, which was re-defined as 'dimension'. A 'remote laboratory' is thus any laboratory within a certain range of the proximal dimension, with the values of the other dimensions unrelated. A 'simulated batch' laboratory has constraints in both the data source and synchronicity dimensions, and so on.

How this relates to the definition of a laboratory overall requires expanding the number of dimensions to capture all possible variations in a teaching and learning activity to create a conceptual n-dimensional space within which each teaching and learning activity might be plotted – these dimensions being the attributes to vary listed above, though this is still not exhaustive and even the number of possible dimensions remains an unknown. The conceptual model however is still useful, because it provides a framework within which to discuss what a laboratory is or is not.

The definition of a laboratory, it is thus proposed, is not something which might be stated like a dictionary definition nor something which has specific absolute borders. Laboratories are teaching and learning activities which exist within a reasonable proximity to each other within this n-dimensional space which defines teaching and learning activities. Which dimensions carry the most weight in terms of the n-dimensional proximity, how proximal a laboratory must be in order to be a laboratory, whether there is any linearity or chaotic behaviour with interactions between dimensions are questions left unanswered as out of the scope of this thesis.

While not providing a satisfactory conclusion to the original question, this does however provide the basis for not a definition of laboratory to be reported in the literature for use by the discipline as a whole, but rather a working definition of laboratory such that the scope of this thesis may be clarified, obtained by considering the dimensions already identified and running through a thought experiment of where borders might be along those dimensions.

> A laboratory in engineering education is any teaching and learning activity or distinct sequence of activities in which students actively participate in a process for investigating natural or human phenomena which requires an engagement with some representation of reality external to the student.

- Being a teaching and learning activity or distinct sequence means there is a focus on specific activity(ies) rather than for example a student’s change in perspective over multiple years of a degree.
- Active participation precludes e.g. listening to a lecture, passively watching a demonstration without engagement, or reading a textbook without engagement.
- Natural or human phenomena captures hard sciences like physics laboratories, but also includes for example laboratories involving the study of human participants in a design workshop. The author does not generally distinguish these as not natural phenomena, but having observed this in some literature human phenomena is explicitly included for clarity.
- Investigating implies that there is some question to be asked. The nature of this question is undefined, but includes comparisons of theory to observed reality, deriving theories themselves, whether didactic cookbook style laboratories or inquiry/research style laboratories.
- Engagement with a representation external to the student differentiates between a student e.g. solving mathematical equations and comparing to an answer sheet, and a student using laboratory equipment, accessing a simulation, etc. There must be some input to the student which purports to be some representation of reality, whether this is real or a simulated model.

This definition can be considered the definition for using the Purpose-First Approach and PIPOV Framework – it is far broader than any definition yet encountered in the reviewed literature, but establishes the minimum requirements of a teaching and learning activity for these artifacts to apply.

7.2.2 Expert Review Design Process

As mentioned above, in parallel with the laboratory definition survey, an early exposure of engineering education practitioners to the Purpose First Approach / PIPOV Framework artifact was conducted as part of the Educational Design Research iterations. This contributes to both the

Analysis and Design phases of the EDR process employed. The practitioners were identified as subject matter experts in laboratory design and the current state of research practice as evidenced by their recruitment via having published literature in the field, and this process is considered to represent the ‘expert review’ process described in [109], or ‘expert appraisal’ from [92].

The context for the exposure for the same N = 9 participants was as follows:

- Participant has had an earlier discussion regarding their published laboratory work, including the motivations for the study itself, and any reasons for motivating change in the laboratory.
- Participant has then gone through the (heavily adapted) questionnaire on ‘what is a lab’.
- The interviewer (author) describes the nature of the Purpose-First Approach, discloses perceptions of the state of the literature on laboratory design and evaluation which motivated its development, and shows an early version of the PIPOV framework for discussion.

The backing analysis protocol for these expert sessions was to seek answers to the following questions, with different weight applied to each depending on the flow of conversation and the expert’s engagement with different aspects. This also includes the earlier discussion regarding their published work conducting laboratory design and evaluation. Note these are themes for discussion and were not necessarily asked verbatim or in sequence.

- What were the triggers for the research to be conducted on the laboratory? What variables were being observed, techniques being used? Was it an ad hoc trigger that emerged, or was there a formal evaluation structure?
- What research methodologies or techniques if any were considered for use in the study but decided against? Was any collaboration involved not captured by the author list?
- What are your observations regarding the Purpose-First Approach? Does it align with your thoughts on laboratories?
- Upon being shown the PIPOV framework as a diagram – before I explain this in detail, can you describe what this diagram indicates to you?
- Upon having the Purpose-First Approach, and contents and motivations for PIPOV explained – does this framework sound like something you could apply to laboratory design and evaluation? How might it fit into a study you could run?
- What aspects of the Purpose-First Approach and PIPOV framework do you like or support? Dislike?
- What changes would you recommend making to these to make them more effective, easier to understand, easier to use?

This component of the sessions progressed far closer to expected than the survey, with the participants actively engaging in the discussion and happy to give feedback and suggestions on the aspects under study.

7.2.2.1 Laboratory Studies

Discussions around the laboratory studies run by the participants were varied. Each of the studies was an implementation study of a new or changed laboratory. The studies reported the design (not design process) of the laboratory, as well as evaluation observations of varying scope, type, and method. None of the studies were longitudinal and referred to a single semester of teaching.

Discussions around the methodologies selected aligned with observations in the earlier literature review:

- Little explicit consideration of epistemology or methodology occurred.
- Following the precedent set by existing material in publication venues was often cited as the way in which methods were selected, as well as the ability for their collection without significant preparation time/effort – e.g. affective surveys, assessment results, and written student feedback.
- Lack of expertise in alternative methods was another reason for the selection of research methods. Time constraints inhibiting upskilling or collaborating were stated alongside this in a small number of cases.
- Concerns regarding the review of proposals by institutional ethics review panels and seeking the path of least resistance was also cited as a motivation, also related to time constraints.

Triggers for change involved far more variance:

- Participants were almost universally unprepared for the question, and took significant time to consider and refine their answers.
- Some overall triggers stated were:
    - A new space or piece of equipment providing an opportunity.
    - Taking over a new class from another practitioner. Exploring this trigger proved challenging – possibilities include practitioners wanting to make the teaching their own, or aligning a class to the practitioner's own experience with the topic were proposed but were not able to be confirmed.
    - Reviews, accreditations, audits etc. indicating a need for change
    - Poor results in assessment on a given topic.
    - Student affective responses, especially in centrally-run end of class surveys.
    - In 2 cases active reflective practice were stated as motivators, however these were not explicitly reported in the associated literature.
- Triggers involving explicit evaluation research techniques applied to the laboratories or their contexts prior to the changes were not described in any case, nor were any identified in the literature.

7.2.2.2 Artifact Review

Participants initial response to the PIPOV Framework were intended to inform the likely "minimum viable methodology" that might emerge from casual engagement with the framework. In practice this was primarily centred on misconceptions/misinterpretations of the visual elements of the framework – while most participants focused on subsets of the overall framework this did not appear to indicate any overlooking of the other elements.

Misconceptions particularly centred on 'Validation' and the way this was intended to interact and be integrated into processes, which saw significant revisions over the course of the expert reviews. In earlier revisions participants tended to 'read' the framework from top to bottom – which of course matches the order of the PIPOV acronym and put little weight on the cyclical nature of the framework, nor that validation was associated with more than simply outcomes.

Validation of inputs, processes, and outcomes was thus challenging, with participants not identifying or engaging with the idea of validation being applied partway through the design and running of a laboratory. Validation of purpose was unclear in terms of direction, with interpretations of this association being "validating that the purpose was achieved' rather than validating that the purpose was correct and/or complete.

Upon having the Purpose-First Approach explained along with details regarding the operation and contents of the PIPOV Framework, participants generally responded favourably to the design and the applicability to laboratory design and evaluation. Due to the direct one-on-one identifiable nature of the expert reviews this lack of negative feedback is not considered authoritative as the context may not have been conducive to giving such feedback. As such increased weight was given to the specific content of the discussion by the expert participants rather than compiling any affective response scores, i.e. what were their stated interpretations, what elements did they focus on, what avenues of applicability did they describe, what recommendations for change did they give.

Participants universally indicated that they felt the artifacts could be applied to laboratory design and evaluation. Initial sessions raised questions about precisely how that should be done, which highlighted a need for explicit process support in the artifact. Exemplars to follow were also raised as a need to aid understanding the artifact.

Some concerns were raised about the methods to be used – one participant asked whether this artifact meant quantitative data about the labs was useless, which required clarification regarding triangulation and that each method should be aligned to the purpose and what evidence is needed to validate it. This is likely captured as a need for exemplars with quantitative methods being represented among these.

Responses to these observations as well as areas in which participants focused leading to changes in the PIPOV Framework are detailed as part of the iterative sequence of diagrams described below.

Some of the discussion with participants revealed considerations that were emergent rather than expected. One participant when discussing the iterative process of improvement of the laboratory baked into the framework stated:

> "Yes, and then once you like it, you stop."

Digging into that statement revealed a desire for finding some laboratory that worked and letting that design settle into use, to be left unchanged. This was in direct opposition to the contextual basis of this thesis work, the iterative nature of EDR, as well as continuous improvement in education as the author was familiar with [208], and so was investigated at some length. The participant described their practice of engineering education from a perspective of time, effort, and risk constraints, paired with indefinite potential areas requiring attention.

At this time of ideation, the author was conceptualising laboratory innovations which incorporated elements from their background in software development and data science, leveraging the PIPOV framework. Laboratories that involved computer-mediated equipment or interfaces provide an opportunity for live data capture, similarly to what was done post hoc in section 2.4.4. This live data capture would provide the opportunity for live monitoring of laboratory execution, which would allow for drifts in contexts (students coming in with different knowledge, cultural changes, other external influences) to be detected and serve as a trigger for renewed evaluation. In this way evaluation would cease to be a once-off, project-based activity, but rather a continuous, service delivery activity.

If, however, one considers engineering education as a practice involving, as the participant phrased it "too many pies and not enough fingers," it would seem reasonable that the prospect of what could be interpreted as never-ending work on an educational laboratory design that was never 'finished' could be daunting. This occurred with a participant near the middle of the sequence of experts, and was followed up in subsequent sessions – each supported the perspective as one they shared. One participant shared their opinion that work on continuous improvement programs led to saved time in the short term, but that their institutional context made obtaining this initial outlay of time/cost/people difficult.

This presented two aspects of challenge to the design. First, how – and whether – to accommodate practitioners who wished to apply the artifacts to a pre-determined window of time. Second, how to communicate the process in a way that ensured the worst-case scenario interpretation given by the originally quoted participant of requiring constant design work, rather than establishing a baseline.

The question of accommodating practitioners was aligned with the idea of peripheral engagement with a minimum viable methodology. Experiences through the expert reviews, external engagement with the web survey, experiences in conference workshops to be described later, and most importantly the increasing impacts of COVID-19 on the workload of engineering educators (observed anecdotally by the author as a participant themselves, validated in e.g. [209], and an inevitable part of the casual introductions in every expert review) – led to the conclusion that any effective approach and framework must champion the most effective use of limited time, and support fractional or peripheral engagement.

The most significant active change in response to this was the treatment of supplementary materials in support of each of the elements in the PIPOV framework. While it was initially anticipated that that these would require supporting explanations and taxonomies for the ways in which they do or might vary across designs, and how these should be incorporated into a design and/or evaluation, the conclusions above coupled with observations around the level of focus given to each aspect of the artifacts indicated that this supplementary material appeared more harmful than supportive – rather than seeing an approachable framework, practitioners described a body of knowledge in which significant time and effort must be expended to understand it.

The outcome for this was to hold the supporting material in reserve rather than make it available along with the visual framework. This had little tangible impact on the expert sessions other than the initial visibility, with prior sessions focusing only on the framework image with any details discussed in conversation rather than being read. There was stronger impact however on the group session workshops held using updated materials described below.

Other than validation, time-pressure concerns, and supporting material, several further design changes were made in part due to the expert review sessions along with the early workshops. These changes were around:

- Stakeholders – this is linked to the validation of purpose.
- Designed vs actual.
- Cognitive and physical processes.

Key variants of different framework designs iterated over through this design process will be described.

7.2.2.3 Artifact Design Iterations

Throughout the early design and analysis phases of the meta-EDR process this thesis undertook to develop the Purpose-First Approach and PIPOV Framework artifacts, checkpoints were used to identify design variants of interest – either embodying certain design decisions or used during certain external contact phases. These key variants are described below, referencing the synthesis and empirical motivators behind the design decisions.

The participants described below are summatively the expert participants from the above study, as well as the workshop participants described in chapter 7.2.3 below. The workshops did not reveal new design criteria or motivations for change, but did confirm several identified aspects and gave an indication of how the framework might be used casually.

These key variants are almost all variations of the visual PIPOV framework, with one comprising the discarded visualisations of internal elements. This is due to the nature of the iterations subsequent to the primary literature synthesis stage. The Purpose-First Approach as an artifact embodying a conceptual, procedural, and methodological approach to the design and evaluation of laboratories in engineering education has at no time during this design process encountered any feedback or other empirical indications to suggest modification. The internal components of the PIPOV framework are coupled with the visualisation – not their contents/possible states, which are discovered rather than defined or designed as part of this thesis, but rather their method of communication is inherent to the visualisation. The exception to this being the discarded expanded visualisation.

It is reiterated here that the Educational Design Research meta-methodology employed in this thesis is not one that is founded on confidently proving hypotheses in order to inform improvements – it is used because doing so is infeasible. Empirical evidence in authentic contexts is critical, but this empirical evidence does not and cannot answer specific questions to degrees of confidence that would typically be expected were that question to be the sole focus of a research project employing some other methodology. Instead, iteration coupled with systematic reflection is used to slowly move in the approximate direction of some unknown target, in order to satisfice the initial requirements. The majority of the empirical observations made, and the conclusions made regarding them, are centred on verbs such as 'appears', or 'seems' – a general indication is observed, supported by triangulation via deeper literature synthesis if this is available, and bold changes are made in response.

7.2.2.3.1 Initial Design

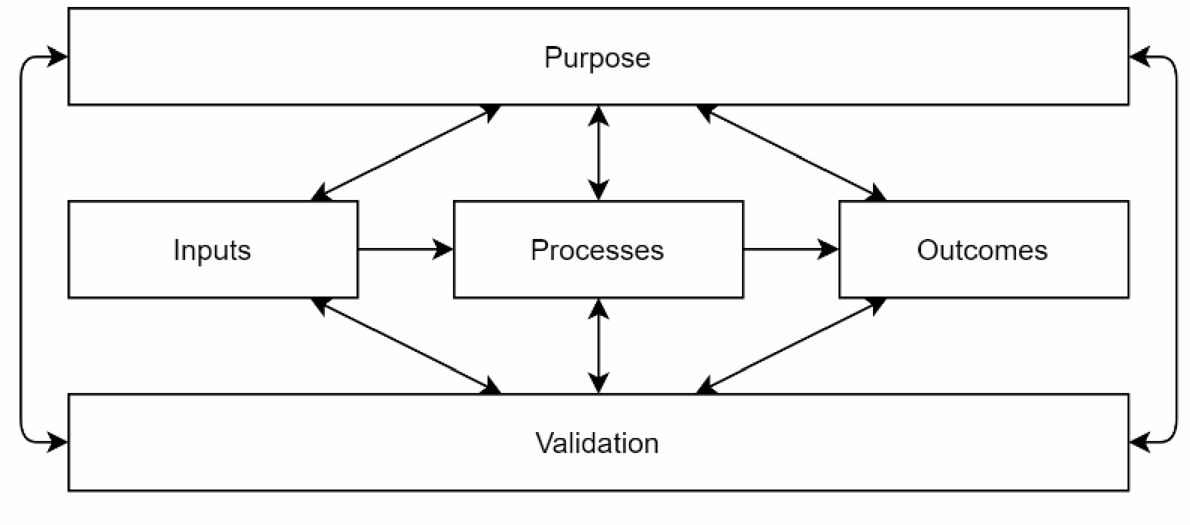


Figure 19: Initial version of the PIPOV Framework visualisation

The initial design of the PIPOV Framework was a purely synthesised design from the literature without a direct empirical source. This initial framework was derived from 4 primary contributors:

- Constructive alignment
- The Input-Process-Output framework system analysis [197]

- Nickerson et al's "A model for evaluating the effectiveness of remote engineering laboratories and simulations in education" [37]
- Brinson's "Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research" [21]

Further informed by the epistemological, procedural, and methodological foundation of Educational Design Research as embodied by the Purpose-First Approach component of the artifact.

These 4 contributions and the Purpose-First Approach remained as constants through the design iterations of this framework, and so the detailed discussion above may be referred to for the design of these elements.

The author consulted with their supervising academics to discuss and refine the intended designs both during the synthesis developing the initial design above, and throughout the design process.

The visualisation of these elements was intended to communicate process relationships between them. Inputs represent the design of the laboratory, these inputs lead to processes that occur while the laboratory is being conducted, which leads to outcomes. Each of these must be informed by the purpose for which the laboratory is being conducted, and each element must be validated.

Within each element in this initial design are the aspects included in the contributing models and frameworks. Inputs contains the various interface, proximity, synchronicity, and similar design attributes of a laboratory. Processes contain actions, motivation, engagement, and cognition, framed by constructivism/social constructivism as developed in constructive alignment. Outcomes are those observable dependent variables measured via observations, assessments, survey instruments, etc. Purposes include those presented in papers such as Feisel and Rosa [6], and categorised in Brinson [21].

7.2.2.3.2 Explicit Stakeholders

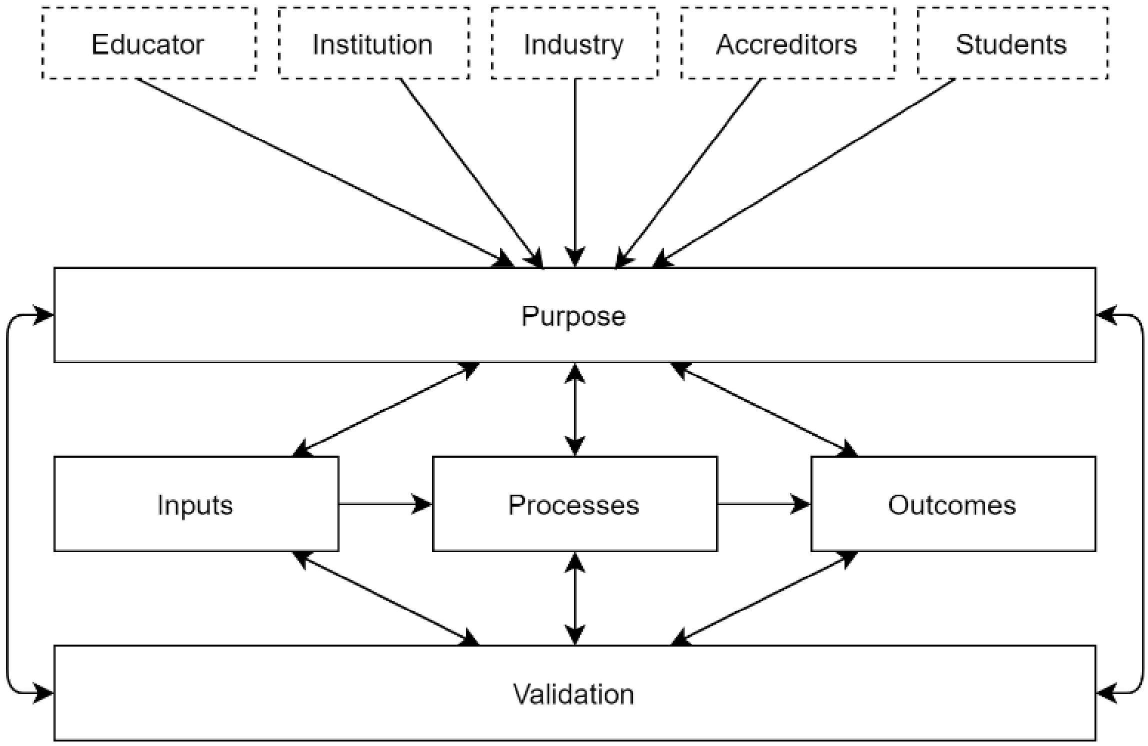


Figure 20: Iteration - Explicit Stakeholders

An early identified gap in the framework visualisation was that while Purpose-First lent itself to Purpose being the first element of a laboratory being designed or evaluated using the PIPOV framework, it was exceptionally difficult for practitioners to articulate purposes without support. The most pressing contextual hurdles were “where do the purposes of a laboratory come from,” and “how do purposes of a laboratory differ between people/entities,” which were not questions practitioners appeared to ask without prompting.

Instead, some practitioners stated the natural phenomena the laboratory was investigating, e.g. “this laboratory teaches students about thermodynamics.” Some identified communication or teamwork skills as the purpose. Others stated that they ran the laboratories for accreditation purposes, or to promote student engagement, or as an advertisement for a ‘practical’ university. At least one was willing to admit that they ran the labs because the academic they inherited their class from had run them.

By adding a layer of stakeholders as inputs to the purpose, participants were able to start those conversations and begin listing purposes from known sources – course handbooks, industry or government motivated skills or schemes and employability, accreditation graduate outcomes, student motivation and engagement, etc. By providing a ‘way in’ practitioners appeared far more comfortable articulating their purposes. This was already provided earlier by way of video conference conversation, however adding it to the visualisation itself provided a similar initiating effect. Note that this stakeholders layer was often extracted and shown separately as in the snippet in Figure 21 rather than being omitted in the subsequent designs.

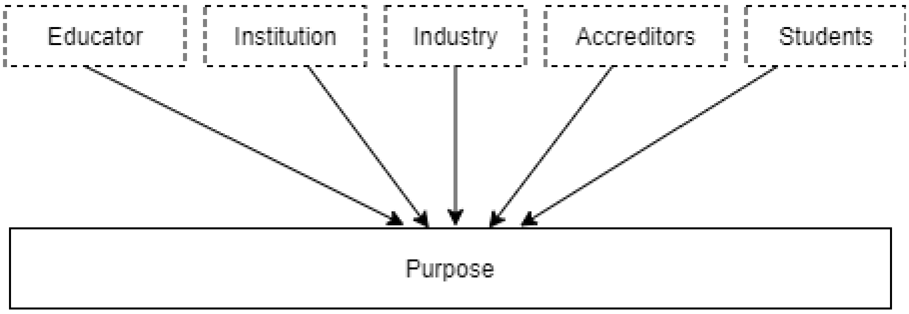


Figure 21: Iteration - Stakeholders and Purpose in isolation

This also helped with the idea of time/cost/effort constraints as raised above. Because the educator is themselves a stakeholder in what they need a laboratory to do and under what conditions, this allows a pathway for those constraints to be captured by PIPOV and inform the laboratory design and evaluation.

7.2.2.3.3 Decomposed Processes

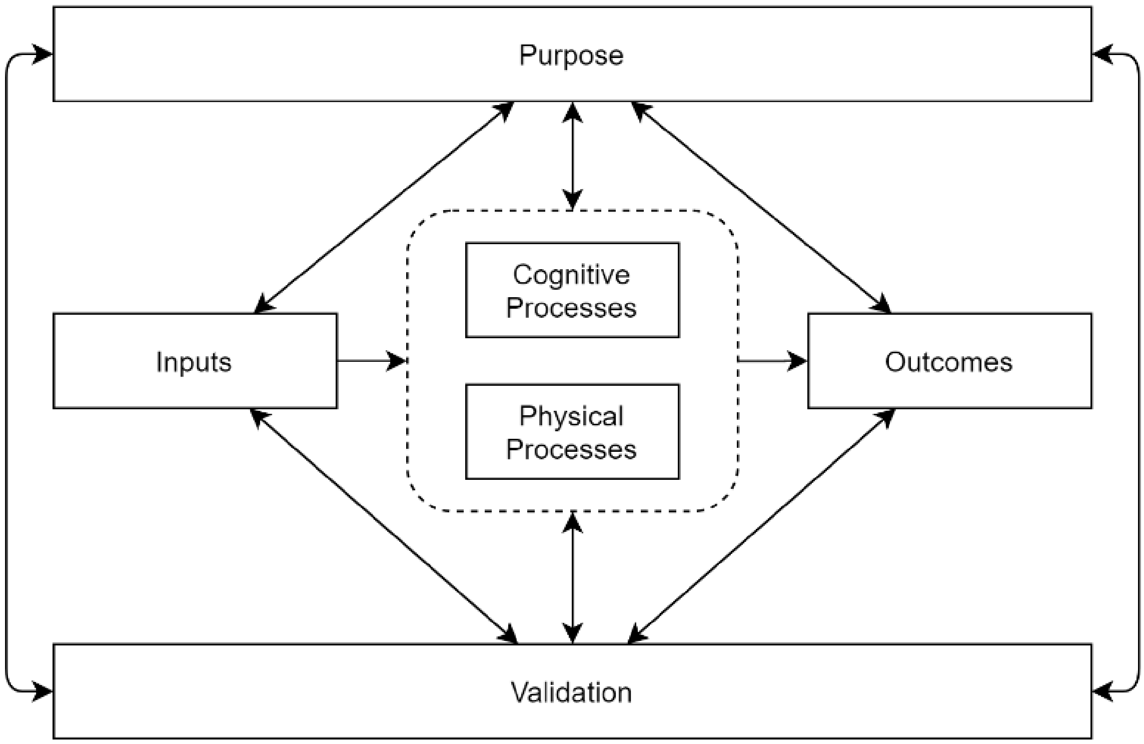


Figure 22: Iteration - Decomposed Processes

Note that these designs were particularly fluid between participants rather than a linear, rigid sequence, and some sessions involved the discussion of multiple design candidates, for example this design and the previous design were both discussed and contrasted in 2 practitioner expert reviews.

Processes has proven a challenging aspect to laboratory design and evaluation. It is by far the most abstract of the aspects of laboratories (though in PIPOV it also includes the most concrete, as physical processes). The contributing literature specifically within engineering education is markedly more vague about what is part of this element than any other, with Nickerson et al [37] for example simply splitting it into 'motivation' and 'cognition', treating these largely as black boxes rather than avenues for direct investigation and intervention.

By explicitly calling out cognitive and physical processes the scope of processes was intended to be made clearer, and an understanding of what these attributes are asking for could be gained. In practice this design made understanding significantly worse – or rather, it exposed poor understanding at the onset and turned it into an early blocker rather than something that could be dealt with later. In earlier versions participants skipped over processes implicitly, just as early researchers did. They understood processes needed to occur, but were used to black box education processes and so did not 'peer inside' pandora's box. By expanding this and explicitly calling out cognitive processes as the central point of interest, academics who were unfamiliar and possibly felt uncomfortable with the idea of cognitive processes lost focus on the other aspects of the framework.

Engineering education practitioners are used to being experts in a domain, and may experience unfamiliar alienation when presented with something in 'their' discipline that they are not or less familiar with compared to their engineering discipline work [65], [210]. This means that while highlighting the inclusion of both cognitive and physical processes was core to the framework, it was necessary to 'massage' that concept in more gently and without as prominent a display.

7.2.2.3.4 Desired vs Actual, Nebulous Validation

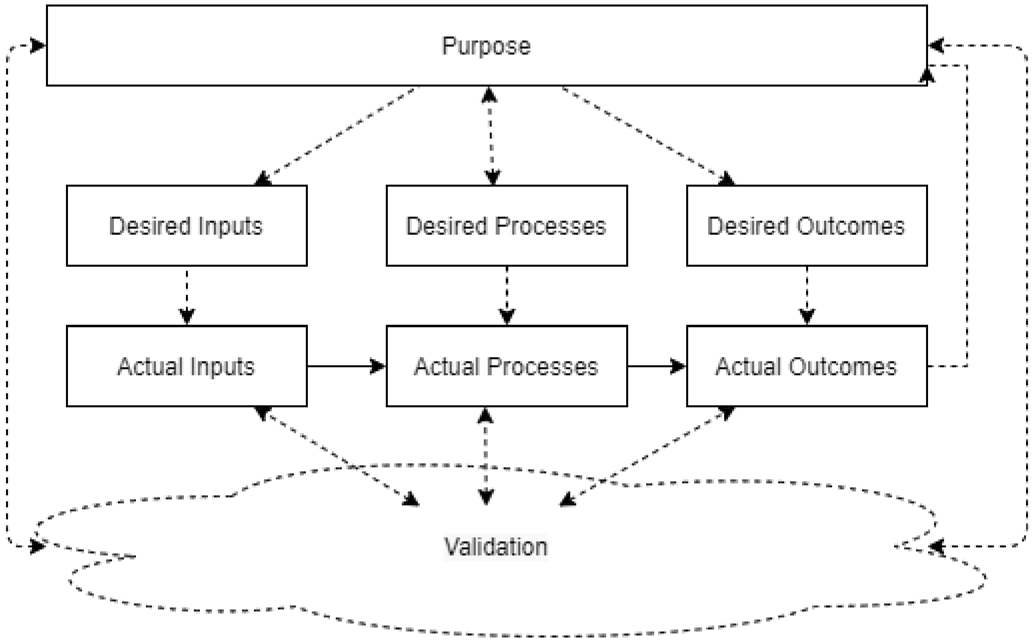


Figure 23: Iteration - Desired vs Actual

This version attempted to resolve processes in another way, as part of a large structural change to the framework. A key generalisability/transferability issue that is present in engineering education research broadly is the assumption that what was supposed to happen, did happen. Take for instance a set of instructions for a student undertaking a quiz. The instructions were designed as part of a research project by the instructor, and lay out what the student should do. The student is in a rush however, and just skips to the quiz in class's web portal. The instructor continues to report their project with their intended quantitative analysis, which has no empirical evidence that their instructions were indeed followed, but states "Students were given instruction document X" and in their analysis assume this inherently translates to "Students followed instruction document X".

This issue was initially raised for Inputs in particular as part of conversations with practitioners during the expert review. When describing their laboratory some practitioners stated e.g. "the students worked in groups of 3 and submitted individual work", to which the author queried "how do you know the students worked in groups?" – the answer to this and similar questions was invariably along the lines of "because that is what I told them to do on the laboratory run sheet".

By explicitly calling out this difference between the intentions an educator might have when designing a laboratory, and the reality of that laboratory in practice, significant avenues of investigation are revealed. Did the students work in groups, or did 1 student dominate and 2 passively watch. Did students told to use equipment in a certain way do so? Were the assumptions made about the students' prior knowledge accurate? This has the side effect of capturing a lot of nuance about processes, without front-loading the idea of cognitive psychology being a necessary component to understanding laboratories – allowing these ideas to be introduced more gradually as practitioners became more comfortable with the overall framework.

The distinction between dashed and solid arrows is meaningful – solid arrows represent an inherent relationship – actual inputs will inherently and inevitably lead to actual processes regardless of a designer's support of that relationship. The dashed arrow relationship between desired and actual inputs however is far less convenient, and will only exist if it is effectively designed to do so. This nuance was not often discussed until quite late in the expert review sessions, but appeared to at least trigger curiosity. A few practitioners suggested a legend or labels; however, this was intentionally left off as the distinction was too nuanced to be conveyed with a noun or single statement. The intention was for curious practitioners to ask the question and dig deeper into the explanatory material (or verbal discussion with the author as in the case of the expert reviews), and in this it was largely successful.

Validation being made a cloud instead of a box was a subtle change recommended by one of the practitioners themselves – when discussing the evaluation applications of PIPOV their assumption was that there was a mandated validation instrument or assessment to follow, and suggested a less defined way of communicating it. The practitioner's assumption was not unique, and came up multiple times in the expert reviews and especially during the workshops, however this framework visualisation change was ineffective in resolving this.

7.2.2.3.5 Encompassing Validation

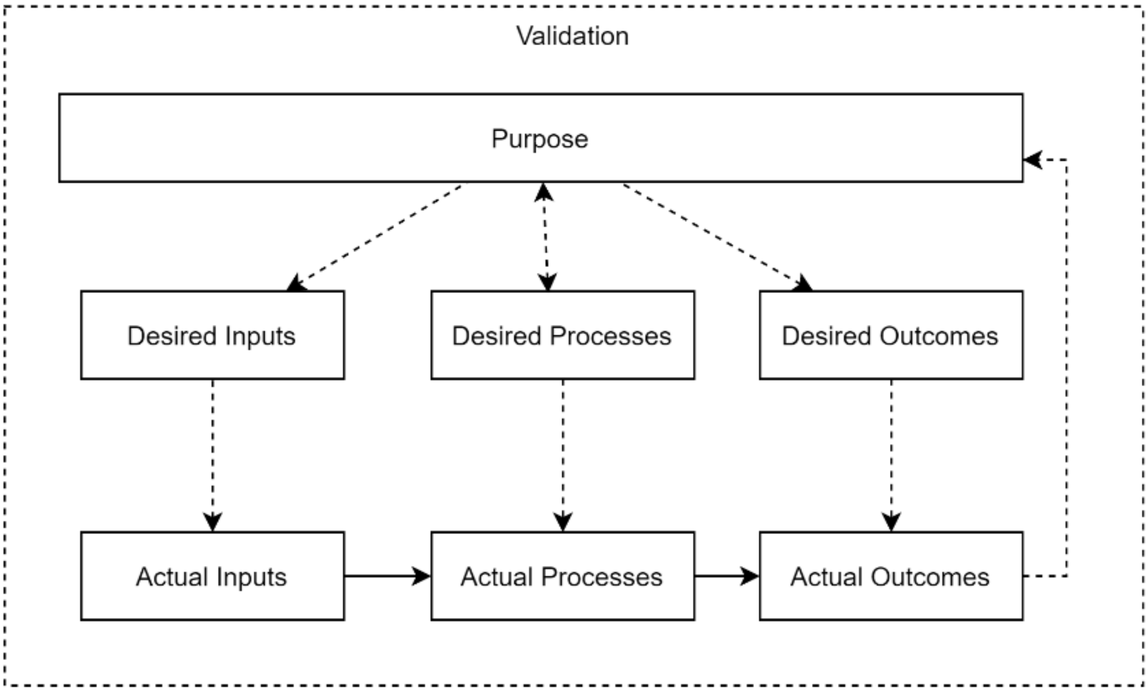


Figure 24: Iteration - Encompassing Validation

Continuing to identify validation as a challenging concept to communicate as described in the expert review discussions above, more drastic modifications were made to the presentation of validation as an element of the visualisation. This redesign, paired with a different order of explaining the framework verbally to match the new design, was intended to resolve the issues of validation conforming to summative research practices performing single discrete data collection and analysis steps rather than iterative, and in particular to validation being applied to more than just the final outcomes.

In practice within the expert reviews this change mitigated some issues while aggravating others. Validation was shifted away from an end-of-study step, but was no longer a point of strong focus and was discussed almost as an afterthought, which was the opposite of the intended effect.

7.2.2.3.6 Incorporated and Transitional Validation

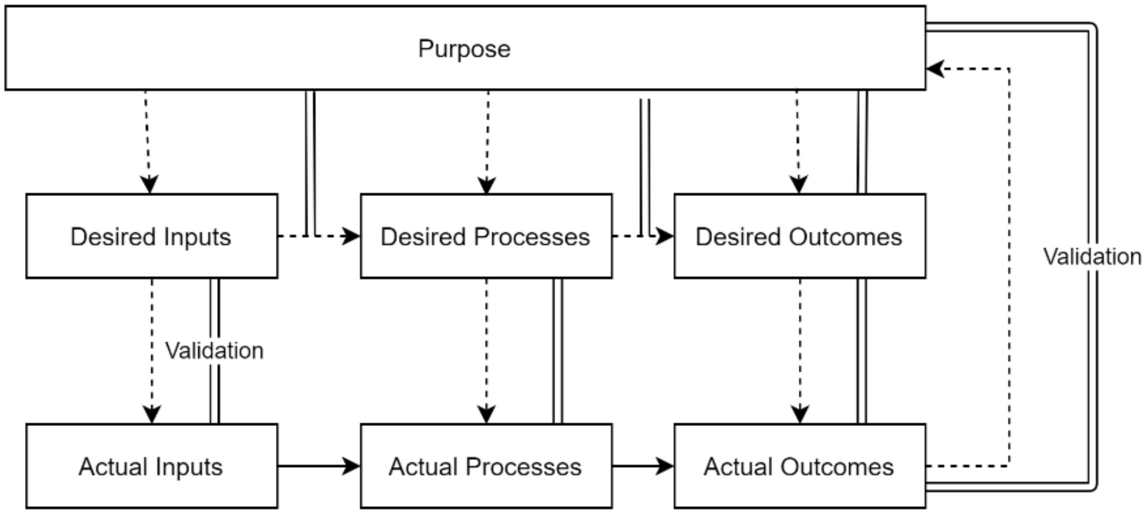


Figure 25: Iteration - Incorporated and Transitional Validation

Having largely settled on what would become the final attribute structure of the framework, most design focus was on the relationships between them, and practitioners being able to effectively interpret these. Validation was the most critical of these, being by far the most misunderstood component as described above, e.g. being associated only with outcomes, coming at the end of a process, how and when it should be applied to inputs and processes not being understood. A label was retained for Validation partly for historical reasons and to retain the 'PIPOV' acronym visually within the framework, but also to place additional weight to it given the issues in the immediately preceding design.

Converting validation from an element of the framework to a type of relationship within the framework proved to be highly effective at resolving this issue. Participants were still unclear as to precisely what, e.g. 'validate between desired processes and actual processes' meant, and far worse was e.g. 'validate between purpose and the connection between desired processes and desired outcomes', but they were immediately aware that this was something the framework called on designers to do and could ask for an explanation. This explanation provided a highly effective vehicle for discussing pre-outcome evaluation in further studies.

7.2.2.3.7 Nebulous + Explicitly Validated Stakeholders

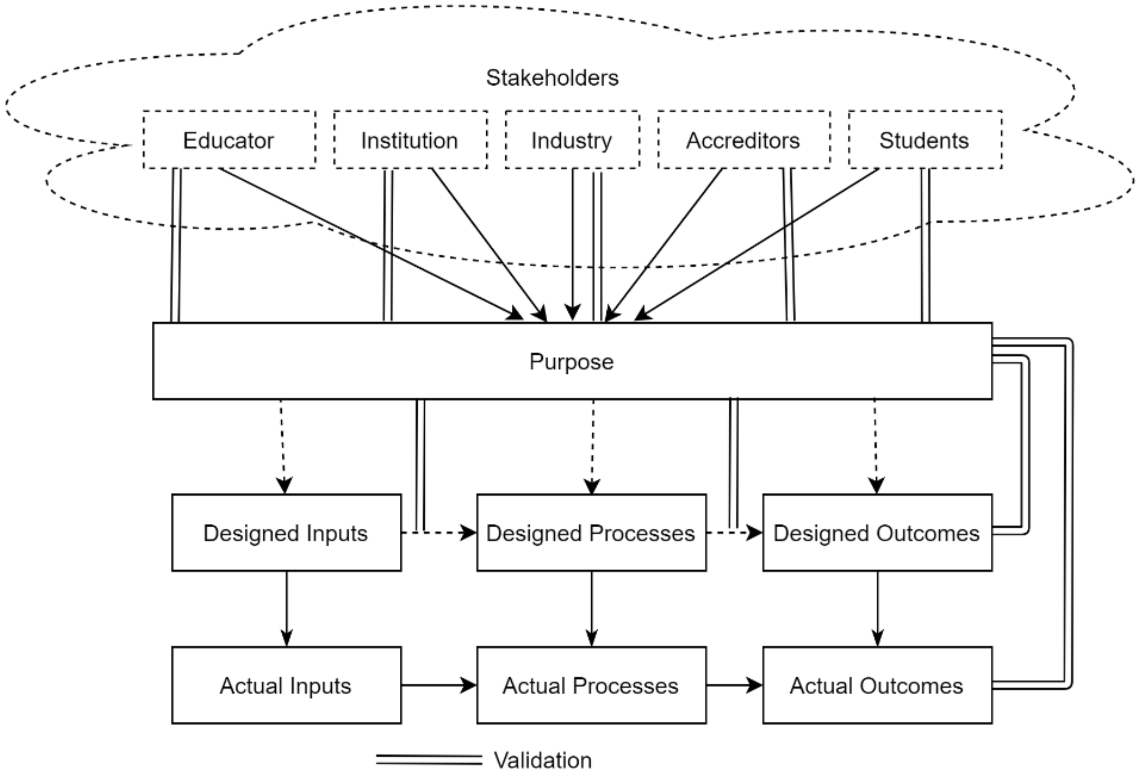


Figure 26: Iteration - Validated Stakeholders

The above represents the state of the framework at the conclusion of the expert review sessions, workshops, and after the collaboration study was well established. The next revision of the visualisation would be the final reflection based on the outcomes of the collaboration and both validation studies. The changes made to the above were the explicit and paired relationships between stakeholders and the purpose, adding a cloud graphic to stakeholders as a form of 'visual etc.', and removing the Validation labels down to a legend. The term 'desired' was swapped out for 'designed', to embed the idea of explicit decision-making rather than potentially unexamined intentions.

The stakeholder changes being the most significant variation were motivated by discussions with practitioners, especially in the workshops, regarding what to do about the purpose of a laboratory. This was partly triggered by the idea of validating the purpose after the run of the laboratory, and the idea of unintended yet beneficial outcomes. Understanding that in order to 'promote' a purpose from unintended yet beneficial to an intended purpose required calling back to the stakeholders in a laboratory to confirm that should be the case.

Adding the visual concept that there may be more stakeholders was in response to observations that society/sustainability/environmental concerns were not explicitly captured, other educators or professional staff, etc. These were generally considered to be contained within existing listed stakeholders (society for example captured in the requirements placed on engineering education via government and/or professional body accreditation), but as these practitioners clearly felt this was an oversight it was necessary to provide a way of allowing space for more without cluttering the diagram with a detailed breakdown of all stakeholders. This appears to have been effective with this comment not recurring in the more in-depth empirical studies.

Validation being extracted was part of a long series of minor tweaks to this – removing, labelling all, using a legend, etc. – the correct balance for this element proved elusive, and the author considers it likely that this is related to personal considerations of different readers without a single general solution. The presence of a legend as depicted above caused some participants to ask where the other line types were listed, the use of a full legend led to a regression to earlier misconceptions regarding the relationship between designed and actual, the complete removal led to participants asking: "what is the V in PIPOV?". The each-line label method was eventually selected, though without conclusive evidence over any other.

7.2.2.3.8 Discarded Attribute Contents

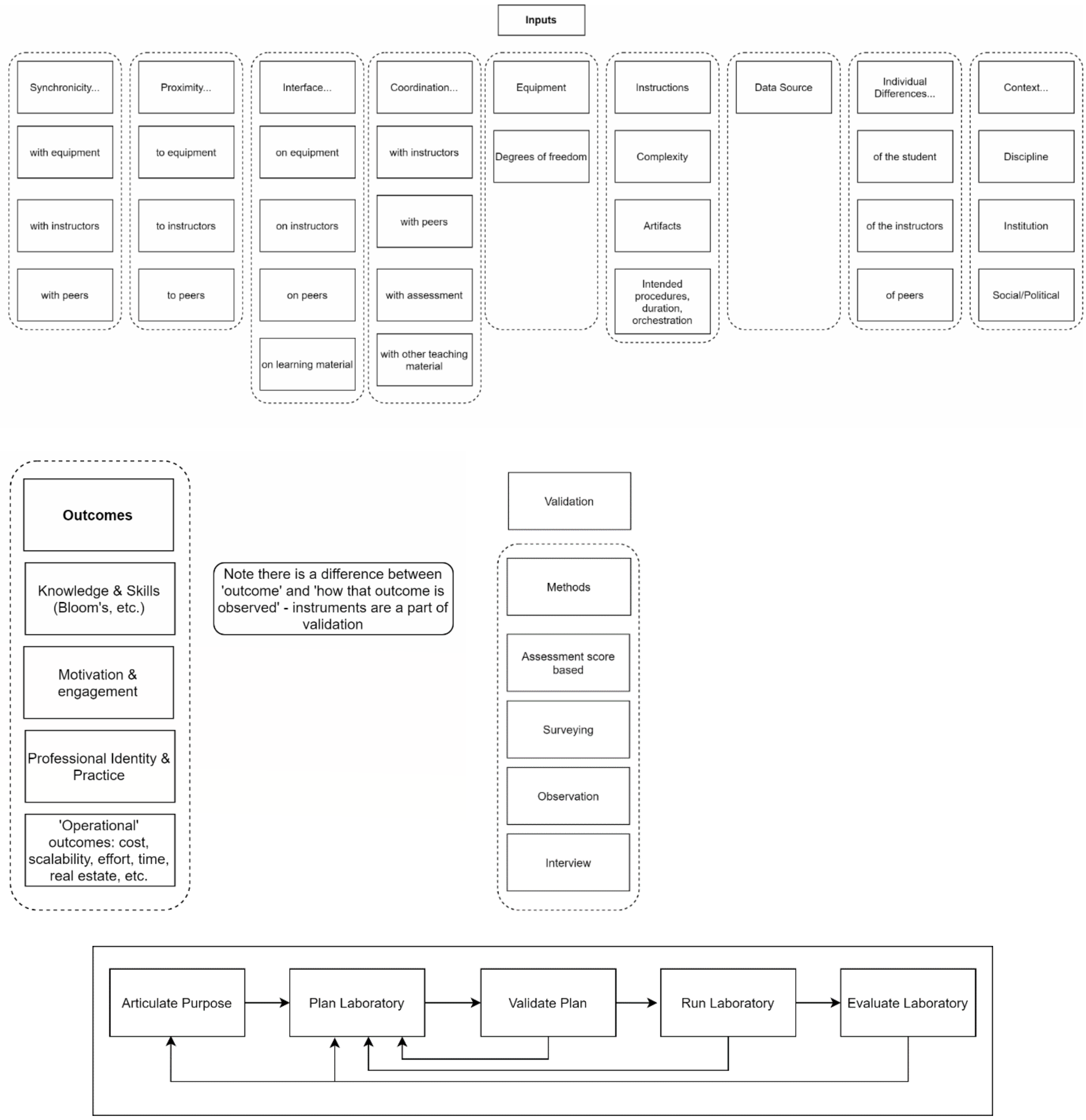


Figure 27: Iteration - Discarded Attribute Contents

An early intention on the part of the author regarding the communication of the Purpose First Approach and PIPOV Framework was to develop a web portal for this design and evaluation research on laboratories. This was partly motivated by the authors background in web development, but also inspired by research education tools such as the Sage Methods Map [211], which provides an approachable interface on a complex hierarchy of contained terms and concepts. Rather than the fully-expressed models as in Nickerson et al [37] this would allow for complexity to be hidden until the reader was ready to dive in. The drafted PIPOV framework iterations each included these contained elements, with examples depicted above, which would have been clicked through in a web interface to explain the specifics within each element.

As described earlier, the expert review sessions and especially the workshops were essentially concept-saturated with just the initial framework – it was clear there was too much cognitive load associated with presenting the framework and these reference materials for participants to encounter in a first instance. As a result, these aspects were withheld, and instead the author as the facilitator answered questions about the contents of each element or different relationships as they arose. It is likely that if the Purpose-First Approach and PIPOV Framework saw significant adoption without the author's direct involvement that access to this reference material would be more appropriate and valuable.

7.2.2.3.9 Final Version

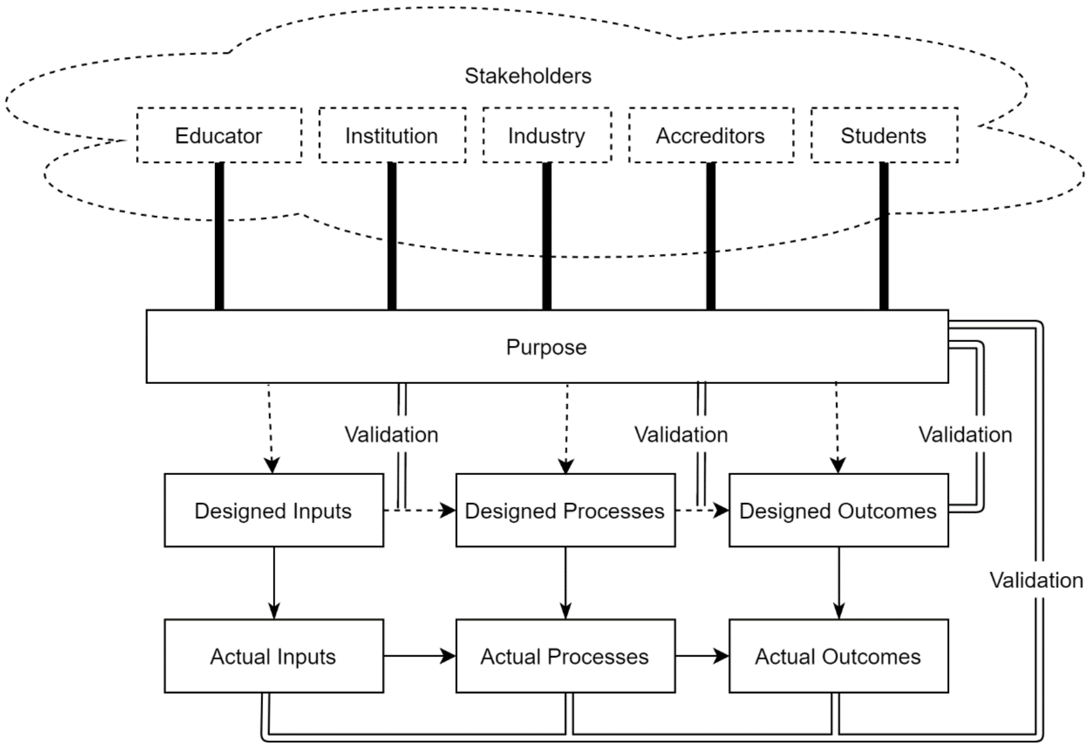


Figure 28: Iteration - Final Version

The 'final' version of the PIPOV framework is something of a misnomer, given the ongoing iterative nature of the methodology. It represents the state of the framework at the point in the collaboration and validation studies at the moment of reporting in this thesis. It is fundamentally similar to the prior state, with some minor conceptual changes.

The busy relationships between stakeholders and purpose were deemed distracting and adjusted to a new type of relationship – with similar intentions of stimulating questions rather than attempting to convey the full meaning directly. This was motivated by informal discussions of the framework's aesthetics with collaboration study participants. The relationship line between Purpose and Stakeholders was given its own characteristics to call out that the relationship is different to the dashed, solid, and validation lines, and similarly invite curiosity about the nature of that relationship rather than assumptions.

Validation of the actual inputs and purposes was made explicit, to more prominently call out opportunities for laboratory evaluation in the style of 'were the inputs to this laboratory what I intended them to be', etc. This was motivated by the existing laboratory modification field test, and the opportunities for this style of evaluation which were identified as part of that process. Validation labelling as previously mentioned was selected without significant evidence in favour of or against various other tested designs.

7.2.2.3.10 Design conclusions

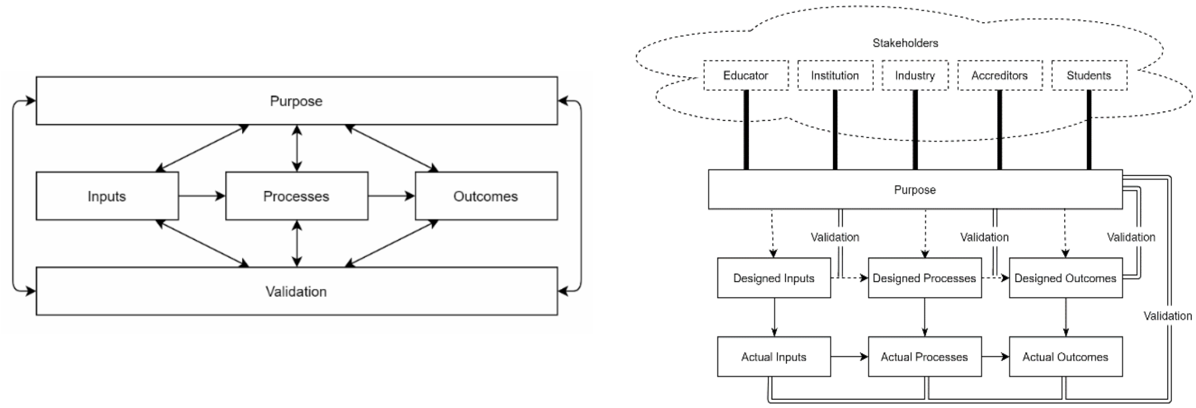


Figure 29: Iteration - End-to-end comparison

The design changes described above may appear trivial in isolation – the incremental process however has resulted in large changes overall. The designs have a markedly different application and motivate very different responses in the engineering education practitioners who apply them. Contrast the initial and final states - the final state has certainly paid for its clarity and specific motivations in complexity, however it also communicates a highly effective perspective of laboratories as teaching and learning activities, and implicitly leads practitioners down the path of educational design research on their laboratory without requiring them to undertake upskilling/education in methodological theories or cognitive psychology, incorporates cognitive alignment, widens the perspective of a designer to incorporate different sources of purpose, and furthermore supporting the effective communication of laboratory designs and research outputs leveraging the framework.

Practitioners encountering the refined version did not require the surface level explanations necessary to understand the initial design, and began by asking deeper questions regarding transition validation, or the nature of post-outcome purpose refinement. This should be seen as an example of a transferable artifact which can be applied in new contexts to design and evaluate laboratories in engineering education.

7.2.3 Broader Evaluation Workshops

Overlapping with the tail end of the expert reviews, the Purpose-First Approach and PIPOV framework were presented at relevant conferences as part of workshops, with the dual purpose of disseminating the work to the engineering education community, as well as obtaining feedback and observations as evaluation data for use in refining the design. The observations made during each of these workshops did not reveal any unencountered design motivations not obtained from the deeper one-on-one discussions of the expert reviews. As a result, the design iterations described above can be seen as an output of these also, and the workshop details are reported here separately.

Their evaluation value was largely in supporting the smaller sample size expert review findings, as well as providing confidence for particular less strongly advocated observations. The most impactful of these being the supplementary material change, which was a mild reaction during some expert review sessions, but which was far more strongly observed during the workshops. An example of some of this supplementary material is provided in Appendix F, specifically that used at the AAEE workshop, with the Australian university workshop material being slight conversions of this.

Both of these workshops involved participant observation following formal research protocols. Ethical review of the study design was approved under protocol number 2019/1040 by the University of Sydney.

7.2.3.1 AAEE Workshop, 2020

This workshop was run as part of the Australasian Association for Engineering Education 2020 Conference, which was originally planned for face-to-face sessions in Sydney, Australia. Converting to online delivery in response to the COVID-19 pandemic, this necessitated a change in the original plan for this workshop similarly to the emergency online conversion of the expert review sessions.

Delivery of the 90-minute workshop involved the use of exemplar laboratories as recommended by some of the expert reviews. These exemplars were presented through the lens of the PIPOV Framework to show how these laboratories fit into each category, and how the Purpose-First Approach and PIPOV Framework could be used to answer questions about what to do with these laboratories in the face of motivations for change. The active participation portion of the workshop involved mapping the participant's own laboratories, or intended design for laboratories, to the Framework for discussion. Note that the PowerPoint supplement in Appendix F was delivered in components – a combination of presentation slides and text heavy slides that would have been workshop handouts converted for online viewing.

The workshop participation was mixed, with N = 11 participants with highly varied engagement with the process. N = 6 of the participants were interested to hear about the work but did not present their own laboratories to use with the framework and were listeners only. N = 2 of the participants engaged in discussion on the exemplar laboratories and with other participants' laboratories, but did not have their own to work with. N = 3 remaining participants had their own laboratories to consider against the process.

Two of these laboratories were good vehicles for demonstrating the Purpose-First Approach – there was little complexity in the laboratories as they were described, both being standard traditional laboratories – proximal, hands-on, and with no equipment/space/teaching innovations in the current design. The most significant impact of applying the Framework to these laboratories was identifying that the participants examining these laboratories did not yet have an articulated purpose behind them. This led to subsequent questions being unanswerable, e.g. what data to evaluate and why, until the participants regressed back to earlier steps in the process.

One of the laboratories was a laboratory with existing published evaluation research, with the practitioner comparing methodologies directly. This laboratory was evaluated using a pre- and post-test between groups quantitative experimental design with no qualitative considerations. Similarly, to the less detailed laboratories, more significance was placed on the issue of laboratory purpose, as this was previously identified as a research limitation when evaluating the laboratory. Research alignment issues had previously occurred which limited the ability of the data to answer laboratory evaluation questions due to not targeting the relevant learning which was the laboratory's purpose.

The majority of the supplementary material was not used by participants during the workshop. Even leveraging exemplars and having the author presenting directly to the participants, the Approach and Framework visual were sufficient to fully engage the participants in learning about the methodology, without remaining capacity or time to engage with the nuanced details.

7.2.3.2 Australian University, 2021

A stand-alone workshop was delivered to a teaching and learning group within the engineering faculty of an Australian university. This workshop was delivered in two parts, combining an online workshop similar in style to the AAEE workshop but with a different participant activity, and followed up with a face-to-face visit to meet with specific interested participants.

The participant activity was refined into specific questions around laboratory purposes, and their application to their existing practice – with a shorter duration due to the overall time constraints of the session, 60 minutes, with N = 12 participants. Participants were asked to reflect on the laboratories they had been involved with teaching, and to list which ones they knew the purposes for. The participants did this before and after a presentation discussing the Purpose-First Approach and the PIPOV framework, and it was discovered that many of these participants, similarly to what was previously identified in the expert review, presented simply the handbook-entry intended learning outcomes. They largely missed purposes from any other stakeholders in the Framework, and upon reflecting of these the second time were able to identify significantly more and broader purposes.

8 FIELD TESTING

8.1 FIELD TEST I: NEW LABORATORY DESIGN

8.1.1 Context

The first field test was conducted with a group of academics in chemical engineering at an Australian university through 2022. The study was conducted as a direct email-based recruitment and fact-finding discussion, followed by a face-to-face visit to campus for a focus group design session, and including follow-up email discussions and several subsequent remote video conferences.

The validation studies were originally planned to occur longitudinally as with the collaboration study. The impact of the COVID-19 pandemic on engineering education however meant that academics had very little time for formal design and evaluation, which came through strongly during the expert reviews. While the author's initial expectation was that COVID-19 would trigger an explosion in remote laboratory adoption, with the exception of those who happened to be already poised to develop them as in the collaboration study interest in remote laboratories appeared rather to drop.

Engineering education practitioners ironically did not have the time or capacity to engage with teaching innovations due to the pressures of converting to emergency online teaching. Simulations promoted by commercial vendors promising to take on the workload was anecdotally far more popular. Engagement with recruitment attempts for studies was met with apologies, mentioning austerity freezes on new initiatives, or giving future dates for when capacity might later open up. This did however mean that when the effects of the pandemic receded that there was some resurgence of interest in laboratory improvement projects among academics seeking to correct flaws identified through 3 years of rapidly converted teaching.

One of these projects, the subject of the first field test, was the development of a pilot plant, a set of integrated chemical engineering equipment which would allow students to conduct full-lifecycle industrial chemical engineering processes, rather than operating different pieces of equipment in isolation, or only experiencing large scale processes as simulations.

The pilot plant was a multi-purpose set of equipment which was intended for use in multiple classes throughout the chemical engineering degree structure. It was also intended as a vehicle for multiple teaching and technological innovations, including:

- Digital twinning
- Augmented Reality
- Remote access
- Data capture on the equipment states and interactions for learning analytics
- Student co-design of teaching and learning

8.1.2 Design Session

After obtaining basic information about the pilot plant via email discussions described above, the design session was run with the author and N = 3 academics involved in the development of the pilot plant and associated laboratories. The session was run following a series of phases based on the original expert review protocol:

- Discussing the nature of the laboratory
- Describing the Purpose-First Approach and PIPOV Framework
- Analysing the laboratory against the PIPOV Framework.
    - In the validation sessions this step was modified to instead be conducted by the participants rather than the author, allowing them to make their own use of the provided framework with guidance and discussion from the author.
- Asking the participants to reflect on the process.

The pilot plant was already well into the design stage during this session, though the classes using it as an educational laboratory were in the early stages of design.

The purpose motivating the creation of the laboratory was a lack of authenticity observed in smaller scale laboratories. Chemical engineering graduates would otherwise be required to transfer mental models built around single pieces of equipment to a context of many hundreds of pieces of equipment used in some chemical processing plant. This was coupled with an examination of the authenticity of students physically manipulating the equipment directly, rather than establishing the relationship between engineers and technicians as often observed in industry – with engineers often not directly manipulating equipment, but needing to effectively communicate their intentions to technicians.

The equipment was intended to be used in two classes initially, with more planned to be added after these were successfully tested – a 2nd year heat and mass transfer class, and a 3rd year process control class. These were intended to be taken by streams of students such that the initial participants in the 2nd year class would revisit the equipment in the 3rd year. A major argument behind this repeat visitation was the overheads in initially training students to use the equipment, which needed multiple visits in multiple classes to be a viable cost of class time.

8.1.3 Discussion

The pilot plant field test raised several aspects for reflection regarding the Purpose-First Approach and PIPOV Framework. The first of these is a fundamental difference in alignment that was possible when using the term 'laboratory' between people with different perspectives. The fluidity of the term laboratory was well established early in the work on this thesis, however there were axes of variability that were a level of variation not originally considered in the original analysis. This is partly due to the incoming conclusions which motivated this thesis work in the first place – that laboratories must be considered within their authentic contexts. The misalignment was that a perspective existed such that 'laboratory' could refer to some piece of dedicated laboratory equipment which was simultaneously used in multiple subjects, rather than a teaching and learning activity which might *use* some equipment.

This contrasts with the truss used in [39], which while used in 2 different classes was considered to be equipment which was used in 2 different laboratories with different purposes. Given the far larger investment of time, money, space, and effort required to develop equipment such as the pilot plant compared to the truss, it is entirely reasonable that this equipment would need to be designed to serve multiple laboratories to be viable.

This requires a shift in thought to accommodate the laboratory equipment as a point of intersection between n teaching and learning activities – each with their own purposes, some of which will flow through to the equipment to be complementary or competitive with the others, and with purposes behind the laboratory which may not be directly associated with any current teaching and learning activity, e.g. future-proofing, or allowing use outside of formal class contexts such as a makerspace.

Attempts to reasonably capture this visually in the PIPOV Framework proved inappropriate, as the complexity immediately grew exponentially, and simple diagrams communicated little. Instead, situations such as this call for a portfolio of related PIPOV Frameworks tied to specific laboratories with the equipment in common. This has the added benefit of allowing futureproofing to be designed according to more systematic speculation, with example future laboratories given an example PIPOV Framework analysis to identify the most likely requirements of the equipment. Given a suspected feature for the equipment to accommodate, expand this into a full laboratory design example and consider what possible consequences of this affordance might be – particularly dependent features not already present.

Another aspect for reflection related to this is the possibility for cross-class integration with the 'same' laboratory. Again, this seems best served visually using multiple instances of the framework rather than somehow integrating them, but it raises particularly interesting questions around the idea of purpose. Up to this point, all reviewed and speculated laboratories had been designed to serve a common set of purposes, which were designed for all students homogeneously. Recognising this it is immediately apparent that this is inauthentic: students coming in with different levels of knowledge, different learning needs, etc. will require different things of the laboratory, or not require it at all. This is different in nature to learning objectives/outcomes, which remain true even if a student had no need of the class (by the end of this lesson you will be able to… remains true even if you were already able to demonstrate the knowledge). There are similar considerations regarding equitability and inclusion – the purpose of the laboratory might align or misalign with specific individual needs.

Using the cross-year common laboratory idea as a vehicle, there are some ways to resolve this apparent gap. If the student groups are discrete enough, e.g. different year cohorts, the analysis can involve 2 different laboratory designs, that happen to share equipment as in the previous observation and where the student groups are themselves inputs as design criteria for the other students and vice versa. If the student groups are less well-defined, un-measurable, or subject to changes midway through a laboratory, then these instead must be identified or declined as purposes for the laboratory, such as accommodations for different individual access requirements for laboratory equipment where possible, or alternatively identifying any circumstances where accommodation may not be possible.

An observation regarding the participants' use of the PIPOV Framework was not to 'fill in' the framework with a design, but rather thinking the laboratory through each aspect of the framework and write notes applying learnings from these steps to their own generated artifacts, in this case schematics and handbook entries. This may have been motivated by the misalignment mentioned between the participants' perception of laboratory and the perspective embedded within the framework, or it may have been motivated by the design of the laboratory existing for the participants as being embodied by those existing design documents, and having an already-established method of note-taking as part of their design process.

A specific design aspect which was raised by the PIPOV Framework process was the possibility of demonstrator dashboards and/or smart tutoring. This was raised during discussion around the multiple interfaces for the pilot plant (proximal, augmented reality, remote) as well as the analytics data capture, looking at real-time interventions. This grew out of discussions looking at real-time interventions from a perspective of laboratory evaluation. Demonstrator dashboards would be based on real-time and historical analytics data from the equipment regarding student use, in order to flag concerns or opportunities for demonstrator intervention. Smart tutoring systems would perform similar tasks, but with the system conducting the intervention rather than a human demonstrator.

Real-time interventions were a component of future user interface plans in the Edinburgh study, and at the time of this design session were becoming an increasingly prominent and defining aspect of the framework aligned to its white-box approach to teaching and learning activities – if the laboratory design can be modified mid-run in response to observations made mid-run, then the teaching of students can be changed while they are learning in response to direct observations. This is of course how direct teaching communication often works, but is limited in a laboratory with didactic instructions, time constraints, or high student-to-teacher ratios. Technologically-enabled laboratories, which have been far more common than traditional laboratories when analysing with the PIPOV Framework, present the opportunity for a revitalisation of this bespoke teaching practice.

The overall perception of the Purpose-First Approach and PIPOV Framework were positive. Participants observed that the methodology more closely aligned with their expectations of quality design practices than other methods of evaluation – with one participant observing that the iterative and purpose-driven nature was better aligned with their industry experience than previously encountered academic evaluation. The application of the framework gave clarity to design decisions that had already been made:

- some design decisions having their motivation already at hand, such as the selection of particular equipment to fulfil a complete production cycle allowing use through multiple subjects through an entire degree,
- others requiring a deeper reflection on the intended learning purposes and logistical constraints of the laboratory as motivators, such as the way in which the proposed innovative technological capabilities would be leveraged in teaching and learning,
- and raising some design questions that had either been implicit or not yet considered, such as
    - the integration of laboratory safety training across classes beyond the common thread of the pilot plant equipment,
    - and interestingly a quote from a participant unprompted by the author that led to email and video conference discussions

        “what even IS a laboratory?”

      when considering the interaction between the pilot plant and wider undergraduate laboratories in the degree structure.

Design, evaluation, and teaching work with the laboratories is ongoing.

8.2 FIELD TEST II: EXISTING LABORATORY MODIFICATION

8.2.1 Context

The second field test was conducted with a pair of academics in electrical engineering in an Australian university, also through 2022 and early 2023. The study process followed the same sequence as the first – email recruitment and fact-finding followed by a face-to-face visit for a focus group design discussion, as well as subsequent email discussions and remote video conferencing.

In contrast with the first field test, this study was concerned with changes to an existing laboratory in active use. This provides a window for analysing the Purpose-First Approach and PIPOV Framework applied to a different situation to the previous authentic context studies, one which aligns to the continuous improvement through iteration theme embodied by the artifact.

The laboratory under study was a remote electronics laboratory used in multiple electrical engineering classes. Like the previous field test (which while heavily overlapping this field test in time had its design study earlier), this laboratory *equipment* was used in multiple laboratory *teaching and learning contexts* – however a single class was the focus of the new improvements with other contexts in current use could retain the same experience, and new contexts could be considered later. This class was run in both a face-to-face teaching mode, and an entirely online remote teaching mode – notably designed as online teaching as part of a degree structure pre-dating COVID-19, not emergency remote teaching. This split meant that similarly to the two initial classes for the pilot plant equipment there would need to be both independent and combined consideration of the laboratory equipment design.

The use of the laboratory was expanding to more sessions within the teaching and learning context of interest. This was motivated in part by the handover of the teaching and learning context to a new academic for the face-to-face teaching version of the class leading to renewed interest in the laboratory, and in part by a successful research project run with the laboratory prior to the commencement of the field test by the academic running the online version.

The laboratory equipment comprises a series of electronic components wired up to a switching matrix, which allows for programmatic construction of real electronic circuits connecting the different components with an exponential number of possible configurations – thus allowing students to design and test different circuits without needing to have the components, wires, breadboards etc. physically in front of them. The equipment also includes a function generator, oscilloscope, and digital multimeter for signal generation and observations, and a camera is pointed at the lab equipment so that remote interface data can be cross-checked with the local equipment displays.

The equipment can be connected to via the internet, with a desktop application used to connect via the internet to the equipment and build circuits. These circuits are displayed visually as if they were being constructed with wires, which the system then interprets and builds in an electronically near-identical way. The mapping is not perfect in the fundamental physics sense (e.g. what is 1 resistor in the model might be simulated by multiple resistors in series in reality), but at the level of abstraction and measurement error the undergraduate classes using the equipment require they are indistinguishable.

The research project which had been conducted previously involves a networked communication component, allowing text-based communication with other users logged in to the system. This component was the output of previous enhancement work on the equipment. The research project added a simple trigger-and-response chat bot to this communication component, which leveraged learning analytics data from the interface to trigger support statements about what the student was doing.

This interaction is mostly characterised as a 'pull' model – students must query the chatbot to obtain a response, rather than a 'push' model where the chatbot might react on its own to observed states in the system. It will partially react to observed states and indicate it is available to answer questions, but will not provide answers to issues without being asked. It combines data from keywords in the student's text request with the state of the components the student has input to make pre-configured support statements, such as "The oscilloscope trigger level will cause sweeping."

8.2.2 Design Session

The design session involved the author with the N = 2 academics running the different modes of the class of interest, held face-to-face after the earlier email and video conference discussions. The pre-existing laboratory nature of the design session led to tweaks in the session protocol. Rather than:

- Discussing the nature of the laboratory.
- Describing the Purpose-First Approach and PIPOV Framework
- Participants analysing the laboratory against the PIPOV Framework.
- Asking the participants to reflect on the process.

The phases were instead:

- Running sample experiments with the laboratory, simulating student uncertainty and the support of the implemented chatbot.
- Describing the Purpose-First Approach and PIPOV Framework.
- Using the PIPOV Framework as a frame for laboratory enhancement ideas that had been floated by the participants, or discovering new ideas.
- Asking the participants to reflect on the process.

The purpose of the laboratory in the classes of interest varied between the face-to-face and online modes. Both were interested in using the laboratory as a vehicle for investigating fundamental laws governing electricity, in support of theory-based instruction. Understanding the use of equipment such as oscilloscopes and gaining skills in setting them up to correctly measure a circuit, and in particular recognising the signs indicating they were configured incorrectly was also stated as a purpose for both.

The face-to-face mode included a physical, proximal laboratory involving components, wires, breadboards, etc. The purpose for this over simply using the remote laboratory more was expressed by the academic running this mode as

> "I would never trust an electrical engineer until they have seen the puff of smoke their mistake caused by blowing up a capacitor in a lab."

However, the academic was still increasing the use of the remote laboratory in their class as a complement to the proximal laboratory – the stated purpose for this was due to the severe time and space constraints involved in having the students undertaking a physical laboratory, replacing and testing components, etc., coupled with an observed need for the students to gain more practical skills with electronic equipment.

The online version of the class was similarly increasing its use of the laboratory. This was initially triggered by the development of a second set of equipment which mirrored the first, and allowed double the student capacity and thus more lab sessions in the same amount of time. The purpose motivating the use of the laboratory was thus the same as it had been, but a constraint on that purpose was mitigated.

By far the most significant trigger for enhancement however was the available time and interest of the involved academics. The new academic coming on to the class and taking over the face-to-face mode, coupled with the completion of the online academic's chatbot research project, meant that time, effort, and interest were available to 'spend' on improving the laboratory.

Explaining the Purpose-First Approach and PIPOV Framework went similarly to the previous field test – the most common reaction of engineering education practitioners being one of realisation rather than surprise; "It sounds so obvious" being the usual response, also observed in this case. It aligned with the actual design practices used when developing the prior laboratory chatbot enhancement, though conflicted with the research design used.

The most significant engagement with the framework occurred when discussing the contents of Processes, both Designed and Actual. The academics described the physical processes involved in the laboratory, and the influence of the chatbot on correcting observed mistakes in those physical processes was also raised. However, when the author described cognitive processes and how they might be considered in the conceptual movement from inputs to outcomes this appeared to be a surprise to the academics and led to an expansive discussion of what these processes involve, theories that inform them, and how these could be leveraged in the laboratory.

The concrete outcome of this discussion was an idea to enhance the chatbot in the laboratory to not simply be reactive to a given state of the equipment, but rather to react to the *behaviour of the student* and use that data to identify patterns indicating learning success or difficulties, and intervene the way a human tutor might. The techniques for this were discussed in the context of the prior behavioural identification work [76], with the electronics laboratory having similar interaction logs which could be leveraged for this purpose.

8.2.3 Discussion

Interestingly, for the academic running the online mode, the observed need to have more practical learning was *imported* from observations with other classes, as this class was being run by them for the first time. Initially seeming at odds with the contextualised purpose of the laboratory embedded in the Purpose-First Approach, upon reflection this can be viewed as the incorporation of expert knowledge in the design of a laboratory the same as the selection of some input criteria might be made based on prior knowledge rather than exhaustively testing each possibility. Just as the design attributes can be selected from based on heuristics and judgement, the need for change can also be motivated by the same information from external contexts – so long as its impact on the context it is being applied to is measured within that context.

The 'time, effort, and interest' motivator for change speaks to a frustration observed throughout the interactions with practitioners in this thesis work. Over the course of the project beginning in 2019 after initial scoping and synthesis through to 2022, the author reached out to hundreds of academics via mailing lists, research groups, conferences, direct individual contact via email, networks of thesis supervisors, etc., Of these a response of some kind that usually did not progress to formal study participation occurred with more than 100, out of which the participants in the expert reviews, workshops, collaboration study, and these validation studies were drawn (the total academic participants in this thesis of varying levels of engagement being N = 39, excluding collaborators with participating academics who did not have direct research data collection interactions with the author).

The overwhelming sense from these academics – influenced by COVID-19 perhaps but still present in 2019 – was that they were often quite interested in teaching, and their use of laboratories, but had no capacity to participate in research on them, no time to spend improving the laboratories. This is likely an appropriate response – given the number of moving parts involved in the overall engineering education process, the likelihood that any one laboratory was the most critical aspect to change in that specific context is low.

Seeing the reverse of this effect, that is, open up time and space for academics to explore teaching enhancement and they engage heavily, is a consistent theme observed throughout these empirical studies. The Edinburgh project was supported by institutional desires for change and academics were given the time, space, and funding to achieve it, the same way the previous field test occurred through significant institution investment in support of an observed capability gap.

This confirmed the changes in the design of the PIPOV Framework to exclude supplementary material until it is needed – academics short on time need the briefest, most accessible version of a motivator for change, and then if there is support for doing something about that motivation detailed material can be used at that time.

This idea of PIPOV being a holistic, contextualised representation of an entire laboratory context but still able to be broken down and viewed as subsets was related to this conclusion. The UI evaluation from the Edinburgh project was focused on purpose, the actual inputs of the UI, and the difference between designed and actual cognitive processes within the laboratory. The revealed design decisions to be made in the previous field test were each focused on different aspects, such as the demonstrator dashboard targeting the designed vs actual physical processes in the laboratory. And in the case of this laboratory, the smart tutoring analytics system shifting the focus from purely physical processes to include cognitive.

This means that the peripheral engagement with a minimum viable methodology will likely not follow the same patterns as observed in Bloom’s Taxonomy and Constructive Alignment. Where Bloom’s Taxonomy sees focus on the cognitive domain rather than affective and psychomotor, and Constructive Alignment begins with the central image of 3 elements, the initial engagement with the Purpose-First Approach and PIPOV Frameworks is likely to be based on the aspects of the laboratory the practitioner is interested in. It has however been observed through these studies that these snippets of focus do begin with a consideration of purpose. Either explicitly including it as in the Edinburgh study, or framing the question of how to design the aspects of interest based on what the purpose motivates as in the validation studies.

As with the other field test, design, evaluation, and teaching work with the laboratory is ongoing.

PART 3: DISCUSSION & CONCLUSIONS

9 SUMMATIVE DISCUSSION

This thesis set out to apply the Educational Design Research process/methodology to the development of an artifact in support of the design and evaluation of laboratories in engineering education. This artifact was proposed as a component of the answer for the core research question:

> "How may laboratories in engineering education be effectively designed to meet the purposes their stakeholders require, and evaluated against these purposes?"

Through empirical collaborative development and validation studies, these artifacts were iterated over, investigated, and refined as an answer to the research sub-questions:

1. How do the research methodologies used in evaluative research on laboratories in engineering education support or inhibit answering the questions this research poses?
2. What methodological alternative provides the appropriate characteristics to resolve the identified issues with experimental approaches in the design and evaluation of laboratories in engineering education?
3. Can Educational Design Research be successfully applied to the design and evaluation of laboratories in engineering education?
4. How can practitioners be supported in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?

4.1. Do the Purpose-First Approach and PIPOV Framework support practitioners in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?

The evidence in support of those questions has been derived from ongoing work in authentic contexts of laboratory design and evaluation involving:

- 11 academics involved in the AAEE workshop with light engagement.
- 12 academics involved in the Australian University workshop with light engagement.
- 9 academics involved in the 'what is a laboratory' and expert review sessions with heavy engagement during a single session.
- 2 academics directly involved in the collaboration study, with a further 3 involved on the Edinburgh side with heavy engagement over a period of multiple years.
- 3 academics involved in the new laboratory design field test with heavy engagement over a single session and ongoing engagement.
- 2 academics involved in the existing laboratory enhancement study with heavy engagement over a single session and ongoing engagement.

For 39 participants total, informing different stages of the research. The specific methods used were qualitative as a result of the small numbers – long term engagements nature of laboratory design as a subject of study. These most closely map to case study research as a methodology, within the overarching structure of Educational Design Research as a process and lens for analysis.

The most challenging question for a research study of this level of scope (which could never be fully answered in the space of a single PhD thesis – the development and adoption of Bloom's Taxonomy for example having been ongoing for over 70 years) and iterative nature is… when do you stop? At what point is the artifact under development complete, the learnings about it saturated, the answers to the questions sufficient. The benchmark used to determine the ‘end’ of this thesis work was the identification of a point where the fundamental nature of the research needed to change.

A common thread to all of the above empirical work is that it directly and intimately involved the author. The author was describing the concepts, artifacts, and methodologies. The author was collaborating in the design itself, making observations, and suggesting specific design decisions for the practitioners to consider. This deeply integrated nature of the study led to clear and significant insights into the working of laboratory design processes by participating practitioners, but has a limitation to authenticity. The next step for investigation of the developed artifacts calls for practitioners to make use of the artifacts independently, work on laboratory design and evaluation using the artifacts on their own without direct involvement by the author, and for longitudinal check-ins on this process to discover how the artifact is used, interpreted, or modified in unimpeded contexts.

Given the long-term nature of this research as the next step, this phase of development is considered the best point in an ongoing process to pause new initiatives, reflect overall on the completed iterations, and report the findings as this thesis.

The majority of the individual insights and conclusions are part of the design process detailed above and should be considered in context. The purpose of this summative discussion is to identify learnings that exist across those studies, explore any bias or limitations to the research, and reflect on the research process overall.

One observation that cuts across the participants is they were all engaged with engineering education research. In some cases, experts with decades of experience, in other cases novices starting out, but in all cases interested in engineering education and the design and evaluation of laboratories – hence their agreeing to participate. This may represent a flaw in recruitment, but on reflection seems to instead be authentic, though with a caveat. The practitioners who are likely to champion a new initiative, methodological practice, or be engaged and reflective about their teaching laboratory designs at all are going to be those who are engaged with this research discipline – and so it makes sense to investigate that group as a pool of research participants. The caveat lies in the issue this raises around engaging the unengaged in order to improve their laboratories as well.

This is an aspect which was not studied in the work on this thesis – and likely is something that while associated with the Purpose First Approach as something an engineering education practitioner not engaged in reflective design and evaluative research might not engage with, is likely merely a symptom of needing to engage those practitioners holistically rather than trying to cross that chasm independently through some changes in the artifact design.

A corollary to this is that the laboratories where deeper engagement was available, in the collaborative and validation studies, were all innovative. Whether remote or industrial scale, these were not traditional laboratories undergoing stock standard design. This is related to the above participant type discussion as well as the observation in the literature review that ordinary, stock standard laboratories are not evaluated, studied, or reported in the literature in general. This does present a concern that limits the potential confidence in those studies, but is mitigated by the experiences of the workshop studies and expert reviews, where participants discussed the artifacts’ application to some ordinary laboratories and still engaged with them effectively. It was also present in the existing laboratory enhancement study in the form of the face-to-face mode having a traditional hands-on laboratory, which fit into the framework process just as effectively as the remote laboratory.

Another observation relates to peripheral engagement with the artifact, or the minimum viable methodology. This was observed to be inconsistent among the practitioners, with the first aspects of focus and the aspects of most consistent focus varying heavily between different participants. This was however observed to be largely predicted by the purposes of the laboratories, and the goals of the practitioners. An academic who had just finished a project dealing with learning analytics and monitoring of the physical processes of the laboratory focused on the processes aspect of the PIPOV Framework rather than assessment outcomes. An academic interested in veracity focused on how students' cognitive responses might be derived from the inputs to a laboratory such as the interface. An academic seeking to increase students-as-partners engagement focused on the way different stakeholders might influence the purpose and thus design of a laboratory.

This inconsistency prevented the speculated study on what the impact of a given minimum viable methodology subset of the Purpose-First Approach and PIPOV might have on the design outcomes, but was confidence-raising when contrasted with Constructive Alignment. Because Constructive Alignment is so recognisable by the 3 connecting elements diagram and the idea of stating learning objectives, it has simultaneously shown an ability to improve designs simply through those elements, but has also shown an ability to be co-opted to the detriment of teaching and learning quality, as described in [173]. By not having a consistent anchor, and thus no common subset to be reduced to, it may be that the Purpose-First Approach and PIPOV Framework are more resilient against such misaligned use.

An observation that has been repeatedly made across multiple contexts – several expert reviews, and all 3 of the larger scale studies – was that opening up the laboratory to examination during its run for research and design activity rather than the black-box evaluations some practitioners appeared to expect led to a surge in ideas incorporating the application of learning analytics for direct intervention as an ongoing practice. While there is nothing preventing the use of mid-class interventions in black-box quantitative research, so long as it is prepared ahead of time as part of the research protocol, there was an aspect to the idea of highlighting physical and especially cognitive processes as something that can actually be poked, prodded, and otherwise studied rather than being something:

> "… Almost magical that happens between the lab guide being handed to the student and the assessment being handed in."

As a participant in the existing laboratory enhancement field test stated.

The nature of this connection is not immediately apparent. It was a specific target of modification in the framework iterations, and so in this sense this design decision has been confirmed. But a consideration of context suggests that perhaps this is associated with the increase in learning analytics as a feature of engineering education research and practice, rather than something that was drawn out only by the application of the artifacts. Noting that both the new laboratory design and existing laboratory enhancement field test laboratories both had either pre-existing or planned learning analytics components prior to the influence of the artifacts. A late speculation added to this is a consideration of the influence of generative AI – there is an opportunity moving forward to test this theory by seeing if there is a shift towards generative AI rather than learning analytics as an aspect of focus.

Another finding that occurred across multiple studies was a consideration of the ‘end state’ of a laboratory design. The different empirical studies were not long enough to reach that state within the scope of this thesis – and given the innovative nature of the laboratories involved is likely to take significantly longer than any ordinary laboratory. Questions were however asked, first in the expert reviews as described, and subsequently in the deeper studies, as to what an ‘end state’ looked like.

Discussing this with the participants and considering the nature of iterative design processes and the input of material from the literature review, the consensus was aligned with the idea of learning analytics, though not necessarily leveraging this. During the run of the different iterations of the laboratory, observation methods would need to be developed to measure the different attributes of interest and gauge whether the design was working. Upon establishing that something was working as a conclusion to those observations, this method of observation may be a ‘finger on the pulse’ for the laboratory. If context variables out of the control of the laboratory designer changed, e.g. if students began entering the class with different pre-existing knowledge, then this pulse observation would be likely to change.

This presents a strategy to maximise educator efficiency with laboratories that are ‘finished’ as was claimed in the expert reviews. At some point the laboratory has satisficed the requirements, and the attention of the educator is moving on to other fires or other opportunities as the case may be. In order to deal with the possibility that these contexts are not fixed, and may change from semester to semester, it is necessary to develop baseline monitoring, which is unintrusive to the students and time-effective for the educators. As with many aspects of the artifacts, this was recognised in existing practice by one of the academics in the collaboration study – describing monitoring through conversations with lab demonstrators and technicians to see how classes were progressing and how students were going. This had in the past discovered context changes, such as a change in equipment software caused by an automatic vendor update had led to differences in experiment results.

The distinction between this existing practice and the processes that can be derived from the Purpose-First Approach is based on intentionality. The observations can be based on the observations that led to the design decisions in the first place, and so provide a direct connection to the aspects in the laboratory that are likely to influence them. They are also observations where the previous impact of an identified issue, and thus some idea of the size of the change indicating a problem, are also known from prior experience. Finally, they are explicit, and recorded as part of the iterations of the laboratory itself, meaning it is not a task that can be as easily overlooked.

9.1 Educational Design Research & Reflective Practice – breaking down silos

A potential challenge to the novel and substantive nature of the Purpose-First Approach, the PIPOV Framework, and Educational Design Research more broadly is the observed similarities between EDR and what might be described as a normal reflective practice. These similarities are not accidental, they arise out of similar motives and serve similar purposes. There are further similarities between engineering discipline practice also, which shares the iterative reflective nature of both EDR and reflective practice.

Like many of the terms used in this field that have existing layman’s definitions and then various research-specific definitions, ‘reflective practice’ is a heavily over-loaded term. It generally involves the active consideration of the holistic intentions, experiences, and observations of the practitioner to recognise strengths and weaknesses, and to use that recognition to move towards some improvement or ideal. It is in fact not separable from Educational Design Research – indeed

McKenney & Reeves in [212] identify reflective practice as a 'core process' of Educational Design Research. But a core process, rather than the entirety of the concept. Educational Design Research influences both beyond and within the framework of reflective practice, by providing a structured approach to the iterative phases of educational design research activity, especially the interaction between contextual reflection and the research understanding that arises out of these. The Purpose-First Approach gives an anchor to which the reflective practice must align, and the PIPOV Framework gives an internal structure to the reflections themselves. A potentially illustrative analogy for this is that Educational Design Research is to reflective practice as engineering is to design. The one contains the other, but the structure, definition and identification of rigour, and guidance through the process is provided by the former. The latter can also be contained in other ways – reflective practice can be conducted without the structure of EDR just as design might be conducted in an architectural or marketing context rather than engineering.

The similarities with engineering discipline practice also have concrete links with EDR and the artifacts of this thesis – and indeed is why this thesis work gravitated towards them as the most appropriate solutions. The earliest history of Educational Design Research could be characterised as a group of education researchers in mathematics and other non-engineering education disciplines taking *the practice of engineering design*, and applying it to education as an 'artifact to be engineered'. Indeed, a synonym in the design research list detailed in 2.6.5.4 is 'engineering research'. This was earlier identified as one of the key reasons from the literature that indicated practical inquiry methodologies, and EDR first among them, as having the best supporting evidence for adoption to address the overall goals of laboratory design and evaluation.

The question that arises, and which motivated the subsequent work of this thesis beyond this identification, was that if a) an engineering approach is the best way to design and evaluate laboratories in engineering education, and b) engineering educators are often deeply versed in engineering design approaches – why then is this approach scarce to the point of being non-extant in the literature on these laboratories? Even to the point of non-engineering design approaches being used to design and evaluate teaching and learning activities that leverage or teach engineering design approaches.

It is present in the motivation for the PIPOV framework, but it is worth calling out explicitly here: a key component of this work lies not in introducing something which is atomically novel, but rather the novelty lies in breaking down the artificial siloing which exists between these clearly inter-relatable domains of knowledge and practice. It appeared in the interactions with academics at every step of introducing EDR and later the supporting artifacts – various forms of "of course – this makes so much sense". The Purpose-First Approach and the PIPOV Framework do not seek to teach academics how to undertake engineering design; this requires significant curricula, learning support, and years of learning and practice. Fortunately, it doesn't have to, as engineering education practitioners already have this expertise. What was speculatively required and then this speculation validated through the introduction of first EDR by the author directly and later through the use of the PIPOV artifact, is that academics need to be informed and convinced that they can and should apply their *engineering discipline* knowledge to their laboratory design and evaluation, in both a teaching practice and educational research context.

10 CONTRIBUTIONS & CONCLUSIONS

10.1 RESEARCH QUESTIONS

Based on the literature review, synthesis, empirical studies, and discussions presented, the Purpose First Approach and accompanying PIPOV Framework are considered to be well-evidenced as an answer to the research question posed by this thesis:

> "How may laboratories in engineering education be effectively designed to meet the purposes their stakeholders require, and evaluated against these purposes?"

Laboratories in engineering education may be effectively designed and evaluated through the application of the Purpose-First Approach and the PIPOV Framework as a proxy for the application of an Educational Design Research approach.

The component sub-questions which were answered were:

1. How do the research methodologies used in evaluative research on laboratories in engineering education support or inhibit answering the questions this research poses?
2. What methodological alternative provides the appropriate characteristics to resolve the identified issues with experimental approaches in the design and evaluation of laboratories in engineering education?
3. Can Educational Design Research be successfully applied to the design and evaluation of laboratories in engineering education?
4. How can practitioners be supported in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?

4.1. Do the Purpose-First Approach and PIPOV Framework support practitioners in their application of Educational Design Research to the design and evaluation of laboratories in engineering education?

These sub-questions are considered variously below as they apply to different external contexts.

10.2 RESEARCH CONSEQUENCES

The progressive elaboration which is a hallmark of Educational Design Research combines partial answers which move towards an overall answer, without ever reaching a 'perfect' answer in the positivist sense. The consequences to knowledge and practice from the above case studies and the overall artifacts are here described against three categories.

- Consequences to the research design and iteration which feed into the design for the next case study

10.2.1 Consequences to the research design and iteration which feed into the design for the next case study

This category is included for completeness, but is instead given treatment in chapter 5, as well as in each subsequent phase's chapter. The collective outcomes and residual uncertainty from each of these consequences is the source of the consequences in the following two summative categories.

10.2.2 Consequences to future work in research

This category might have two interpretations – the consequences for work attempting to answer similar questions to this thesis itself are detailed in 10.5. This category is focused on the research on laboratories in engineering education of the sort that forms the body of literature for the literature review and Critical Review. Here the recommendations are clearly described through the conclusions of those reviews, and the EDR and artifacts used in this thesis – research on laboratories in engineering education needs to move beyond the widely used methodologies to more appropriate approaches, in order to be capable of answering their intended research questions with internal and external validity. As later stated in 10.4, this is not limited only to laboratories, but rather much of the current research work, particularly that which focuses on close-context teaching and learning activities.

This draws primarily on the earlier components of this thesis work, specifically the literature synthesis. It appears reasonable to require a working knowledge of EDR to be held by researchers undertaking formal study in this area, such that the nuances of different traditions can be paired with the specific questions under study. The limitations on adoption and scalability that the following thesis work addressed does not apply to this category.

In linking in this out-of-discipline body of literature to the engineering education body of literature, and identifying the affordances of this work which is relevant to the questions currently being asked in the literature, the consequences for researchers are highlighted in the need to leverage practical inquiry approaches as an answer to validity concerns and improve research quality, replicability, and effectiveness.

Research contortion is explicitly called out as a potential flaw in mainly but not uniquely in experimental comparative research designs, which can be interpreted to destroy external validity by directing the research to an invalid, contorted representation of reality. It is not sufficient or beneficial for research to hold the control of nuisance or confounding variables as the primary goal when the variable of 'was this a research project or not' becomes an orders of magnitude greater confounding variable. Again, use of alternative approaches that do not seek this control and instead triangulate is the observed consequence to research.

10.2.3 Consequences to future work in laboratory design and evaluation practice

The consequences to practice are more specific, and are the focus of the empirical studies and the artifacts these leverage. Educational Design Research has been shown in the literature to provide valid and reliable design improvements through iterative evaluation and feedback. EDR has been empirically shown as able to be applied to the design and evaluation of laboratories in engineering education. And finally, the use of the Purpose-First Approach and the PIPOV Framework has been shown to allow practitioners without a foundation in EDR to leverage EDR techniques when designing and evaluating their laboratories.

Thus, the consequences for practice are for engineering educators to use the Purpose-First Approach and PIPOV Framework to conduct their laboratory design and evaluation, and to avoid research contortions and other sources of internal invalidity that will inhibit their ability to conduct this design and evaluation – even if external validity for dissemination is not a concern. There is no need to develop new skills from outside engineering practice; the key consequence is to recognise that existing engineering design skills are valid methods to use, and that artifacts exist to support that use.

10.3 Research Contributions

Calling back to the opening section 1.3, this thesis presents the following contributions to knowledge:

- Literature synthesis drawing knowledge from multiple external disciplines and collecting this into a coherent body of knowledge on laboratories in engineering education and how they might be designed and evaluated.
- A systematic review on practical inquiry methods applied to laboratories in engineering education, section 2.6.5.4
- Identification and definition of the 'research contortion' effect on research in engineering education, section 3.2.
- Describing and defining the Educational Design Research methodology through a structure of named research traditions and distinguishing literature in order to support its investigation in spite of the mixed terminology issues.
- The Purpose-First Approach and PIPOV Framework as artifacts supporting the design and evaluation of laboratories in engineering education.
- A lens for describing the scope of what is considered a laboratory teaching and learning activity.
- Design decisions made in the process of designing and refining the artifacts, and the literature and empirical observations which motivated this as an Educational Design Research output.
- Identification of white-boxed design and evaluation of teaching and learning activities incorporating mid-way observation and intervention via learning analytics as a critical opportunity in engineering education laboratory practice

10.4 Scope

It became clear to the author early in the empirical studies portion of this thesis work that the output of the synthesis work, i.e. the Purpose-First Framework and PIPOV Framework, are not actually unique to laboratories. There is no conceivable barrier to their application to other teaching and learning activities, and in other disciplines with an existing issue with misapplied quantitative experimentation in education. The one aspect which does tie laboratories to the artifacts is the focus on the teaching and learning activities in isolation – laboratories in engineering education are often considered with a level of independence from the wider context due to their increased time and cost requirements, and the need for different rooms, equipment etc. from other teaching and learning. This was a weakness for laboratories in that they *should* be considered in context with lectures, tutorials, self-directed work, etc. – but does tie into the artifacts which take a systems analysis approach rather than e.g. frameworks looking at learner trajectories over multiple sessions, or other more integrated lenses on educational practice.

10.5 Future Work

There is significant future work to be done in this space, with the scope of the issue large enough that any single project can only start a movement in the right direction. The previously described stopping point for the iterations of this thesis work leads to the most obvious next step: use of the generated artifacts by practitioners in engineering education designing and evaluating their laboratories without direct involvement by the author. This will hopefully lead to continued adoption, refinement, and collaboration leading to improved outcomes for laboratories overall.

Another understudied aspect is the quantitative aspects of laboratory design and evaluation. This is not to say run experimental designs, but the data used for empirical observations in this thesis was largely the author being embedded in the laboratory design and evaluation processes, contributing to discussions, taking notes and journals, and various interview sessions, with relatively small numbers of participants. There are quantitative aspects of interest such as detailed timings of different processes, but more interesting would be collective quantitative data of a large sample of participants, allowing for triangulation of the qualitative data collected between them against data such as the number of iterations before stability, types and number of observational methods and instruments used, or coding sizes of design decisions from existential to structural to incremental tweaks and plotting their occurrence over time.

There is a need to conduct an analysis of decision-making *across* methodologies rather than just using the Purpose-First Approach and PIPOV – including no methodology at all, i.e. unstudied informal laboratory designs. This type of laboratory design can be considered another form of dark data, as it is work being done and observations being made which is not reported in the literature.

Related to this is an opportunity to examine the contextual pressures on practitioners enabling or inhibiting formal design and evaluation work at all. Time, scope, space, institutional culture – any number of influences may determine engagement in formal design and evaluation, and lowering these barriers may be even more effective than refining the designed artifacts could be.

There is an aspect of laboratory research related to informal design and evaluation, and contextual pressures, revealed as a disclosed observation by a participant during the expert reviews. This observation was around practices of laboratory research which might preclude the adoption of the Purpose-First Approach and PIPOV Framework by practitioners who practiced post-hoc evaluation research. This is a research practice described as following an informal design practice, developing a laboratory based on prior experience, gut feel, interest, opportunity with particular equipment, etc., but then seeking empirical evidence specifically in support of this pre-established conclusion.

Two motivations for this were observed, one being the time pressures inhibiting deliberate reflective evaluation design and implementation, and the other being the dissonance between the design data being informally collected and the practitioner's expectation of what 'good' research is supposed to look like. The consequences of this style of research are significant, with implicit p-hacking being the most benign that could be expected. The first motivation is part of the future work identified above, but the second speaks in favour of the adoption of the Purpose-First Approach and the PIPOV Framework, to allow a structured outlet for the design work already being done.

The ultimate future work would of course be an evaluation of students undertaking laboratories designed and evaluated using these methods, showing their continued improvement over time, but is likely dependent on the completion of the above work and more in establishing large scale adoption and experience with the artifacts.

10.6 Final words

The discipline of engineering education is dealing with new challenges on an ever-increasing basis. Exploding student numbers in some areas, shrinking in others, the globalisation of education, COVID, generative AI. The need for quality teaching and learning activities to help meet these challenges is clear, but the ability of current research practices to achieve this quality is not. The siloed nature of engineering education as a research discipline has hindered its potential progress compared to more integrated and mature disciplines like nursing education, and limited the benefits of significant research effort. The answers to many of the questions we are asking, or support in investigating them, are already available for the taking. If there is a meta-conclusion to this research thesis, it is to take this lesson of importing knowledge from these other disciplines, rather than staying within the silo walls.

REFERENCES

[1] J. R. Brinson, "Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research," *Comput Educ*, vol. 87, pp. 218–237, Sep. 2015, doi: 10.1016/j.compedu.2015.07.003.

[2] ABET Engineering Accreditation Commission, "Criteria for accrediting engineering programs 2018-2019," 2017, *Baltimore, Maryland, USA*. [Online]. Available: http://www.abet.org/wp-content/uploads/2017/12/E001-18-19-EAC-Criteria-11-29-17-FINAL.pdf

[3] P. Martinez-Jimenez, L. Salas-Morera, G. Pedros-Perez, A. J. Cubero-Atienza, and M. Varo-Martinez, "OPEE: An Outreach Project for Engineering Education," *IEEE Transactions on Education*, vol. 53, no. 1, pp. 96–104, Feb. 2010, doi: 10.1109/TE.2009.2024931.

[4] G. S. Coutinho, N. A. Stites, and A. J. Magana, "Understanding faculty decisions about the integration of laboratories into engineering education," in *2017 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2017, pp. 1–9. doi: 10.1109/FIE.2017.8190605.

[5] J. Pleasants and J. K. Olson, "What is engineering? Elaborating the nature of engineering for K-12 education," *Sci Educ*, vol. 103, no. 1, pp. 145–166, Jan. 2019, doi: 10.1002/sce.21483.

[6] L. D. Feisel and A. J. Rosa, "The Role of the Laboratory in Undergraduate Engineering Education," *Journal of Engineering Education*, vol. 94, no. 1, pp. 121–130, Jan. 2005, doi: 10.1002/j.2168-9830.2005.tb00833.x.

[7] I. Newton, *The Correspondence of Isaac Newton*, vol. 192, no. 21. Cambridge University Press, 1959. doi: 10.1017/9781108699945.

[8] W. E. Rowe, "Positionality," *The Sage Encyclopaedia of Action Research*. Sage, 2014.

[9] C. Hampton, D. Reeping, and D. S. Ozkan, "Positionality Statements in Engineering Education Research: A Look at the Hand that Guides the Methodological Tools," *Studies in Engineering Education*, vol. 1, no. 2, p. 126, Mar. 2021, doi: 10.21061/see.13.

[10] L. D. Feisel and G. D. Peterson, "A Colloquy on Learning Objectives for Engineering Educational Laboratories," in *Proceedings of the 2002 American Society for Engineering Education Annual Conference and Exposition*, 2002.

[11] F. P. . Whitman, "The Beginnings of Laboratory Teaching in America," *Science (1979)*, vol. 8, no. 190, pp. 201–206, 1898.

[12] W. W. Knox, "The Demonstration Method versus the Laboratory Method of Teaching High-School Chemistry," *The School Review*, vol. 35, no. 5, pp. 376–386, 1927.

[13] J. S. Peck, "The Function of the Laboratory in Engineering Education," Columbia University, 1936.

[14] D. Comley, "Some Aims Of Laboratory Work," *Education + Training*, vol. 4, no. 5, pp. 28–29, May 1962, doi: 10.1108/eb015126.

[15] J. Ma and J. V. Nickerson, "Hands-on, simulated, and remote laboratories," *ACM Comput Surv*, vol. 38, no. 3, pp. 7-es, Sep. 2006, doi: 10.1145/1132960.1132961.

[16] S. Nikolic, M. Ros, K. Jovanovic, and Z. Stanisavljevic, "Remote, simulation or traditional engineering teaching laboratory: a systematic literature review of assessment implementations to measure student achievement or learning," *European Journal of Engineering Education*, vol. 46, no. 6, pp. 1141–1162, Nov. 2021, doi: 10.1080/03043797.2021.1990864.

[17] R. Raman, K. Achuthan, V. K. Nair, and P. Nedungadi, "Virtual Laboratories- A historical review and bibliometric analysis of the past three decades," *Educ Inf Technol (Dordr)*, vol. 27, no. 8, pp. 11055–11087, Sep. 2022, doi: 10.1007/s10639-022-11058-9.

[18] A. Hofstein and V. N. Lunetta, "The Role of the Laboratory in Science Teaching: Neglected Aspects of Research," *Rev Educ Res*, vol. 52, no. 2, pp. 201–217, Jan. 1982, doi: 10.3102/00346543052002201.

[19] A. Hofstein and V. N. Lunetta, "The laboratory in science education: Foundations for the twenty-first century," *Sci Educ*, vol. 88, no. 1, pp. 28–54, Jan. 2004, doi: 10.1002/sce.10106.

[20] J. Ma and J. V. Nickerson, "Hands-on, simulated, and remote laboratories," *ACM Comput Surv*, vol. 38, no. 3, pp. 7-es, Sep. 2006, doi: 10.1145/1132960.1132961.

[21] J. R. Brinson, "Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research," *Comput Educ*, vol. 87, pp. 218–237, Sep. 2015, doi: 10.1016/j.compedu.2015.07.003.

[22] R. E. Machol, "The Computer Revolution in Engineering Education," *IRE Transactions on Education*, vol. 4, no. 2, pp. 67–70, 1961, doi: 10.1109/TE.1961.4322187.

[23] J. A. Chambers and J. W. Sprecher, "Computer assisted instruction," *Commun ACM*, vol. 23, no. 6, pp. 332–342, Jun. 1980, doi: 10.1145/358876.358881.

[24] M. Budhu, "VIRTUAL LABORATORIES FOR ENGINEERING EDUCATION," *International Conference on Engineering Education*, Mar. 2002.

[25] S. K. Ong and M. A. Mannan, "Virtual reality simulations and animations in a web-based interactive manufacturing engineering module," *Comput Educ*, vol. 43, no. 4, pp. 361–382, Dec. 2004, doi: 10.1016/j.compedu.2003.12.001.

[26] B. B. Aktan, C. A. Bohus, L. A. Crowl, and M. H. Shor, "Distance learning applied to control engineering laboratories," *IEEE Transactions on Education*, vol. 39, no. 3, pp. 320–326, 1996, doi: 10.1109/13.538754.

[27] P. Orduna *et al.*, "Classifying online laboratories: Reality, simulation, user perception and potential overlaps," in *2016 13th International Conference on Remote Engineering and Virtual Instrumentation (REV)*, IEEE, Feb. 2016, pp. 224–230. doi: 10.1109/REV.2016.7444469.

[28] P. J. Ogren and T. P. Jones, "Laboratory Interfacing Using the LabVIEW Software Package," *J Chem Educ*, vol. 73, no. 12, p. 1115, Dec. 1996, doi: 10.1021/ed073p1115.

[29] V. Potkonjak *et al.*, "Virtual laboratories for education in science, technology, and engineering: A review," *Comput Educ*, vol. 95, pp. 309–327, 2016, doi: 10.1016/j.compedu.2016.02.002.

[30] M. P. Strzys *et al.*, "Physics holo.lab learning experience: Using smartglasses for augmented reality labwork to foster the concepts of heat conduction," *Eur J Phys*, vol. 39, no. 3, Mar. 2018, doi: 10.1088/1361-6404/aaa8fb.

[31] K. Bowen and M. O'Malley, "Haptic interfaces for a labview-based system dynamics course," *ASEE Annual Conference and Exposition, Conference Proceedings*, pp. 147–152, 2006.

[32] Z. Nedic, "Demonstration of Collaborative Features of Remote Laboratory NetLab," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 9, no. S1, p. 10, Jan. 2013, doi: 10.3991/ijoe.v9iS1.2368.

[33] A. Srivastava and P. Yammiyavar, "Augmenting tutoring of students using Tangible Smart Learning Objects: An IOT based approach to assist student learning in laboratories," in *2016 International Conference on Internet of Things and Applications (IOTA)*, IEEE, Jan. 2016, pp. 424–426. doi: 10.1109/IOTA.2016.7562765.

[34] P. J. Grise., K. I. Epstein, and T. G. Lukin, "Educational Radio: A Review of the Literature," pp. 1–51, 1974, [Online]. Available: https://files.eric.ed.gov/fulltext/ED095904.pdf

[35] A. W. Bates, "Television, Learning and Distance Education," *Journal of Educational Television*, vol. 14, no. 3, pp. 213–225, Jan. 1988, doi: 10.1080/0260741880140305.

[36] J. DeBoer, C. Haney, S. Z. Atiq, C. Smith, and D. Cox, "Hands-on engagement online: using a randomised control trial to estimate the impact of an at-home lab kit on student attitudes and achievement in a MOOC," *European Journal of Engineering Education*, vol. 44, no. 1–2, pp. 234–252, Mar. 2019, doi: 10.1080/03043797.2017.1378170.

[37] J. V. Nickerson, J. E. Corter, S. K. Esche, and C. Chassapis, "A model for evaluating the effectiveness of remote engineering laboratories and simulations in education," *Comput Educ*, vol. 49, no. 3, pp. 708–725, Nov. 2007, doi: 10.1016/j.compedu.2005.11.019.

[38] E. Lindsay and M. C. Good, "Effects of laboratory access modes upon learning outcomes," *IEEE Transactions on Education*, vol. 48, no. 4, pp. 619–631, Nov. 2005, doi: 10.1109/TE.2005.852591.

[39] J. Burridge, "The Evaluation of Practical Laboratories in Engineering and Science Education," The University of Sydney, 2018.

[40] Sahara Labs, "Sahara Labs GitHub Repository." Accessed: Jun. 21, 2018. [Online]. Available: https://github.com/sahara-labs

[41] M. G. Moore, "Editorial: Three types of interaction," *American Journal of Distance Education*, vol. 3, no. 2, pp. 1–7, Jan. 1989, doi: 10.1080/08923648909526659.

[42] Z. Nedic and J. Machotka, "Remote Laboratory NetLab for Effective Teaching of 1 st Year Engineering Students *," *International Journal of Online Engineering*, pp. 1–6, 2007.

[43] Y. Chang, E.-S. S. Aziz, Z. Zhang, M. Zhang, and S. K. Esche, "Usability Evaluation of a Virtual Educational Laboratory Platform," in *ASEE Annual Conference and Exposition, Conference Proceedings*, American Society for Engineering Education, Jun. 2015, pp. 24–36. doi: 10.18260/p.24973.

[44] Y. Chang, E.-S. Aziz, Z. Zhang, M. Zhang, and S. K. Esche, "Evaluation of a video game adaptation for mechanical engineering educational laboratories," in *2016 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2016, pp. 1–6. doi: 10.1109/FIE.2016.7757670.

[45] L. Foreman-Peck and C. Winch, *Using Educational Research to Inform Practice*. Routledge, 2010. doi: 10.4324/9780203847619.

[46] J. Burridge and D. Lowe, "Breaking down the Laboratory Supertype Conflation," in *2018 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2018, pp. 1–6. doi: 10.1109/FIE.2018.8658725.

[47] S. Nikolic, M. Ros, K. Jovanovic, and Z. Stanisavljevic, "Remote, simulation or traditional engineering teaching laboratory: a systematic literature review of assessment implementations to measure student achievement or learning," *European Journal of Engineering Education*, vol. 46, no. 6, pp. 1141–1162, Nov. 2021, doi: 10.1080/03043797.2021.1990864.

[48] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 4th ed. Newbury Park: Sage, 2017.

[49] R. Millsap and A. Maydeu-Olivares, *The SAGE Handbook of Quantitative Methods in Psychology*. 2012. doi: 10.4135/9780857020994.

[50] W. Shadish and J. Luellen, "Quasi-Experimental Design," in *Handbook of Complementary Methods in Education Research*, 3rd ed., Routledge, 2006.

[51] D. L. STUFFLEBEAM, "THE USE OF EXPERIMENTAL DESIGN IN EDUCATIONAL EVALUATION," *J Educ Meas*, vol. 8, no. 4, pp. 267–274, Dec. 1971, doi: 10.1111/j.1745-3984.1971.tb00936.x.

[52] P. Dugard and J. Todman, "Analysis of Pre-test-Post-test Control Group Designs in Educational Research," *Educ Psychol (Lond)*, vol. 15, no. 2, pp. 181–198, Jan. 1995, doi: 10.1080/0144341950150207.

[53] D. Campbell and J. Stanley, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH*. Dallas: HOUGHTON MIFFLIN COMPANY, 1963.

[54] M. B. Brewer and W. D. Crano, "Research Design and Issues of Validity," *Handbook of Research Methods in Social and Personality Psychology*, pp. 11–26, 2014, doi: 10.1017/cbo9780511996481.005.

[55] C. H. Rhoads, "The Implications of 'Contamination' for Experimental Design in Education," *Journal of Educational and Behavioral Statistics*, vol. 36, no. 1, pp. 76–104, Feb. 2011, doi: 10.3102/1076998610379133.

[56] E. Marsden and C. J. Torgerson, "Single group, pre- and post-test research designs: Some methodological concerns," *Oxf Rev Educ*, vol. 38, no. 5, pp. 583–616, Oct. 2012, doi: 10.1080/03054985.2012.731208.

[57] J. Rogers and A. Révész, "Experimental and quasi-experimental designs," in *The Routledge Handbook of Research Methods in Applied Linguistics*, Routledge, 2019.

[58] G. M. Sullivan, "Getting Off the 'Gold Standard': Randomized Controlled Trials and Education Research," *J Grad Med Educ*, vol. 3, no. 3, pp. 285–289, Sep. 2011, doi: 10.4300/JGME-D-11-00147.1.

[59] T. C. Reeves, “Design research from a technology perspective,” *Educational Design Research*, pp. 52–66, 2006, doi: 10.4324/9780203088364-13.

[60] E. Salas Blas, “Pre-experimental designs in psychology and education: a conceptual review,” *Liberabit*, vol. 19, no. 1, pp. 133–141, 2013.

[61] A. Morris and R. Langari, *Measurement and Instrumentation*. Elsevier, 2016. doi: 10.1016/C2013-0-15387-1.

[62] J. C. Stanley, “A Common Class of Pseudo-Experiments,” *Am Educ Res J*, vol. 3, no. 2, pp. 79–87, Mar. 1966, doi: 10.3102/00028312003002079.

[63] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[64] J. P. T. Higgins, S. Green, and A. Ben Van Den, “Cochrane Handbook for Systematic Reviews of Interventions,” *International Coaching Psychology Review*, vol. 15, no. 2, pp. 123–125, 2020, doi: 10.53841/bpsicpr.2020.15.2.123.

[65] M. Borrego, “Conceptual difficulties experienced by trained engineers learning educational research methods,” *Journal of Engineering Education*, vol. 96, no. 2, pp. 91–102, Apr. 2007, doi: 10.1002/j.2168-9830.2007.tb00920.x.

[66] D. KEMBER, “To Control or Not to Control: The question of whether experimental designs are appropriate for evaluating teaching innovations in higher education,” *Assess Eval High Educ*, vol. 28, no. 1, pp. 89–101, Jan. 2003, doi: 10.1080/02602930301684.

[67] G. Thomas, “Experiment’s persistent failure in education inquiry, and why it keeps failing,” *Br Educ Res J*, vol. 47, no. 3, pp. 501–519, Jun. 2021, doi: 10.1002/berj.3660.

[68] D. W. Surry and D. Ensminger, “What’s Wrong with Media Comparison Studies ?,” *Educational Technology*, no. August, pp. 32–35, 2001.

[69] Y. Rogers, N. Yuill, and P. Marshall, “Contrasting Lab-Based and in-the-Wild Studies for Evaluating Multi-User Technologies,” in *The SAGE Handbook of Digital Technology Research*, 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, 2013, pp. 359–373. doi: 10.4135/9781446282229.n24.

[70] J. Lazar, J. H. Feng, and H. Hochheiser, “Research Methods in Human-Computer Interaction,” *Research Methods in Human-Computer Interaction*, pp. 1–560, 2017, doi: 10.1016/b978-0-444-70536-5.50047-6.

[71] J. Pisapia, J. Schlesinger, and A. Parks, “Learning Technologies in the Classroom: Review of the Literature,” 1993.

[72] TecQuipment Academia, “Redundant Truss STR17.” Accessed: Jun. 13, 2018. [Online]. Available: https://www.tecquipment.com/redundant-truss

[73] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, “G*Power: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.,” *Behav Res Methods*, vol. 39, no. 2, pp. 175–191, 2007, doi: 10.3758/BF03193146.

[74] J. Cohen, "A power primer," *Psychol Bull*, vol. 112, no. 1, pp. 155–159, 1992, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-11944272254&partnerID=40&md5=20438740c8abff73cbc85727d262ef29

[75] J. Burridge and J. Kay, "Dark Data and Learning Analytics," in *Australian Learning Analytics Summer Institute Proceedings*, 2019.

[76] J. Burridge, D. Lowe, K. Willey, and J. Kay, "Defeating Hawthorne in tech-enabled education : Passive observation of student behaviour with a remote laboratory The Laboratory System," *AAEE2019: 30th Annual Conference of the Australasian Association for Engineering Education*, 2019.

[77] L. B. Buck, M. Towns, and S. Lowery Bretz, "Characterizing the Level of Inquiry in the Undergraduate Laboratory," *J Coll Sci Teach*, vol. 38, no. 01, pp. 52–58, 2008.

[78] G. L. Salinger, T. F. of America, and C. for T. E. R. G. University, "Engineering in the K-12 Curriculum," in *Initiative in Technology Education: Comparative Perspectives*, 2003.

[79] B. Jane and R. Tytler, "TECHNOLOGY EDUCATION, SCIENCE, AND SCIENCE EDUCATION: EXPLORING THE RELATIONSHIP," in *Initiative in Technology Education: Comparative Perspectives*, 2003.

[80] N. Gericke, P. Högström, and J. Wallin, "A systematic review of research on laboratory work in secondary school," *Stud Sci Educ*, vol. 59, no. 2, pp. 245–285, Jul. 2023, doi: 10.1080/03057267.2022.2090125.

[81] P. R. Jeffries, "Designing simulations for nursing education," *Annual Review of Nursing Education*, vol. 4, pp. 161–177, 2006.

[82] S. L. Hanshaw and S. S. Dickerson, "High fidelity simulation evaluation studies in nursing education: A review of the literature," *Nurse Educ Pract*, vol. 46, p. 102818, Jul. 2020, doi: 10.1016/j.nepr.2020.102818.

[83] S. Prion, "A Practical Framework for Evaluating the Impact of Clinical Simulation Experiences in Prelicensure Nursing Education," *Clin Simul Nurs*, vol. 4, no. 3, pp. e69–e78, Oct. 2008, doi: 10.1016/j.ecns.2008.08.002.

[84] A. W. Astin, "Educational Assessment and Educational Equity," *American Journal of Education*, vol. 98, no. 4, pp. 458–478, Aug. 1990, doi: 10.1086/443972.

[85] C. Foronda, S. Liu, and E. B. Bauman, "Evaluation of Simulation in Undergraduate Nurse Education: An Integrative Review," *Clin Simul Nurs*, vol. 9, no. 10, pp. e409–e416, Oct. 2013, doi: 10.1016/j.ecns.2012.11.003.

[86] A. Frandsen and S. Lehn-Christiansen, "Into the black-box of learning in simulation debriefing: A qualitative research study," *Nurse Educ Today*, vol. 88, p. 104373, May 2020, doi: 10.1016/j.nedt.2020.104373.

[87] K. Illeris, "A comprehensive understanding of human learning," in *Contemporary Theories of Learning*, 2018.

[88] J.-M. Koivisto *et al.*, "Design-based research in designing the model for educating simulation facilitators," *Nurse Educ Pract*, vol. 29, pp. 206–211, Mar. 2018, doi: 10.1016/j.nepr.2018.02.002.

[89] S. Barab and K. Squire, "Design-Based Research: Putting a Stake in the Ground," *Journal of the Learning Sciences*, vol. 13, no. 1, pp. 1–14, Jan. 2004, doi: 10.1207/s15327809jls1301_1.

[90] T. Plomp, "Educational design research: An introduction," in *Educational Design Research Part A: An Introduction*, 2013, pp. 11–50.

[91] R. Macdiarmid, R. Winnington, T. Cochrane, and E. Merrick, "Using educational design research to develop authentic learning for graduate entry nursing students in New Zealand," *Nurse Educ Pract*, vol. 51, p. 102965, Feb. 2021, doi: 10.1016/j.nepr.2021.102965.

[92] S. McKenney and T. C. Reeves, *Conducting Educational Design Research*. Second edition. | New York : Routledge, 2019. | "[First edition: Routledge, 2018. doi: 10.4324/9781315105642.

[93] J. Creswell, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, 3rd ed. SAGE Publications, Inc, 2012.

[94] M. Sandelowski, C. I. Voils, and G. Knafl, "On Quantitizing," *J Mix Methods Res*, vol. 3, no. 3, pp. 208–222, Jul. 2009, doi: 10.1177/1558689809334210.

[95] V. Clarke and V. Braun, "Thematic analysis," *J Posit Psychol*, vol. 12, no. 3, pp. 297–298, May 2017, doi: 10.1080/17439760.2016.1262613.

[96] D. W. Shaffer, "Quantitative ethnography," no. April, p. 473, 2017.

[97] M. T. Anguera, A. Blanco-Villaseñor, J. L. Losada, P. Sánchez-Algarra, and A. J. Onwuegbuzie, "Revisiting the difference between mixed methods and multimethods: Is it all in the name?," *Qual Quant*, vol. 52, no. 6, pp. 2757–2770, 2018, doi: 10.1007/s11135-018-0700-2.

[98] R. Vivek and Y. Nanthagopan, "Review and Comparison of Multi-Method and Mixed Method Application in Research Studies," *European Journal of Management Issues*, vol. 29, no. 4, pp. 200–208, Dec. 2021, doi: 10.15421/192119.

[99] R. B. Johnson, A. J. Onwuegbuzie, and L. A. Turner, "Toward a Definition of Mixed Methods Research," *J Mix Methods Res*, vol. 1, no. 2, pp. 112–133, Apr. 2007, doi: 10.1177/1558689806298224.

[100] R. B. Johnson and A. J. Onwuegbuzie, "Mixed Methods Research: A Research Paradigm Whose Time Has Come," *Educational Researcher*, vol. 33, no. 7, pp. 14–26, Oct. 2004, doi: 10.3102/0013189X033007014.

[101] A. Bakker, *Design Research in Education*. Routledge, 2018. doi: 10.4324/9780203701010.

[102] A. Minichiello and L. Caldwell, "A Narrative Review of Design-Based Research in Engineering Education: Opportunities and Challenges," *Studies in Engineering Education*, vol. 1, no. 2, p. 31, Feb. 2021, doi: 10.21061/see.15.

[103] Sage Research, "Applied Research," Methods Map. [Online]. Available: https://methods.sagepub.com/methods-map/applied-research

[104] Sage Research, "Action Research," Methods Map. [Online]. Available: https://methods.sagepub.com/methods-map/action-research

[105] S. Kemmis and R. McTaggart, *The Action research planner*. Waurn Ponds, Vic: Deakin University : distributed by Deakin University Press, 1988.

[106] H. Altrichter, S. Kemmis, R. McTaggart, and O. Zuber-Skerritt, “The concept of action research,” *The Learning Organization*, vol. 9, no. 3, pp. 125–131, Aug. 2002, doi: 10.1108/09696470210428840.

[107] N. Denzin, Y. Lincoln, and L. Smith, Eds., *Handbook of Critical and Indigenous Methodologies*. 2008.

[108] P. Cobb, J. Confrey, A. DiSessa, R. Lehrer, and L. Schauble, “Design Experiments in Educational Research,” *Educational Researcher*, vol. 32, no. 1, pp. 9–13, Jan. 2003, doi: 10.3102/0013189X032001009.

[109] J. Van Den Akker, K. Gravemeijer, S. McKenney, and N. Nieveen, *Educational Design Research*. Routledge, 2006.

[110] T. M. Kister, “Improving the information development process: A refined iterative development model,” *Tech Commun*, vol. 63, no. 3, pp. 186–211, 2016.

[111] C. E. Lindblom, “The Science of ‘Muddling Through,’” *Public Adm Rev*, vol. 19, no. 2, p. 79, 1959, doi: 10.2307/973677.

[112] K. J. Vicente, C. M. Burns, and W. S. Pawlak, “Muddling Through Wicked Design Problems,” *Ergonomics in Design: The Quarterly of Human Factors Applications*, vol. 5, no. 1, pp. 25–30, Jan. 1997, doi: 10.1177/106480469700500105.

[113] C. Bereiter, “Design research for sustained innovation,” *Cognitive studies, bulletin of the Japanese cognitive …*, vol. 9, no. 3, pp. 321–327, 2002, doi: 10.11225/jcss.9.321.

[114] J. W. Creswell, *Qualitative Inquiry and Research Design: Choosing among Five Traditions*, 1st ed. SAGE Publications, Inc, 1997.

[115] K. Gravemeijer, “Educational Development and Developmental Research in Mathematics Education,” *J Res Math Educ*, vol. 25, no. 5, pp. 443–471, Nov. 1994, doi: 10.5951/jresematheduc.25.5.0443.

[116] T. Cochrane and J. Munn, “Integrating Educational Design Research and Design Thinking to Enable Creative Pedagogies,” *Pacific Journal of Technology Enhanced Learning*, vol. 2, no. 2, pp. 1–14, May 2020, doi: 10.24135/pjtel.v2i2.58.

[117] W. Sandoval, “Conjecture Mapping: An Approach to Systematic Educational Design Research,” *Journal of the Learning Sciences*, vol. 23, no. 1, pp. 18–36, Jan. 2014, doi: 10.1080/10508406.2013.778204.

[118] A. Collins, D. Joseph, and K. Bielaczyc, “Design Research: Theoretical and Methodological Issues,” *Journal of the Learning Sciences*, vol. 13, no. 1, pp. 15–42, Jan. 2004, doi: 10.1207/s15327809jls1301_2.

[119] J. van den Akker, “Principles and Methods of Development Research,” *Design Approaches and Tools in Education and Training*, pp. 1–14, 1999, doi: 10.1007/978-94-011-4255-7_1.

[120] J. C. Dipietro, “Leveraging Educational Design Research To Develop and Refine a Blended Professional Development Intervention,” George Mason University, 2015.

[121] H. G. Larsen and P. Adu, “The Theoretical Framework in Phenomenological Research: Development and Application,” *The Theoretical Framework in Phenomenological Research: Development and Application*, pp. 1–232, 2021, doi: 10.4324/9781003084259.

[122] J. Bernhard, “Insightful learning in the laboratory: Some experiences from 10 years of designing and using conceptual labs,” *European Journal of Engineering Education*, vol. 35, no. 3, pp. 271–287, Jun. 2010, doi: 10.1080/03043791003739759.

[123] J. Bernhard, “What matters for learning in labs? – Experiences from designing for insightful learning in labs based on a symbiosis of American and European thinking,” in *2020 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2020, pp. 1–9. doi: 10.1109/FIE44824.2020.9273812.

[124] H. W. Fennell, J. A. Lyon, A. J. Magana, S. Rebello, C. M. Rebello, and Y. B. Peidrahita, “Designing hybrid physics labs: combining simulation and experiment for teaching computational thinking in first-year engineering,” in *2019 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2019, pp. 1–8. doi: 10.1109/FIE43999.2019.9028390.

[125] R. R. Gutierrez *et al.*, “Combining hands-on and virtual experiments for enhancing fluid mechanics teaching: A design-based research study,” *Computer Applications in Engineering Education*, vol. 30, no. 6, pp. 1701–1724, Nov. 2022, doi: 10.1002/cae.22551.

[126] N. Rumler, S. Staude, and N. Friese, “Development of an Innovative Learning Environment for Engineering Education,” in *Advances in Intelligent Systems and Computing*, vol. 544, 2017, pp. 254–264. doi: 10.1007/978-3-319-50337-0_22.

[127] S. W. Tho and Y. Y. Yeung, “Remote Laboratory System for Technology-Enhanced Science Learning: The Design and Pilot Implementation in Undergraduate Courses,” in *Proceedings of the 22nd International Conference on Computers in Education*, 2014.

[128] A. L. Brown, “Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings,” *Journal of the Learning Sciences*, vol. 2, no. 2, pp. 141–178, Apr. 1992, doi: 10.1207/s15327809jls0202_2.

[129] E. Baumgartner *et al.*, “Design-Based Research: An Emerging Paradigm for Educational Inquiry,” *Educational Researcher*, vol. 32, no. 1, pp. 5–8, Jan. 2003, doi: 10.3102/0013189X032001005.

[130] A. Collins, “Toward a Design Science of Education,” in *New Directions in Educational Technology*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 15–22. doi: 10.1007/978-3-642-77750-9_2.

[131] A.-K. Carstensen and J. Bernhard, “Design science research – a powerful tool for improving methods in engineering education research,” *European Journal of Engineering Education*, vol. 44, no. 1–2, pp. 85–102, Mar. 2019, doi: 10.1080/03043797.2018.1498459.

[132] “CORE Conference Rankings: Frontiers in Education.” [Online]. Available: http://portal.core.edu.au/conf-ranks/2027/

[133] A. E. Kelly, “Design-Based Research in Engineering Education,” in *Cambridge Handbook of Engineering Education Research*, Cambridge University Press, 2014, pp. 497–518. doi: 10.1017/CBO9781139013451.032.

[134] S. McKenney and T. C. Reeves, "Educational design research: Portraying, conducting, and enhancing productive scholarship," *Med Educ*, vol. 55, no. 1, pp. 82–92, Jan. 2021, doi: 10.1111/medu.14280.

[135] M. Borrego, E. P. Douglas, and C. T. Amelink, "Quantitative, qualitative, and mixed research methods in engineering education," *Journal of Engineering Education*, vol. 98, no. 1, pp. 53–66, Jan. 2009, doi: 10.1002/j.2168-9830.2009.tb01005.x.

[136] A. Biglan, "The characteristics of subject matter in different academic areas.," *Journal of Applied Psychology*, vol. 57, no. 3, pp. 195–203, Jun. 1973, doi: 10.1037/h0034701.

[137] K. Kelly and B. Bowe, "Qualitative research in engineering education," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2011.

[138] B. Williams, P. C. Wankat, and P. Neto, "Not so global: a bibliometric look at engineering education research," *European Journal of Engineering Education*, vol. 43, no. 2, pp. 190–200, Mar. 2018, doi: 10.1080/03043797.2016.1153043.

[139] A. Valentine, "Do Australian Engineering Education Researchers Publish More Educational or Non-Educational Research?: A Bibliometric Analysis," in *31st Annual Conference of the Australasian Association for Engineering Education*, Engineers Australia, 2020.

[140] L. Malmi *et al.*, "How authors did it – a methodological analysis of recent engineering education research papers in the European Journal of Engineering Education," *European Journal of Engineering Education*, vol. 43, no. 2, pp. 171–189, Mar. 2018, doi: 10.1080/03043797.2016.1202905.

[141] T. J. Moore, K. M. Tank, A. W. Glancy, and J. A. Kersten, "NGSS and the landscape of engineering in K-12 state science standards," *J Res Sci Teach*, vol. 52, no. 3, pp. 296–318, Mar. 2015, doi: 10.1002/tea.21199.

[142] S. Brophy, S. Klein, M. Portsmore, and C. Rogers, "Advancing Engineering Education in P-12 Classrooms," *Journal of Engineering Education*, vol. 97, no. 3, pp. 369–387, Jul. 2008, doi: 10.1002/j.2168-9830.2008.tb00985.x.

[143] L. Varpio, E. Paradis, S. Uijtdehaage, and M. Young, "The Distinctions Between Theory, Theoretical Framework, and Conceptual Framework," *Academic Medicine*, vol. 95, no. 7, pp. 989–994, Jul. 2020, doi: 10.1097/ACM.0000000000003075.

[144] G. Bordage, "Conceptual frameworks to illuminate and magnify," *Med Educ*, vol. 43, no. 4, pp. 312–319, Apr. 2009, doi: 10.1111/j.1365-2923.2009.03295.x.

[145] M. Yu and S. M Smith, "Grounded Theory: A Guide for a New Generation of Researchers," *International Journal of Doctoral Studies*, vol. 16, pp. 553–568, 2021, doi: 10.28945/4836.

[146] J. Piaget, *The Psychology of Intelligence.*, 3rd ed. Littlefied, Adams and Co, 1960.

[147] L. S. VYGOTSKY, *Mind in Society*. Harvard University Press, 1980. doi: 10.2307/j.ctvjf9vz4.

[148] J. S. Bruner, *Toward a Theory of Instruction*. Harvard University Press, 1974.

[149] B. Skinner, *About Behaviorism*. Knopf Doubleday Publishing Group, 1974.

[150] J. Piaget, *The Origins of Intelligence in Children*. International Universities Press, 1952.

[151] D. Wood, J. S. Bruner, and G. Ross, "THE ROLE OF TUTORING IN PROBLEM SOLVING," *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89–100, Apr. 1976, doi: 10.1111/j.1469-7610.1976.tb00381.x.

[152] J. Biggs, C. Tang, and G. Kennedy, *Teaching For Quality Learning At University*, 5th ed. Open University Press, 2022.

[153] P. Armstrong, "Bloom's Taxonomy," Vanderbilt University Center for Teaching. [Online]. Available: https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

[154] L. W. Anderson, D. R. Krathwohl, and B. S. Bloom, *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. 2001.

[155] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, "The Classification of Educational Goals," *Taxonomy of educational objectives*, p. 207, 1956.

[156] B. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, *Handbook I: cognitive domain*. New York: David McKay, 1956.

[157] M. J. Booker, "A Roof without Walls: Benjamin Bloom's Taxonomy and the Misdirection of American Education," *Academic Questions*, vol. 20, no. 4, pp. 347–355, Dec. 2007, doi: 10.1007/s12129-007-9031-9.

[158] R. Case, "The Unfortunate Consequences of Bloom's Taxonomy," *Social Education*, vol. 77, no. 4, pp. 196–200, 2013.

[159] J. Biggs and K. F. Collis, *Evaluating the Quality of Learning: The SOLO Taxonomy*. 1982.

[160] L. Fink, *Creating significant learning experiences: An integrated approach to designing college courses*. Jossey-Bass, 2013.

[161] D. A. Kolb, "Experiential Learning: Experience as The Source of Learning and Development," *Prentice Hall, Inc.*, no. 1984, pp. 20–38, 1984, doi: 10.1016/B978-0-7506-7223-8.50017-4.

[162] P. M. Newton, "The Learning Styles Myth is Thriving in Higher Education," *Front Psychol*, vol. 6, Dec. 2015, doi: 10.3389/fpsyg.2015.01908.

[163] C. Riener and D. Willingham, "The Myth of Learning Styles," *Change: The Magazine of Higher Learning*, vol. 42, no. 5, pp. 32–35, 2010, doi: 10.1080/00091383.2010.503139.

[164] S. Kurt, "Kolb's Experiential Learning Theory & Learning Styles," Educational Technology Consulting Services. [Online]. Available: https://educationaltechnology.net/kolbs-experiential-learning-theory-learning-styles/

[165] I. P. Pavlov, "Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex," *Ann Neurosci*, vol. 17, no. 3, Jun. 2010, doi: 10.5214/ans.0972-7531.1017309.

[166] D. A. Kolb and M. Plovnick, "The experiential learning theory of career development," Cambridge, 1974.

[167] M. Abdulwahed and Z. K. Nagy, "Applying Kolb's Experiential Learning Cycle for Laboratory Education," *Journal of Engineering Education*, vol. 98, no. 3, pp. 283–294, Jul. 2009, doi: 10.1002/j.2168-9830.2009.tb01025.x.

[168] J. Biggs, “What the Student Does: teaching for enhanced learning,” *Higher Education Research & Development*, vol. 18, no. 1, pp. 57–75, Apr. 1999, doi: 10.1080/0729436990180105.

[169] G. Gallagher, “Aligning for learning: Including feedback in the Constructive Alignment model,” *All Ireland Journal of Higher Education*, vol. 9, no. 1, p. 3012, 2017.

[170] M. Paskevicius, “Conceptualizing Open Educational Practices through the Lens of Constructive Alignment,” *Open Praxis*, vol. 9, no. 2, p. 125, 2017, doi: 10.5944/openpraxis.9.2.519.

[171] J. Biggs, “Enhancing teaching through constructive alignment,” *High Educ (Dordr)*, vol. 32, no. 3, pp. 347–364, Oct. 1996, doi: 10.1007/BF00138871.

[172] G. Ruge, O. Tokede, and L. Tivendale, “Implementing constructive alignment in higher education – cross-institutional perspectives from Australia,” *Higher Education Research & Development*, vol. 38, no. 4, pp. 833–848, Jun. 2019, doi: 10.1080/07294360.2019.1586842.

[173] C. Loughlin, S. Lygo-Baker, and Å. Lindberg-Sand, “Reclaiming constructive alignment,” *European Journal of Higher Education*, vol. 11, no. 2, pp. 119–136, Apr. 2021, doi: 10.1080/21568235.2020.1816197.

[174] G. Wiggins and J. McTighe, *Understanding by design*. ASCD, 2005.

[175] J. E. Corter, S. K. Esche, C. Chassapis, J. Ma, and J. V Nickerson, “Process and learning outcomes from remotely-operated, simulated, and hands-on student laboratories,” *Comput Educ*, vol. 57, no. 3, pp. 2054–2067, 2011, doi: 10.1016/j.compedu.2011.04.009.

[176] J. R. Brinson, “A Further Characterization of Empirical Research Related to Learning Outcome Achievement in Remote and Virtual Science Labs,” *J Sci Educ Technol*, vol. 26, no. 5, pp. 546–560, Oct. 2017, doi: 10.1007/s10956-017-9699-8.

[177] D. Basu and V. K. Lohani, “Learning and engagement with an online laboratory for environmental monitoring education,” *European Journal of Engineering Education*, 2023, doi: 10.1080/03043797.2023.2175199.

[178] W. M. Mcdonald, R. L. Dymond, V. K. Lohani, and D. S. Brogan, “A Continuous, High-Frequency Environmental Monitoring System for Watershed Education Research,” *Journal of Engineering Education Transformations*, vol. 28, no. 4, p. 11, Apr. 2015, doi: 10.16920/jeet/2015/v28i4/70385.

[179] W. C. Newstetter and M. D. Svinicki, “Learning Theories for Engineering Education Practice,” in *Cambridge Handbook of Engineering Education Research*, Cambridge University Press, 2014, pp. 29–46. doi: 10.1017/CBO9781139013451.005.

[180] S. Verstege, J.-P. Vincken, and J. Diederen, “Blueprint to design virtual experiment environments,” *Computers and Education Open*, vol. 2, p. 100039, Dec. 2021, doi: 10.1016/j.caeo.2021.100039.

[181] Y. Rogers, N. Yuill, and P. Marshall, “Contrasting Lab-Based and in-the-Wild Studies for Evaluating Multi-User Technologies,” in *The SAGE Handbook of Digital Technology Research*, 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, 2013, pp. 359–373. doi: 10.4135/9781446282229.n24.

[182] J. McCambridge, J. Witton, and D. R. Elbourne, “Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects,” 2014. doi: 10.1016/j.jclinepi.2013.08.015.

[183] T. E. Pinelli, “Distinguishing Engineers from Scientists-The Case for an Engineering Knowledge Community,” *Sci Technol Libr (New York, NY)*, vol. 21, no. 3–4, pp. 131–163, Nov. 2001, doi: 10.1300/J122v21n03_09.

[184] R. Gluga, J. Kay, R. Lister, S. Kleitman, and T. Lever, “Over-confidence and confusion in using bloom for programming fundamentals assessment,” in *SIGCSE’12 - Proceedings of the 43rd ACM Technical Symposium on Computer Science Education*, New York, NY, USA: ACM, Feb. 2012, pp. 147–151. doi: 10.1145/2157136.2157181.

[185] The University of Edinburgh, “Technician Commitment.” [Online]. Available: https://technicians.ed.ac.uk/technician-commitment

[186] UK Institute for Technical Skills and Strategy, “Technician Commitment.” [Online]. Available: https://www.techniciancommitment.org.uk

[187] E. Lindsay, S. Murray, D. K. Liu, D. Lowe, and C. G. Bright, “Establishment reality vs. maintenance reality: how real is real enough?,” *European Journal of Engineering Education*, vol. 34, no. 3, pp. 229–234, Jun. 2009, doi: 10.1080/03043790902902906.

[188] C. Hodges, S. Moore, B. Lockee, T. Trust, and A. Bond, “The Difference Between Emergency Remote Teaching and Online Learning,” 2020.

[189] Quanser, “QUBE – Servo 2.” [Online]. Available: https://www.quanser.com/products/qube-servo-2/

[190] “SCQF LEVEL 9.” [Online]. Available: https://scqf.org.uk/level-9/

[191] “Undergraduate Course: Control and Instrumentation Engineering 3 (SCEE09002).” [Online]. Available: http://www.drps.ed.ac.uk/21-22/dpt/cxscee09002.htm

[192] D. Reid, J. Burridge, D. Lowe, and T. Drysdale, “Open-source remote laboratory experiments for controls engineering education,” *International Journal of Mechanical Engineering Education*, vol. 50, no. 4, pp. 828–848, Oct. 2022, doi: 10.1177/03064190221081451.

[193] “Practable Experiment Examples.” [Online]. Available: https://practable.io/experiment-examples.html

[194] D. Hansson, “Software has bugs. This is normal.” [Online]. Available: https://world.hey.com/dhh/software-has-bugs-this-is-normal-26d5fd06

[195] B. Laugwitz, T. Held, and M. Schrepp, “Construction and Evaluation of a User Experience Questionnaire,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5298 LNCS, 2008, pp. 63–76. doi: 10.1007/978-3-540-89350-9_6.

[196] “Practical Engineering Education Conference 2021.” [Online]. Available: https://sites.google.com/sheffield.ac.uk/pee/programme/session-3

[197] APA Dictionary of Psychology, “input-process-output model (IPO model).” [Online]. Available: https://dictionary.apa.org/input-process-output-model

[198] G. V Glass and F. S. Ellett, "Evaluation Research," *Annu Rev Psychol*, vol. 31, no. 1, pp. 211–228, Jan. 1980, doi: 10.1146/annurev.ps.31.020180.001235.

[199] "Cambridge Dictionary: Evaluation." [Online]. Available: https://dictionary.cambridge.org/dictionary/english/evaluation

[200] "Cambridge Dictionary: Assessment." [Online]. Available: https://dictionary.cambridge.org/dictionary/english/assessment

[201] F. D. Davis, "User Acceptance of Information Systems: The Technology Acceptance Model (TAM)," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[202] M. Fowler, "Two Hard Things." [Online]. Available: https://martinfowler.com/bliki/TwoHardThings.html

[203] R. Munroe, "Standards." [Online]. Available: https://xkcd.com/927/

[204] B. Freisleben Hartmann, D., Kielmann, T., "Parallel Incremental Raytracing of Animations on a Network of Workstations," *1998 International Conference on Parallel and Distributed Processing. Techniques and Applications*, vol. III, pp. 1305–1312, 1998.

[205] A. H. Panhwar, S. Ansari, and Shah Asif Ali, "Post-positivism: an effective paradigm for social and educational research | Panhwar | International Research Journal of Arts & Humanities (IRJAH)," *International Research Journal of Arts & Humanities*, vol. 45, no. 45, pp. 253–260, 2017.

[206] I. Harel and S. Papert, "Constructionism : research reports and essays, 1985-1990," 1991.

[207] J. E. Corter, J. V. Nickerson, S. K. Esche, C. Chassapis, S. Im, and J. Ma, "Constructing reality: A study of remote, hands-on and simulated laboratories," *ACM Transactions on Computer-Human Interaction*, vol. 14, no. 2, 2007, doi: 10.1145/1275511.1275513.

[208] L. Park, Sandra; Hironaka, Stephanie; Carver, Penny; Nordstrum, "Continuous Improvement in Education," *Carnegie Foundation for the Advancement of Teaching*, no. White Paper; Advancing Teaching-Improving Learning, pp. 459–479, 2013.

[209] A. Shekh-Abed and N. Barakat, "Transition of Engineering Education During the COVID-19 Pandemic," in *2021 World Engineering Education Forum/Global Engineering Deans Council (WEEF/GEDC)*, IEEE, Nov. 2021, pp. 29–35. doi: 10.1109/WEEF/GEDC53299.2021.9657167.

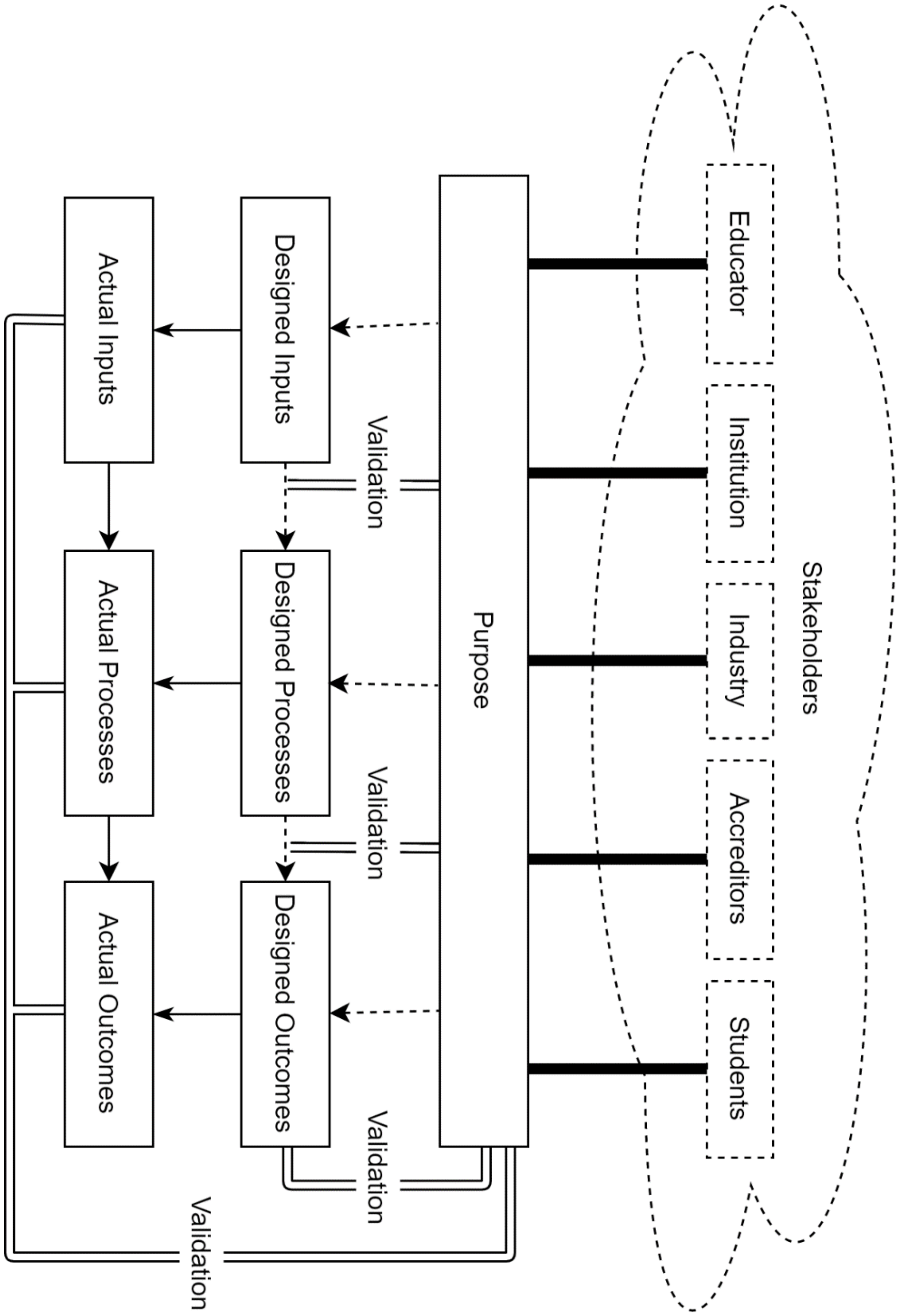
[210] J. Turns, M. Eliot, R. Neal, and A. Linse, "Investigating the Teaching Concerns of Engineering Educators," *Journal of Engineering Education*, vol. 96, no. 4, pp. 295–308, Oct. 2007, doi: 10.1002/j.2168-9830.2007.tb00940.x.

[211] Sage Research, "Methods Map." [Online]. Available: https://methods.sagepub.com/methods-map

[212] S. McKenney and T. C. Reeves, *Conducting Educational Design Research*. Second edition. | New York : Routledge, 2019. | "[First edition: Routledge, 2018. doi: 10.4324/9781315105642.

APPENDICES

A. REFERENCE PIPOV DIAGRAM



B. Interview Protocol

INTERVIEW PROTOCOL – ENGINEERING EDUCATOR INTERVIEWEE

OPENING:

Hi <name of interviewee>, my name is <name of interviewer> and I will be conducting your interview today.

PIS/PCF

Before we begin, I have an information statement for you that describes what the interview is for, what data is being collected, and it explains your rights regarding your data and consent. Please take some time to read this, and if you have any questions feel free to ask me.

[Face to face – hand over printed copy of PIS]

[Zoom]

I will be sending you this file in the chat window in Zoom. Please let me know if you have received it. If you prefer, I can also email this file to you as well.

[After participant has indicated they are ready to proceed]

Based on that information statement I also have a consent form for you to let me know if you are happy to continue, as well as to indicate whether you are happy for quotes from this interview to be published, whether you are interested in further participation, and if you would like to receive feedback on the outcomes of this research.

[Face to face – hand over printed copy of PCF to sign]

[Zoom]

I have placed a link in the chat window in Zoom (https://educational-laboratoriesresearch.sydney.edu.au/interview_ee_pcf link to be created for digital equivalent of printed PCF). Please use that link in order to record your consent and the other decisions I mentioned if you are happy to do so.

[If consent not granted]

I understand. No data has been collected yet, so we can simply end the interview here - thank you for your time.

[If consent granted]

Thank you. Please remember you can choose not to answer questions you don’t want to, and you can end the interview any time you choose by asking me. You can also choose to withdraw your consent, and I will delete the recordings taken up to that point.

RECORDINGS

[Face to face]

As mentioned in the information statement the interviews will be audio recorded using this device [indicate dedicated voice recorder]. You may also see me writing notes. I am happy to show you the notes I am taking at any time.

[Zoom]

As mentioned in the information statement the interviews will be video and audio recorded using the Zoom record feature. You may also see me writing notes. I am happy to show you the notes I am taking at any time.

[start recording]

TOPIC 1: DESCRIPTION OF LABORATORY

I would like to talk to you about your laboratory, [name or descriptor of laboratory of interest]. Can you please tell me about that laboratory?

Here the follow up questions by necessity will be driven by the participant's description rather than being structured – the information intended to be elicited is:

- *The equipment and activities the laboratory involves*
- *How the laboratory fits into the wider unit of study it is situated in*
- *What the educator believes purpose of the laboratory to be*
- *How well the laboratory ran or is expected to run*

TOPIC 2: FRAMEWORK

Thank you for telling me about your laboratory. One of the ways we are looking to understand laboratories and help educators to think about and communicate their laboratories is with the use of a framework, or you might prefer to think of it as a model. This model divides up the laboratory into Purpose, Inputs, Processes, Outcomes, and Validation.

Based on the information on your laboratory we have discussed so far, I can see aspects of…

Here again the questions will be driven by the participant. The researcher will take the answers to Topic 1 and divide them into the 5 areas of the framework, and where gaps exist will ask questions of the educator i.e.

- *Given that purpose might include the intended learning outcomes of the laboratory, the organisational constraints, etc., is there anything else you might include as a purpose of your laboratory?*
- *Inputs can include the equipment, the student demographics, the lab guide students are given, the environment the laboratory takes place in – can we go through some of these for your laboratory?*
- *Processes include the physical actions that take place as part of the laboratory, but also the cognitive processes of the laboratory as well – so what is taking place, or intended to take place, in the minds of the students undertaking it. Do you have some insight or is there an intention in the laboratory for these sorts of processes?*

- *Outcomes might be obvious outcomes such as a lab report or an exam score, but they could also be more abstract such as an acquired skill, an affective outcome, engagement, etc. What outcomes like these might you expect, or have you observed?*
- *Validation is the question of 'how do you know' for the previous 4 aspects of your laboratory. For purpose, inputs, processes, and outcomes, what measures do you have to indicate that these are the case?*

TOPIC 3: REFLECTION

Thank you for going through this with me. Now that we've gone through the framework, I'm interested in how you experienced that process of thinking through your laboratory in that way. I have some specific questions to ask, but before I ask them can you tell me what you think about that process?

Have you ever used a framework or some other formal process for thinking about or designing your laboratories?

[If yes] Can you tell me what you remember about that framework?

What about teaching in general?

[If yes] Do you believe that it influenced your laboratory teaching?

How well do you believe the framework fit your own understanding of laboratories?

Follow ups to this question will depend on the responses – where similarities or differences are identified by the participant they will be queried by the researcher, e.g.

- *Is this something you do think about but with a different name or categorisation, or is it something you have not thought about regarding laboratories before?*
- *What role do you see this aspect having in your understanding of laboratories? That is, how important do you consider it? Where/when do you start thinking about it in your design process?*

Have you communicated your laboratory designs to other educators before? What methods [have you used/ might you intend to use] to do so?

Do you believe the framework we have gone through would be effective for this communication? Can you see any positive or negative aspects to its use?

WRAP-UP

Thank you so much for participating in this interview, that is all the questions I have for you. If you would like to add anything further that you believe is relevant, please feel free.

I will now stop the recording.

[stop device or zoom recording]

That concludes the interview.

[face to face]

The information statement is yours to keep for reference.

If you have any questions you'd like to ask me about the interview, consent, data, etc you can, otherwise thank you again for your participation.

C. SAMPLE PIPOV DESIGN AND DISCUSSION SUPPORT DOCUMENT

Purpose-First Approach

1-page PIPOV Lab Review - Design Communication

Purpose

What purpose is the laboratory intended to serve? Within the class, within the major, within the degree, beyond?

- What learning outcomes should the laboratory contribute to, and to what extent? Understanding of phenomena, technical expertise, soft skills, etc.
- What graduate attributes or lifelong learning outcomes should the laboratory contribute to, and to what extent?
- What affective results is the laboratory intended to achieve? Engagement, enjoyment, curiosity, etc.
- What opportunities or constraints have motivated the laboratory, or limit its design? Equipment, budget, time, etc.
- Consider a wide variety of stakeholders in this section: the educator(s), the students, the institutional context, industry, certifying bodies.

Designed Inputs

What inputs make up the laboratory design?
These will fit different types:

- Context variables - things you cannot change
- Design variables - things you can select or create

In different categories:

- Experiment
  - Equipment
  - Instructions
- Coordination, Interface, & Synchronicity
  - Student-to-Equipment
  - Student-to-Material
  - Student-to-Teacher
  - Student-to-Student
- Integration with teaching context
- Data Source
- Individual Differences

Designed Processes

What processes occur during the laboratory?

- Physical processes - things the students say, do, write
- Cognitive processes - the less tangible processes of understanding and learning

Usually the physical processes are there to facilitate some intended cognitive processes, such as a student recording observations to realise something about some phenomenon, or engaging with peers in a group activity to learn something about social dynamics, etc

The inputs of a laboratory lead to the processes of the laboratory, which lead to the outcomes of the laboratory.

Designed Outcomes

What outcomes do you expect from the laboratory?
These may be:

- Not Observable
  - Proxy outcome(s) exists
  - No proxy outcome(s) exists
- Observable
  - Observed
  - Not Observed

These outcomes may be observed using:

- Lab reports, exams, other assessments
  - Scores
  - Qualitative analysis
- Surveys
- Interviews

These outcomes will likely include a mix of those indicating the achievement of the Purpose(s) of the laboratory, as well as other unintended outcomes.

Validation

What elements of the above design include iterative validation? This should be embedded in the laboratory design itself to ensure the design is working, to provide opportunities for intervention when things change or break, and to allow for iterative improvement and re-alignment of the laboratory over time. PIPOV identifies 8 key opportunities for this validation:

Run - should be embedded in the lab design

- Whether the Designed Inputs match the Actual Inputs
- Whether the Designed Processes match the Actual Processes
- Whether the Designed Outcomes match the Actual Outcomes

Design - should be reconsidered iteratively

- Whether the Stakeholder intentions align with the declared Purposes
- Whether the progression from Inputs to Processes is likely to occur as designed
- Whether the progression from Processes to Outcomes is likely to occur as designed
- Whether the Designed Outcomes, if achieved, mean the Purpose(s) have been achieved
- Whether the Actual Outcomes necessitate a reconsideration of the Purpose(s)

Inputs

- Experiment
- Coordination
- Interface
- Synchronicity
- Integration with teaching context
- Data source
- Individual differences

Inputs - Context Variables

Cannot be changed, might be observable, must considered in the design (which is which will be different depending on the lab)

- Individual differences - student prior/concurrent knowledge, engagement, ‘talent’, available time, learning preferences, etc.
- Equipment (for pilot plant?)

Inputs - Data Source

- Real data vs simulated
- Refers to the root source of the data
- Exists independently of other considerations (particularly synchronicity and interface)

- If simulated - nature of the data as well. Noisy, 'perfect' to theory, etc.

Inputs - Synchronicity

Between:

- Students and teachers - able to ask live questions/make live observations of students, vs asynchronous communication
- Students and other students - working & communicating at the same time, or async
  - Different characteristics of async - short-term being async within the same lab activity, but the potential exists for async across longer time periods as well, e.g. across semesters.
- Students and the equipment - async captures 'batch process' queued jobs
- Students and the learning material - usually synchronous (except certain exam-style conditions)

Inputs - Interface

‘Hands-on’, Computer-mediated (remote, virtual, augmented), between:

- Students and teachers - in the same room, over the internet, etc.
- Students and other students - same again
- Students and the equipment - the classic ‘remote lab’, ‘VR lab’, ‘AR lab’ considerations.
  - Critical for the emergent ‘lab frame’ element - what do the students *believe* about the lab? Is it real?
- Students and the learning material - printed, electronic?
  - Important for data collection.
  - Includes possibilities such as reactive/smart tutoring systems

Inputs - Coordination

How are the students’ interactions coordinated?

- Students and teachers
  - What are the students told by the teachers
  - What are the students allowed to ask
  - What roles are each playing?
    - Demonstrator and student performing activity
    - Simulation of authentic practice, team lead, etc.
    - Assessor
    - Teacher following student instructions

Inputs - Coordination

How are the students' interactions coordinated?

- Students and other students
  - Solo/Group
    - What roles? Peers, lead and underlings, competitors
    - Rotating roles?
    - Prescribed or emergent?

Inputs - Coordination

How are the students' interactions coordinated?

- Students and the equipment
  - Related to the Experiment option
  - What times and mechanisms for accessing the equipment?
    - During class time/business hours/any time
    - Student choice/timetabled/by booking

Inputs - Coordination

How are the students' interactions coordinated?

- Students and the learning material
  - Related to the integration with teaching context
  - Possibilities here a little esoteric
    - Asymmetric knowledge between students within a group
    - Learning material 'gated' (gamification)

Inputs - Experiment

What are the students asked to do?

- 4 main types (more spectrum than fixed)
  - Didactic
    - Cookbook instructions
  - Inquiry-based
    - Question posed, no instructions for obtaining an answer provided
  - Project-based/Problem-based
    - Task assigned, no steps for obtaining the desired result given
  - Research-based
    - No question pre-provided
- These control what the lab guide will include
- Includes what equipment is available
  - With interface and coordination from other aspects
  - Degrees of freedom
  - 'Red herring' aspects - more provided than needed?

Inputs - Integration with teaching context

How does the lab fit into the other teaching?

- Basic considerations: lecture before or after (or unknown), what assessment is done, etc
- More complex options
  - Deeper integrations: lab more closely integrated with other teaching material, lecture/tutorial activities inside lab, lab activities broken up around other material
  - Long term Project-based: lab and other material combine to form part of a larger project

<SUBJECT CODE>

Possible Opportunities

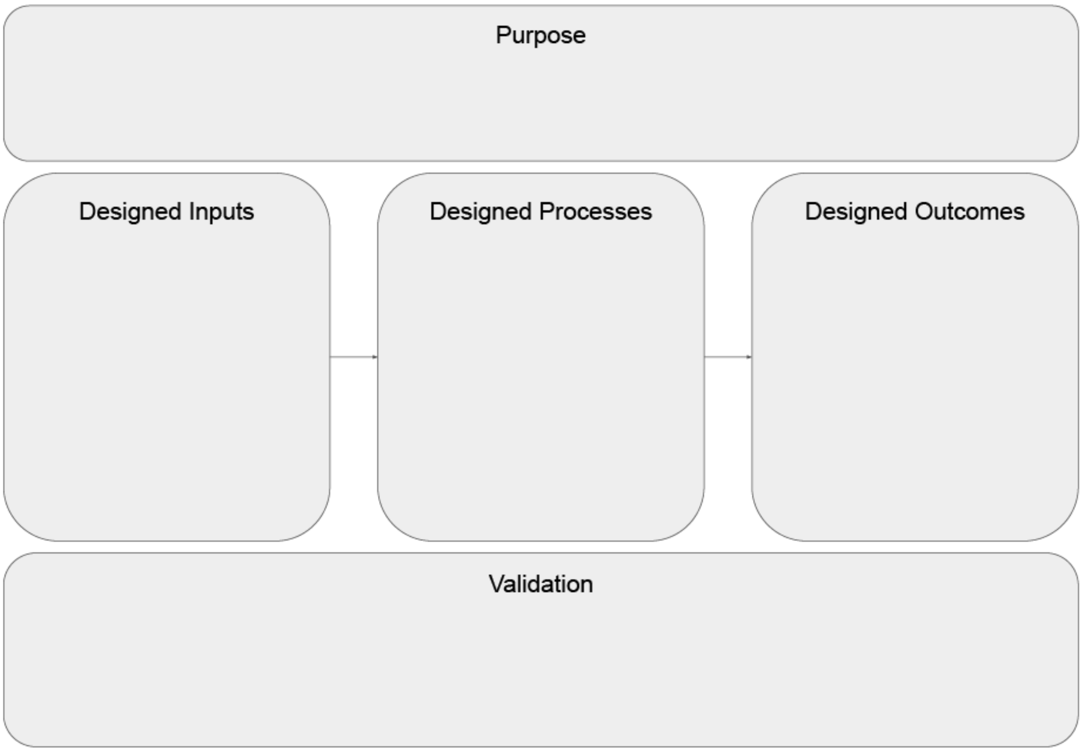
- Existing PBL for potential integration
- LO4 - Critically analyse solution - potential for cross-group peer review

Purpose-First Approach

1-page PIPOV Lab Review - Purpose Articulation - ABC1234 (per-lab)

Purpose-First Approach

1-page PIPOV Lab Review - Design Communication - ABC1234 (per-lab)



<SUBJECT CODE>

Possible Opportunities

- More time to enhance equipment and data processes
  - Remote access
  - AR (in-room augmentation)
  - Demonstrator dashboards
  - Reactive/smart tutoring

Purpose-First Approach

1-page PIPOV Lab Review - Purpose Articulation - ABC1234 (per-lab)

Stakeholders

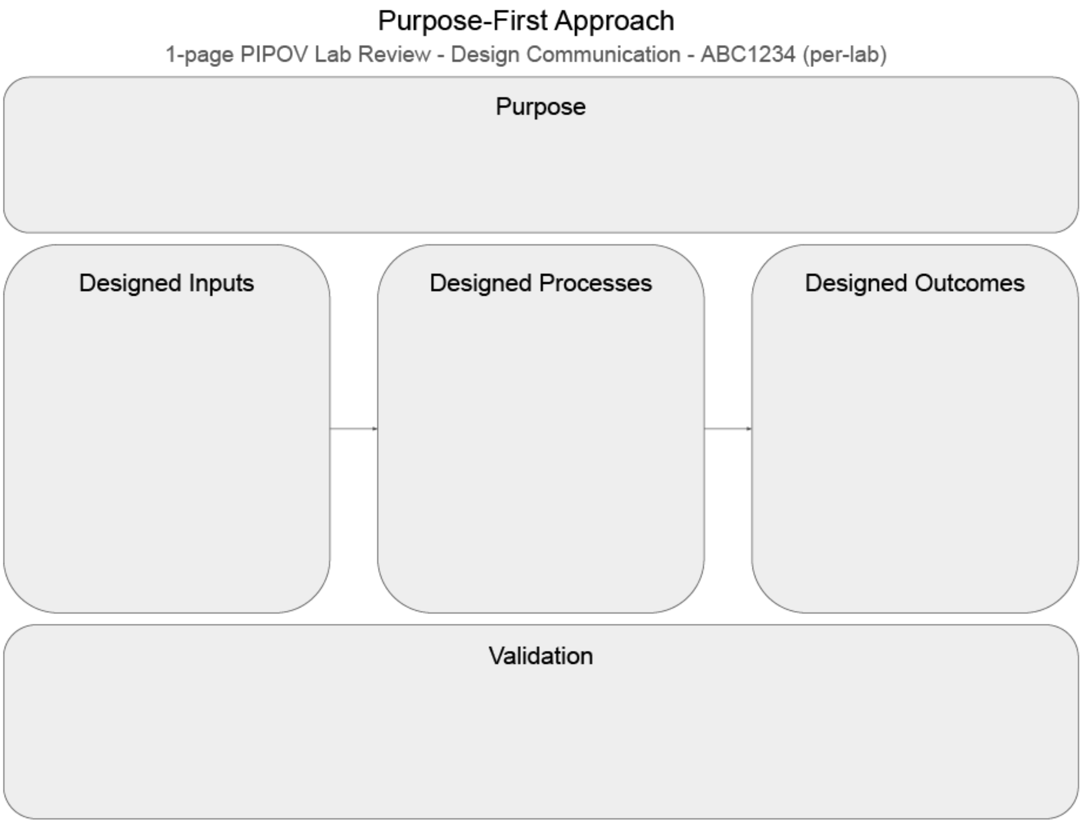
Constraints

Learning Outcomes/Objectives

Graduate Attributes

Current Beneficial Outcomes

Current Undesired Outcomes


Purpose-First Approach
1-page PIPOV Lab Review - Design Communication - ABC1234 (per-lab)
Purpose
Designed Inputs
Designed Processes
Designed Outcomes
Validation

D. What Is A Lab Study - Examples

1

Students attend this class in a room with beakers, flasks, and other chemistry equipment. They have previously been instructed in some properties of acids and bases in a lecture theatre by the professor, and before moving to their individual stations in this room they have watched a teacher use similar equipment to determine the pH of several liquids. They have a worksheet with instructions on how to do this themselves. They then go to their individual stations and carry out these instructions.

2

Students attend this class in a room with desktop computers. They have previously been instructed in some properties of acids and bases in a lecture theatre by the professor, and before beginning their own work they have watched a video of someone using equipment to determine the pH of several liquids. The desktop machines run a simulation program that allows the setup and running of chemistry glassware processes like titration and distillation. The simulation provides instructions on how to use this to determine the pH of simulated chemicals. They then carry out these instructions.

3

Students begin with instruction in the physical properties of subatomic particles in a lecture theatre by the professor. Students are then given a dataset of raw observations from a series of Large Hadron Collider experiments, and must investigate this data to answer a research question they decide on themselves.

4

Students begin with instruction in deflection in civil structures in a lecture theatre by the professor. Students then attend this class in a room with chairs and desks, where they are given a scale model of a bridge with deflection sensors, and a set of weights. They work in groups of 3 by adding the weights on the bridge, and recording the deflection of various members/components of that bridge. They then analyse this data to determine attributes of the material used in the bridge's construction.

5

Students begin with instruction in deflection in civil structures in a lecture theatre by the professor. Students then attend remotely via the internet - there is a website they access that allows them to remotely control robotic equipment that adds real weights to a full-sized bridge, and displays the deflection this causes. They work individually by using the website to add weights on the bridge, and recording the deflection of various members/components of that bridge. They then analyse this data to determine attributes of the material used in the bridge's construction.

6

Students begin with instruction in deflection in civil structures in a lecture theatre by the professor. They are shown a model bridge in this lecture, and told to design an experiment to determine the materials used in the bridge's construction by adding weights to that bridge, and submit this design to the professor. The professor's TA takes the students' submitted designs, takes the bridge model, and runs through each of these experiments by adding weights to the bridge and records the deflection caused. These results are then given to the students in the next class, which they then analyse to determine attributes of the material used in the bridge's construction.

7

Students attend a lecture theatre where the professor delivers a lecture on electrical circuitry. During this lecture the professor has various piece of equipment such as a breadboard and oscilloscope that he passes out for student inspection during the lecture.

8

Students attend a lecture theatre where the professor delivers a lecture on electrical circuitry. During this lecture the students watch the professor build circuits and display the outputs on an oscilloscope, both the circuit and display of which are projected for the whole room to see. The students record their observations about these circuits and outputs to derive certain formulae.

9

Students attend a lecture theatre where the professor delivers a lecture on electrical circuitry. During this lecture, the professor asks the class questions about particular oscilloscope patterns (designing circuits to produce these patterns requiring understanding of particular formulae and circuit behaviour). In response to these questions the students use a mobile app to design different circuits they believe will produce matching output. The students vote on the submitted designs. The students then watch the professor build the winning designs, and display the outputs on an oscilloscope.

10

Students have been learning the properties of steel structures over the course of their degree. A later unit in their degree is wholly comprised of them working in teams of 3 to investigate a property of steel structures of their choosing using various pieces of physical equipment (including steel structures, sensors, experimental scaffolding etc.). They design, carry out, and report on this work themselves with guidance from a faculty mentor.

11

Students begin with instruction in sustainability of the environment and water in a lecture theatre by the professor. Students then go to a nearby marshland to collect water and soil samples. They take them back to a room containing microscopes and associated equipment (slides etc.) and make observations of the samples they have collected. They use these observations to determine the types and amounts of bacteria in their samples, and report on what this means regarding the circumstances of this marshland.

12

Students begin with instruction in sustainability of the environment and water in a lecture theatre by the professor. Students then go to a nearby marshland to collect water and soil samples. They then send these samples to a commercial laboratory, who replies with a list of the types and amounts of bacteria in the samples. The students then write a report on what this means regarding the circumstances of this marshland.

13

Students visit a series of construction sites where buildings, bridges, and tunnels are being constructed. As part of these site visits, the students are making observations in order to answer a research question they have determined about real engineering practice in industry. They analyse these observations to answer their research question and write a report.

14

Students visit a series of construction sites where buildings, bridges, and tunnels are being constructed. These site visits involve the students meeting engineers 'in the wild', and they have a chance to shadow these engineers as they go about their work. No written work is required to be submitted by the student.

15

Students attend a lecture theatre where the professor delivers a lecture on electrical circuitry. Students then attend a class in a room with tables with electrical supply, electrical equipment (sensors, breadboards, power supplies, etc) and a series of black boxes. These boxes contain an electrical circuit the students cannot see, with a series of connections for electrical input and output. Students must use the provided equipment to determine the circuit inside the black box based on its characteristics.

16

Students attend a lecture theatre where the professor delivers a lecture on electrical circuitry. Students then attend a class in a room with tables with electrical supply, electrical equipment (sensors, breadboards, power supplies, etc) and a series of black boxes. These boxes contain a microcomputer the students cannot see that runs a simulation of an electrical circuit, with a series of connections for electrical input and output. Students must use the provided equipment to determine the 'circuit' inside the black box based on its characteristics.

17

Students begin with instruction in some properties of solutions and suspensions in a lecture theatre by the professor. Students then attend a class in a room with a multitude of complex chemistry equipment hooked up to computers running LabVIEW (an application for the running of sensitive/complex laboratory equipment via computer). The equipment is able to (and must) be run wholly through the labVIEW interface. While in this room, but before moving to their individual stations, they watch a teacher use the equipment to determine the properties of several mixed chemicals. They have a worksheet with instructions on how to do this themselves. They then move around to different pieces of equipment in turn to complete these instructions.

18

Students begin with instruction in some properties of solutions and suspensions in a lecture theatre by the professor. Students are told about the institution's equipment room that contains a multitude of complex chemistry equipment hooked up to computers running LabVIEW (an application for the running of sensitive/complex laboratory equipment via computer), but do not visit this room themselves. The equipment is able to (and must) be run wholly through the labVIEW interface. This interface is available to students via screen-sharing over the internet, which students must access in order to control the equipment. They are able to access a video of the professor using this equipment to determine the properties of several mixed chemicals. They are given a worksheet with instructions on how to do this themselves. They use the screen sharing software to access different pieces of equipment in turn to complete these instructions.

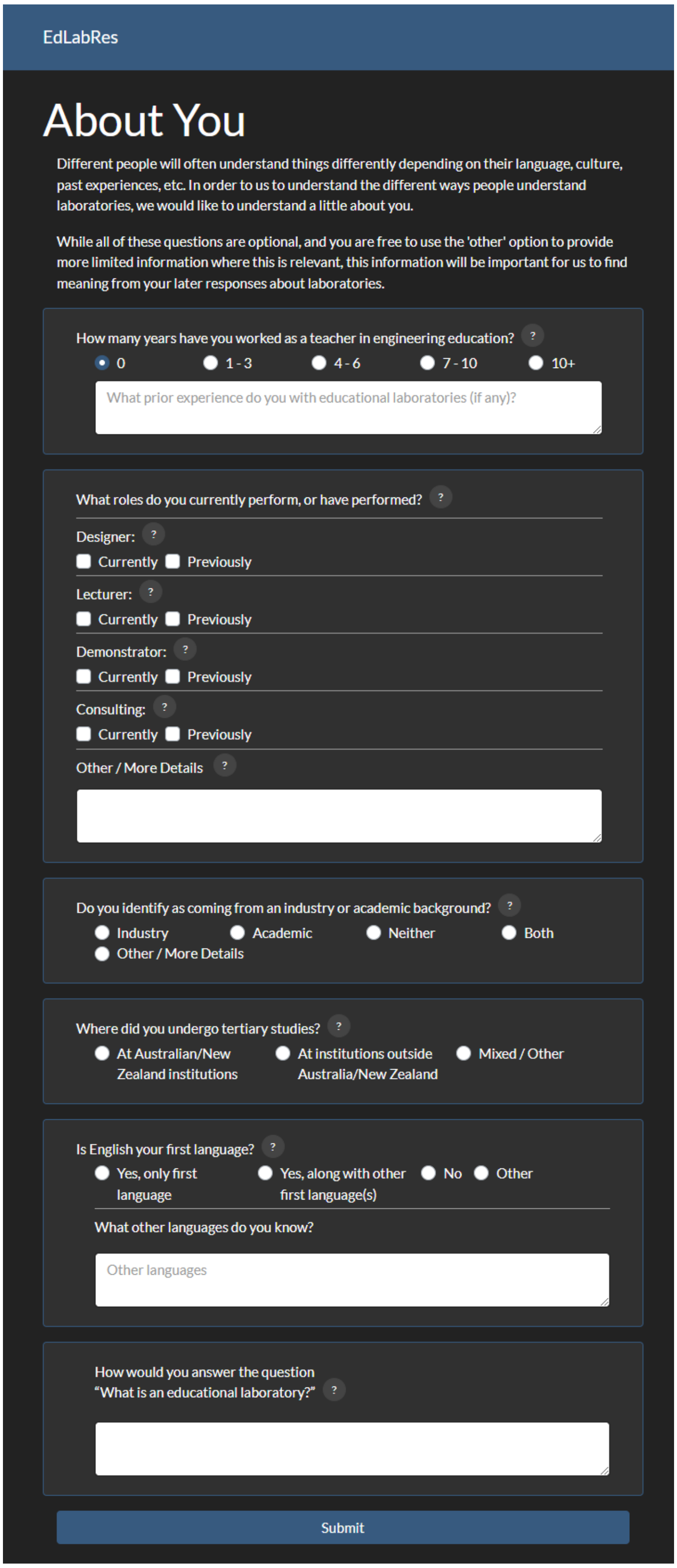
19

A student has been given instruction in some properties of solutions and suspensions in a lecture theatre by the professor. They have not understood some of the concepts in this lecture, and have decided to do some self-investigation to see if their understanding is correct. They make some predictions about the likely results of mixing household chemicals (following all necessary safety precautions of course), and then carry out this mixing. They compare their observations to their predictions to determine if their understanding of the lecture material was correct.

20

A student has been given instruction in some properties of solutions and suspensions in a lecture theatre by the professor. They have not understood some of the concepts in this lecture, and have asked the lecturer for advice. The lecturer has given them a set of lab guides using household chemicals, which they carry out at home using their own equipment. They compare their understanding of the material to their observations as instructed in the lab guides.

E. What Is a Lab – UI and Survey


EdLabRes

About You

Different people will often understand things differently depending on their language, culture, past experiences, etc. In order to us to understand the different ways people understand laboratories, we would like to understand a little about you.

While all of these questions are optional, and you are free to use the 'other' option to provide more limited information where this is relevant, this information will be important for us to find meaning from your later responses about laboratories.

How many years have you worked as a teacher in engineering education? ?
0 | 1 - 3 | 4 - 6 | 7 - 10 | 10+

What prior experience do you with educational laboratories (if any)?

What roles do you currently perform, or have performed? ?

Designer: ?
Currently | Previously

Lecturer: ?
Currently | Previously

Demonstrator: ?
Currently | Previously

Consulting: ?
Currently | Previously

Other / More Details ?

Do you identify as coming from an industry or academic background? ?
Industry | Academic | Neither | Both
Other / More Details

Where did you undergo tertiary studies? ?
At Australian/New Zealand institutions | At institutions outside Australia/New Zealand | Mixed / Other

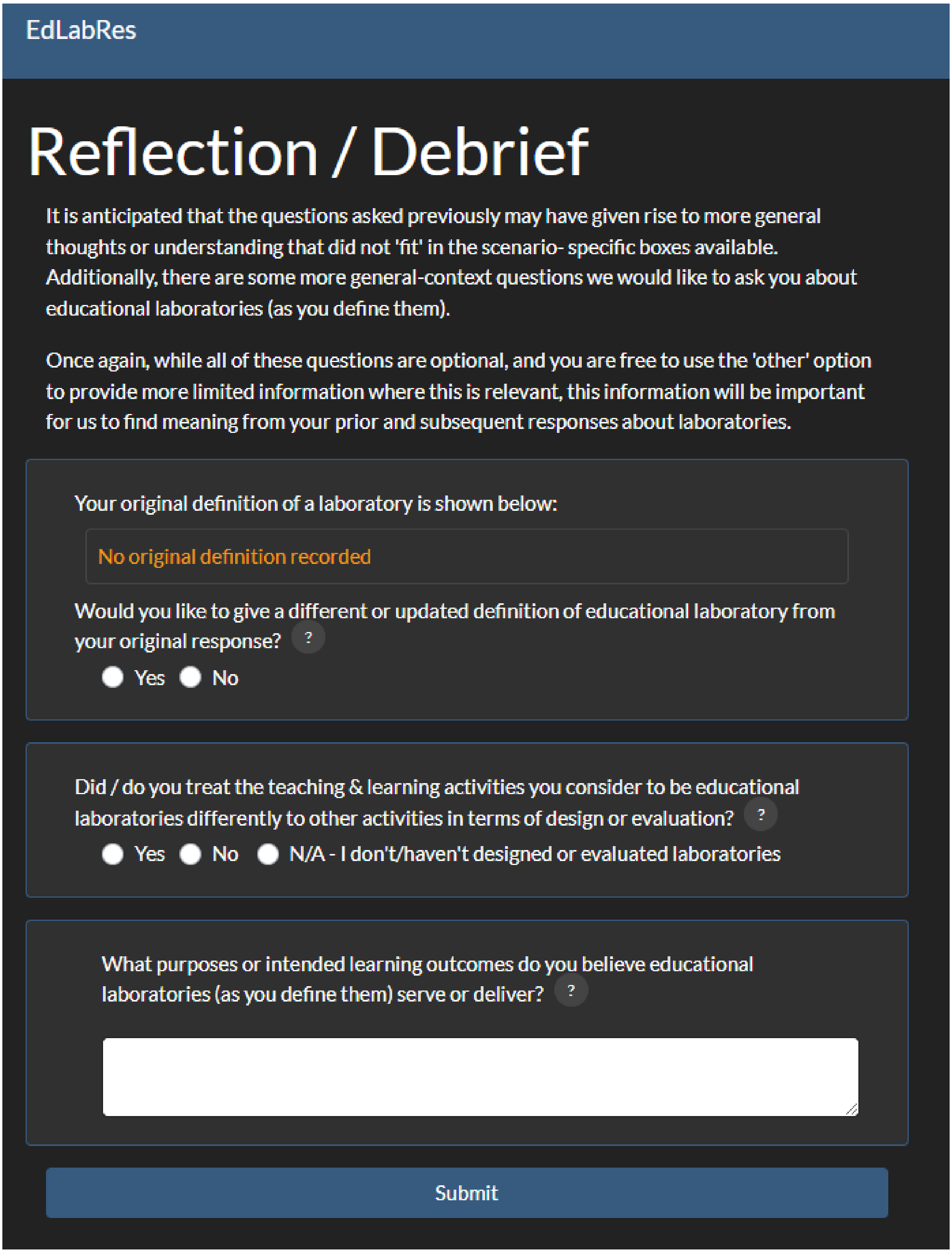
Is English your first language? ?
Yes, only first language | Yes, along with other first language(s) | No | Other

What other languages do you know?

Other languages

How would you answer the question "What is an educational laboratory?" ?

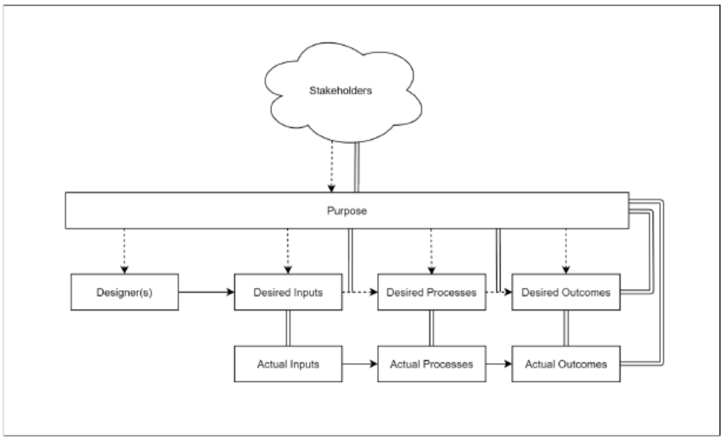
Submit



Note the “No original definition recorded” will instead display the entry from the earlier demographic page for that user if it was entered.

F. PIPOV Workshop Explanatory Materials

PIPOV



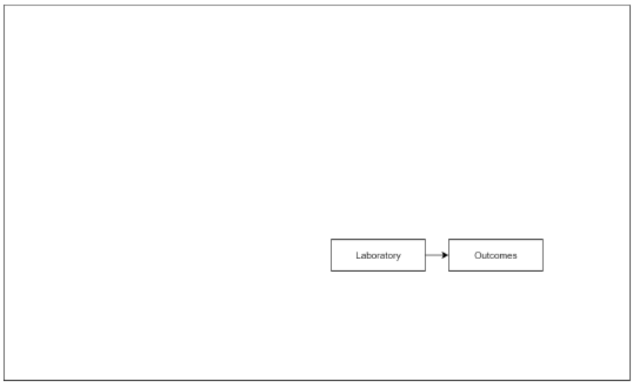
Intro | PIPOV - Framework | Application | Sequence | Outro | 1

So you want to evaluate how well a laboratory is performing – or will perform if you start using it, or make some changes to it. A reasonable place to start is with the outcomes that laboratory yields.

Given a certain **laboratory**, extant or intended, observe the **outcomes**, and from those outcomes you can derive the answer to the question of 'is it working well'.

Labs, however, are far too complex, and exist in far too fragile a context, to be evaluated using this 'black box' methodology. The Purpose-First approach and PIPOV present a method to evaluating labs that acknowledges these challenges, and while not achieving the 'certain knowledge' of true experimental research, maximise the evaluative knowledge that reasonably can be gained about a laboratory.

Note that while this first sequence lays out a progression of changes, these are *not* intended to show a sequence of steps to take – that sequence will be described after the full model has been described.



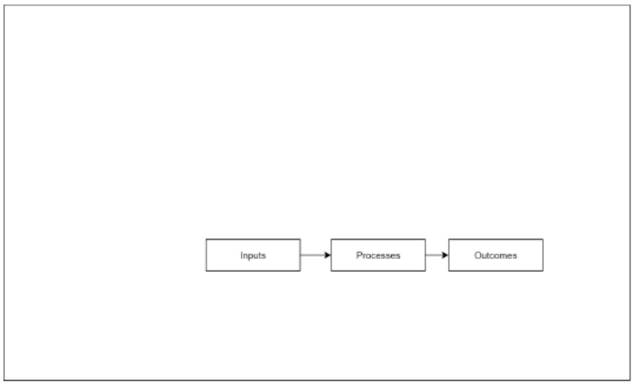
Intro | PIPOV - Framework | Application | Sequence | Outro | 2

PIPOV begins by breaking down the laboratory into its **inputs**, and **processes**. This results in a very similar concept of to the IPO (inputs-processes-outputs) model for system analysis –approaching laboratories as complex systems is a useful starting point for understanding them. This division into the 3 I P and O concepts comprises the overall model proposed earlier by Nickerson et al that PIPOV is informed by.

Inputs are hugely varied, as well as hugely specific. They include the equipment and instructions for the lab, but also the individual differences of the students participating in it, the interfaces on the equipment and between people, the institutional and other contexts that the laboratory exists within, and many other elements. Some of these may be fixed, and some may be chosen.

Processes are both the physical processes that occur during the laboratory (press this button, mix these chemicals, write this report), but also the cognitive processes that occur in the minds of the students when they undertake the lab.

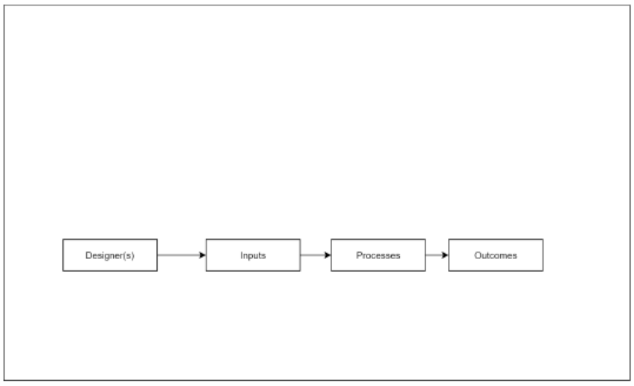
Outcomes are what the laboratory yields. These may include student learning outcomes, time/cost/effort/space constraints, affective outcomes such as enjoyment, engagement, etc.



Intro | PIPOV - Framework | Application | Sequence | Outro | 3

It is important to recognise that as laboratory **designers**, we can directly change *only* the **inputs** to the 'system' – the **inputs**, whatever they happen to be, will determine the **processes** that occur, and the **processes** will yield outcomes.

Further, there will only be a limited set of the overall **inputs** that can be manipulated – for example, you almost certainly could not change the country, institution, or cohort of students the laboratory was run with. But you may have the ability to change whether students work individually or in pairs, whether the laboratory instructions give didactic instructions or inquiry-based, etc. The main purpose (no pun intended) to Purpose-First and PIPOV is to support those design decisions that are available with guidance and evidence to improve the **outcomes** the laboratory yields.



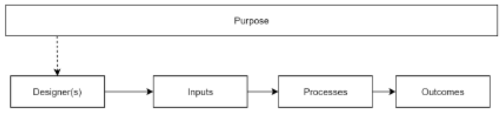
Intro | PIPOV - Framework | Application | Sequence | Outro | 4

Given that we want to optimise the **outcomes**, we need some way of knowing what is desirable, undesirable, and inconsequential. Part of Purpose First is recognising both that this is not a universal truth – i.e. what is desirable for one laboratory in one context may be inconsequential or even undesirable in another – and also that this/these **purpose/s** must be the driving force behind both the design and evaluation of any laboratory.

Before making **input** decisions, and before observing **outcomes**, it is necessary to articulate the **purpose** of the laboratory: why is it being used? What is it intended to provide in terms of learning outcomes? What constraints exist, such as budget, real estate, etc.

Legend note: The 2 types of arrows used here hold semantic value. A solid arrow denotes a relationship that is inherent and unbreakable: e.g. the **processes** of a lab inherently and necessarily yield the **outcomes.** No design can change this – the only way the outcomes can be altered is by also altering the processes that yield them.

A dashed arrow denotes a relationship that is recommended by the PIPOV model itself – it is not guaranteed, and ensuring the consideration of these relationships and the alignment of the elements involved is the method by which PIPOV is applied. Here for example the **purpose** should drive the **designer** when making **input** decisions – but there is nothing *inherently* preventing the designer from simply flipping a coin to make those decisions.



Legend note: The 2 types of arrows used here hold semantic value. A solid arrow denotes a relationship that is inherent and unbreakable: e.g. the **processes** of a lab inherently and necessarily yield the **outcomes**. No design can change this – the only way the outcomes can be altered is by also altering the processes that yield them.

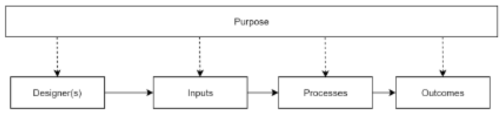
A dashed arrow denotes a relationship that is recommended by the PIPOV model itself – it is not guaranteed, and ensuring the consideration of these relationships and the alignment of the elements involved is the method by which PIPOV is applied. Here for example the **purpose** should drive the **designer** when making **input** decisions – but there is nothing *inherently* preventing the designer from simply flipping a coin to make those decisions.

Intro | PIPOV - Framework | Application | Sequence | Outro

5

Because we cannot treat laboratories as a black box and expect to obtain reliable evaluative results, so too must the purpose of the laboratory be considered at each phase of the system.

The main reason for this is that there will be **outcomes** our **purpose** holds as desirable that cannot be observed (or that the act of observing them is too invasive, damaging, difficult, or error-prone). These **outcomes** can still be promoted, albeit in a less certain way, by including a consideration of the laboratory's **purpose** when designing the **processes** that students participating in the laboratory will undergo, and the **inputs** that yield those **processes.**



Intro | PIPOV - Framework | Application | Sequence | Outro

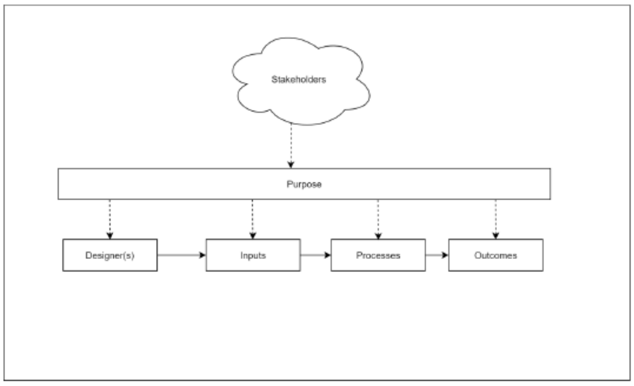
6

Purposes for laboratories are not simply limited to the learning outcomes in the students that a single educator might want them to achieve (though these are of course important!) – they include all the desirable **outcomes** from all parties involved with or affected by the laboratory.

A common style of statement to hear is 'I would have liked to do X, but I couldn't because we needed to do Y'. Sometimes, that is a limitation from **inputs** you do not have control over – no matter how much a lab might benefit from 2 years of previous study this will never happen for a lab delivered in the first semester of a degree. Many times however, it is simply not recognising a competing purpose from a **stakeholder.**

Doing something in a certain way to meet accreditation requirements, using a particular tool that is popular with industry partners, using a simulated data source since there is no available floor space for the necessary equipment – these are all to do with **purposes** that must be considered: meeting an accreditation criterion, producing graduates who are familiar with key industry tools, working within the available real estate of the department. No design or evaluation of a laboratory could hope to be successful without acknowledging and incorporating these **purposes.**

Note that while the **Designer(s)** and **Stakeholders** elements are separate, these should be considered 'roles', not individuals. A **designer** is by definition a **stakeholder**, and while not all **stakeholders** will be involved in the design of the laboratory, many might – including collaborative designs involving teams of academics, industry partners, students, etc.



Intro PIPOV - Framework Application Sequence Outro 7

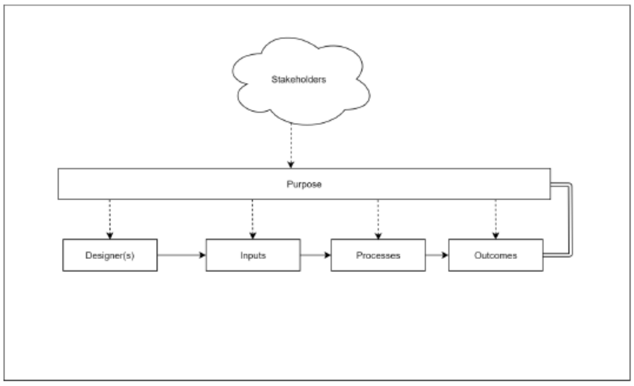
When the laboratory has been run and its **outcomes** have been yielded, it is then necessary to compare those **outcomes** with the **purpose(s)** of the laboratory. This comparison is the validation of the **outcomes**, and the laboratory overall.

Here the difference between **outcomes** and **purposes** becomes important. While effectively all of the desired **outcomes** will be identical to the **purpose** of the laboratory ('students being able to apply x knowledge or y skill', 'the laboratory being delivered within 2 weeks instead of the original 3', etc) the **purpose** may hold statements that do not appear in **outcomes** (i.e., purposes that are *not* met) – and there will be many, many **outcomes** that do not appear as **purposes** (or as negative counterparts to positive purposes).

These unintended **outcomes** will be either positive or negative. For both, a decision must be made as to whether they represent a gap in the previously articulated **purpose**. If an unintended **outcome** is negative, is it desirable to change the laboratory to avoid it? If so, it represents a **purpose** that was originally missed. Conversely for a positive unintended **outcome** – can it be ignored, and a future change in the lab drop it without consequence? If not, it must be included as a **purpose** of the laboratory.

A good example for both is a laboratory whose **purpose** is to teach a specific skill. The skill is learned effectively, and the laboratory may be considered a success. Students work in groups of 3 to complete the laboratory. The negative unintended **outcome** of the laboratory is that students take a significant amount of time to complete the laboratory, and this results in some negative feedback. The positive unintended **outcome** is that students gain significant experience managing work in a team and improve in this area.

It would be necessary in this circumstance to consider whether the increased workload is a 'deal breaker' for the laboratory – and whether a change in the laboratory to solve this that might damage the learning outcomes would be desirable. Conversely, the teamwork skills it provides may be desirable and a reason to not make some future change to the laboratory (such as moving to an online platform where each student has their own equipment to access) or they may not be a relevant **purpose** for the laboratory and so the change may occur regardless. In this way the **purpose** of the laboratory is validated against the **outcomes** as much as the other way around.



Legend note: the double line indicates a 'validation' relationship. This validation may involve checking for differences and handling, or otherwise ensuring alignment that will best serve the overall purpose.

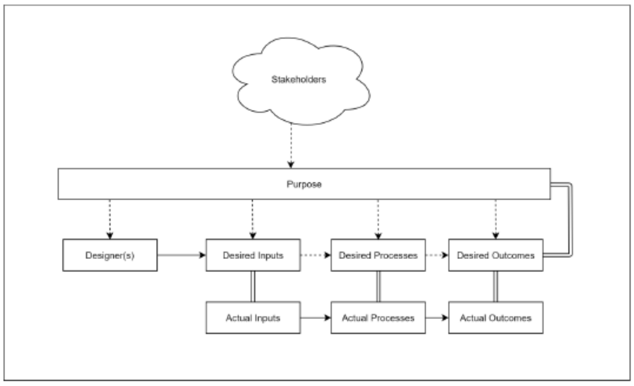
Intro PIPOV - Framework Application Sequence Outro 8

The existence of a difference between the **desirable outcomes** and the **actual outcomes** holds true for all phases of the laboratory. The **actual processes** the students in a laboratory undergo may or may not align with the **desired processes** – and even some **actual inputs** may be different to the **desired inputs** of the laboratory.

For example, the intended design may call for groups of 3 where all members do some of each step in the laboratory – but in reality students may 'divide and conquer' the tasks rather than work together, or 1 student may dominate the group and do everything.

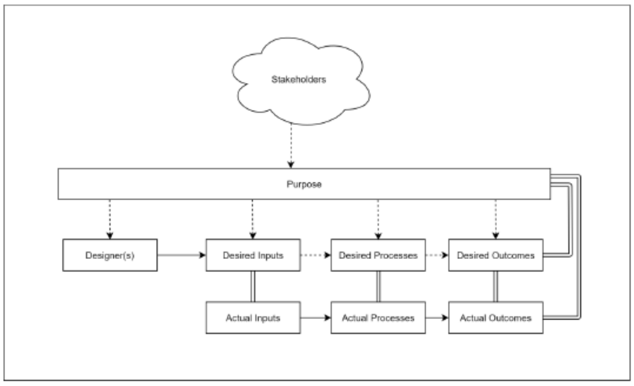
Pay close attention to changes in the solid vs dashed lines here – only the relationships between **designer(s)** to **desired inputs**, and the progression from **actual inputs** to **actual processes** to **actual outcomes** is inherent. There is no direct link between the **designer(s)** and the **actual inputs**, and none of the relationships between **desired inputs, processes, or outcomes** are inherent – alignment of these elements is a key part of applying the PIPOV model, but is in no way guaranteed.

The validation between **desired outcomes** and **purpose** denotes ensuring that the **desired outcomes** fully reflect the **purpose** – i.e. that the statement 'if and only if these outcomes are achieved then the laboratory's purpose has been met' is true.



Intro PIPOV - Framework Application Sequence Outro 9

Thus, the validation between **purpose** and **outcomes** after the laboratory is run must of course consider the **actual outcomes**, rather than simply the **desired outcomes**. In the same way that unintended **actual outcomes** may inform changes to the **purpose**, these should trigger changes in the **desired outcomes**.

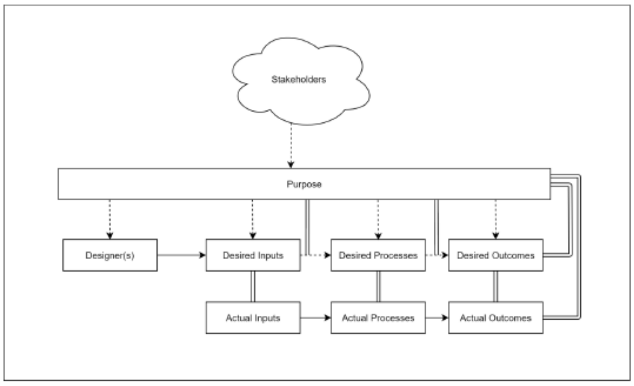


Intro PIPOV - Framework Application Sequence Outro 10

In many if not all circumstances the design of a laboratory must occur with very limited 'certain' knowledge about the likely results it will have. Each occurs in a unique combination of contexts.

It is therefore necessary when designing to validate the alignment of the different phases of the design: "are the **desired pr**ocesses of this design likely to result in the **desired outcomes** and thus serve the **purpose** of the laboratory" and "are the **desired inputs** of this design likely to result in the **desired processes** and thus serve the **purpose** of the laboratory" are 2 key questions that must be answered when designing a laboratory.

In this way those **outcomes** that cannot be observed, or should not be observed due to the likely impact of that observation, can still be supported by the design of the laboratory in a 'best effort' way.

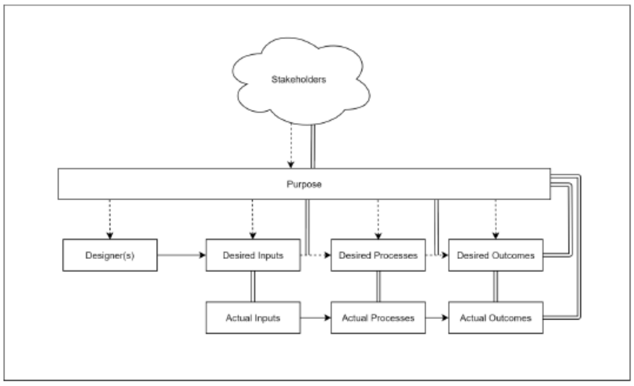


Intro PIPOV - Framework Application Sequence Outro 11

The final relationship in the PIPOV model ensures that the **stakeholders** that generated the original **purpose** are included in the validation of that purpose – both by confirming with them whether or not the **purposes** are being met by the design, as well as involving them in any changes to the **purpose** brought on by unintended **outcomes**.

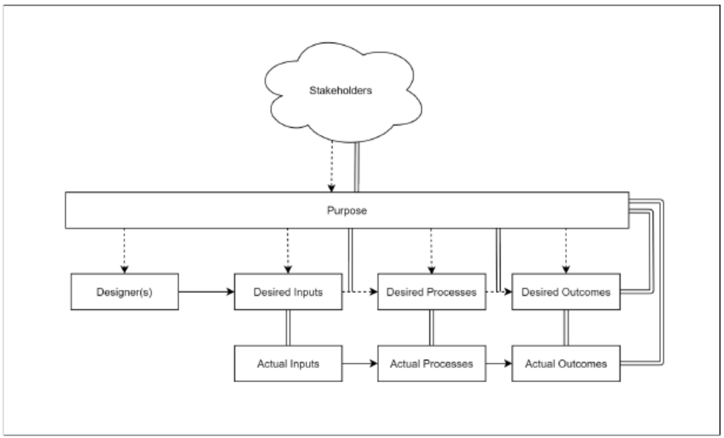
This then is the full PIPOV model – **Stakeholders** inform **Purpose(s)** that inform the **Designer(s)**, who use those to inform their design of **Desired Inputs**, **Processes**, and **Outcomes**, that (ideally) result in matching **Actual Inputs**, **Processes**, and **Outcomes** – and this design is then validated against the **Purpose** along with the **Stakeholders** to confirm whether the laboratory is providing what it is intended to provide.

Purpose
Inputs
Processes
Outcomes
Validation



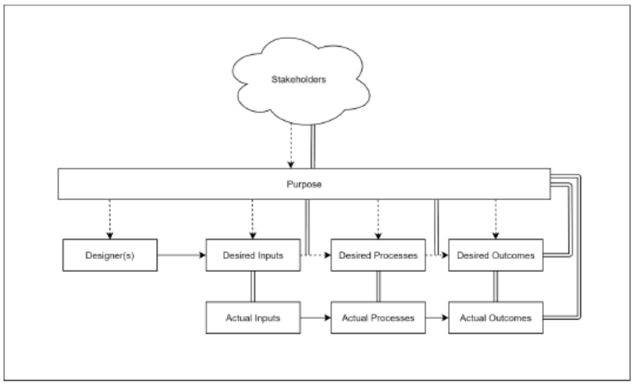
Intro PIPOV - Framework Application Sequence Outro 12

Application



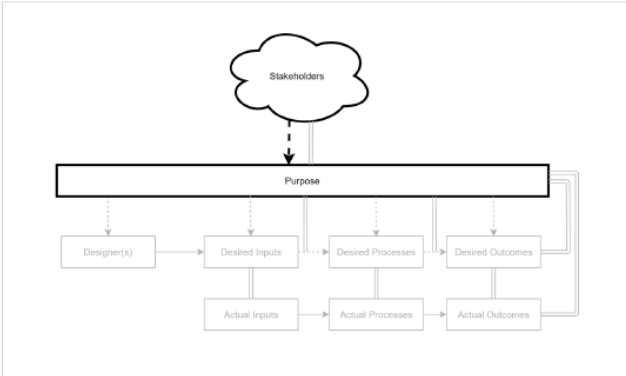
So what does this look like in practice? PIPOV presents a series of design stages that should be followed to yield effective laboratories that serve their intended purposes, that each leverage some of the relationships between the different elements of the model.

These design stages are non-prescriptive in terms of when they should be applied – there is a sequence to them, but it is not linear. This sequence will be described in detail after the stages themselves.



Design stage: Articulate Purpose(s)

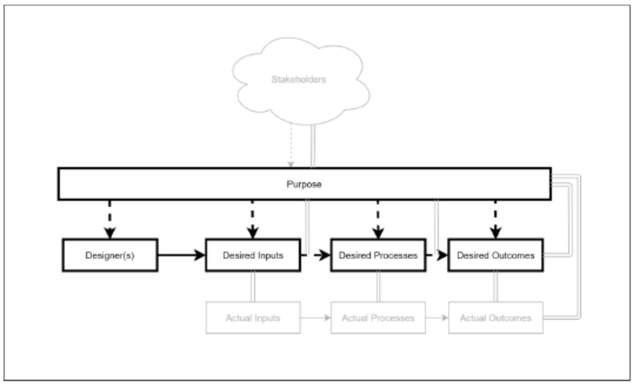
In this stage the various **stakeholders** should be consulted or collaborated with to determine what the **purpose** of the laboratory is. Remember that **stakeholder** and **designer** are not mutually exclusive – the **designer** will of course be involved as a key **stakeholder** in this step. Also remember that these stages do not necessarily occur only once, nor only in the sequence listed here – this stage will potentially be revisited many times over the life of the laboratory.



Design stage: Plan Laboratory

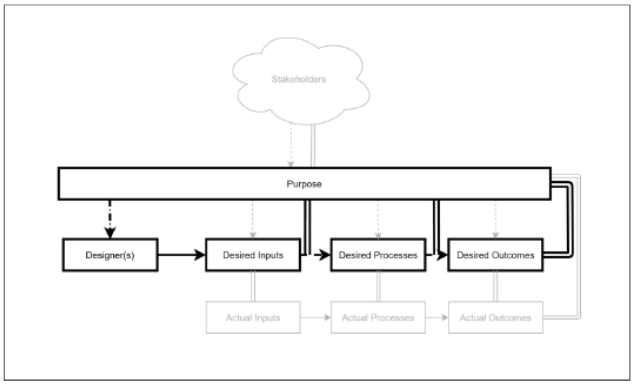
This is the stage where most of the 'design' of the laboratory takes place. It is not called 'design the laboratory' though because the design process involves far more than just drawing up the plans.

Given the **purpose** of the laboratory, the designer must plan **inputs** to yield the **processes** to yield the **outcomes** that the **purpose** requires. This is almost certain to require a combination of very different levels of certainty, from well studied aspects to complete guesswork – this is normal, and the reason that the design process involves validation and modification.



Design stage: Validate Plan

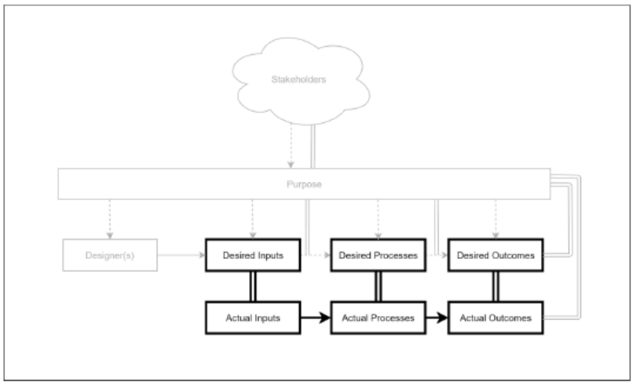
With a plan developed, there should be a deliberate effort to consider the various components against the purpose of the laboratory. Will the **desired outcomes** meet the **purpose**? Are the **inputs**, **processes**, and **outcomes** aligned with the **purpose** to the limit of the available knowledge about them?



Intro PIPOV - Framework Application Sequence Outro 17

Design stage: Run Laboratory

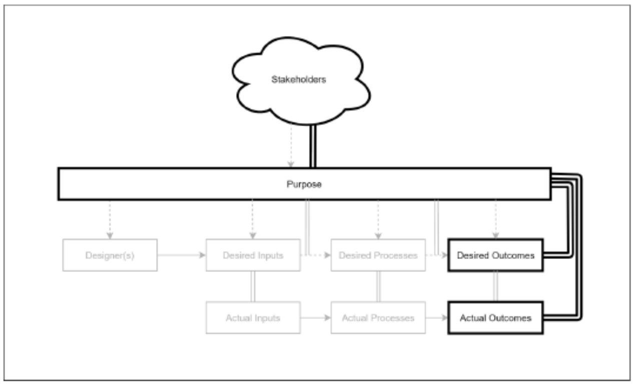
At some point the laboratory must of course be run in its real context, with real students for real results. This provides the opportunity to compare the plan to what has actually occurred ("The best laid schemes o' mice an' men" indicates how unlikely these are to be the same!)



Intro PIPOV - Framework Application Sequence Outro 18

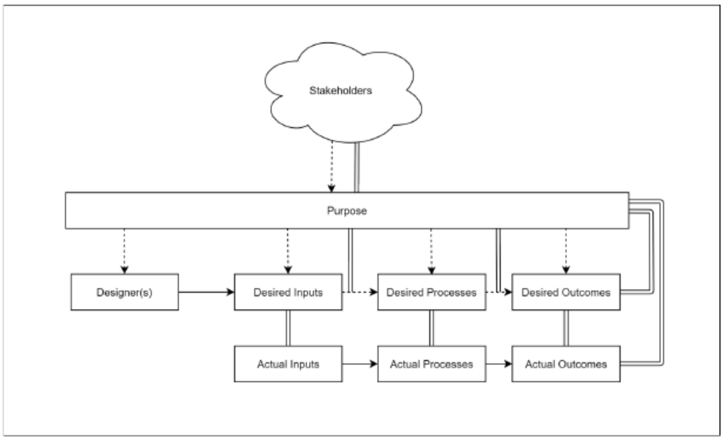
Design stage: Evaluate Laboratory

Once the laboratory has been run, the key question 'did the laboratory achieve its purpose' should be asked. This will involve validating the **outcomes** against the **purpose** – and where unintended **outcomes** lead to changes in the **purpose** these should be validated against the **stakeholders** of the laboratory also.

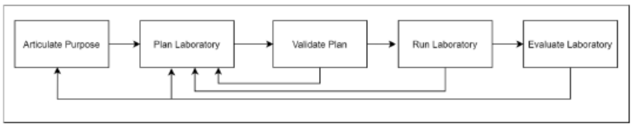


Intro PIPOV - Framework Application Sequence Outro 19

Sequence



Intro PIPOV - Framework Application Sequence Outro 20



The 5 stages of PIPOV combine into a design process that promotes an iterative approach – repeatedly revisiting plans to incorporate feedback.

The primary sequence is to articulate the purpose, plan the laboratory, validate the plan, run the laboratory, and finally to evaluate the laboratory. At its simplest, this process then feeds back into either the purpose or the plan to begin the cycle again.

The nature of laboratories is such that the cycle need not run to its conclusion before feedback can be acted upon however. The plan should be redesigned based on the results of the validation of the plan before the laboratory is run at all – and the laboratory can and should be modified during its run if the early evidence suggests there is a problem.

This is a key way that the purpose first approach is distinct from experimental designs for evaluation – the laboratory should not be held as a constant in the face of evidence to change it. This aligns with what is likely to occur in day to day teaching – if a problem is observed then solutions may be implemented without waiting e.g. for the end of the semester or for some study to be completed.

Intro PIPOV - Framework Application Sequence Outro 21

G. EDINBURGH CONTROL SYSTEMS SURVEY

Remote Labs - User Experience Questionnaire

Page 1: Participant Information Sheet

PARTICIPANT INFORMATION SHEET

You are being invited to take part in research on ‘student perceptions of remote practical work with exploratory user interfaces’. Professor Timothy Drysdale and Mr David Reid at the University of Edinburgh are leading this research. Before you decide to take part it is important you understand why the research is being conducted and what it will involve. Please take time to read the following information carefully.

What is the purpose of the study?

The purpose of the study is to:

1. Provide the practical element of the ‘Controls and Instrumentation 3’ course for Engineering students using a remote lab with interaction via a web based user interface.
2. Evaluate the benefits of different user interface elements or interaction modes.
3. Understand whether varying levels of control of the user interface result in different learning outcomes or student perceptions of their practical work.
4. Utilise student feedback to help improve designs of the user interface.

Why have I been invited to take part?

You are invited to participate in this study because you are enrolled in the ‘Controls and Instrumentation 3’ course and will be using a remote lab designed by this research team to complete the practical element of your course. This research relies on feedback from the student users of the remote labs.

Do I have to take part?

No – it is entirely up to you. If you do decide to take part, please keep this Information Sheet and complete the Informed Consent Form to show that you understand your rights in relation to the research, and that you are happy to participate. If you do decide to take part you are still free to withdraw at any time and without giving a reason. Please note down your participant number (which is on the Consent Form) and provide this to the lead researcher if you seek to withdraw from the study at a later date. Deciding not to take part or withdrawing from the study will not affect your course grade in any way.

1 / 18

What will happen if I decide to take part?

You will be asked a number of questions regarding your experience of working with the remote lab and user interface. These questions will include rating the user interface in terms of its usefulness and ease of use. You will also be asked your opinion on remote laboratory work in general and how you feel it compares to traditional practical work in the lab. You will answer questions on Likert scales (e.g. ranking statements from 'Strongly agree' to 'Strongly disagree' on a 5 point scale) but also be given the chance to provide longer written responses. The questionnaire is entirely online and can be completed in your own time. The questionnaire should take around 20-30 minutes to complete.

As part of the questionnaire, we will also be asking for volunteers for follow up semi-structured interviews to be conducted online (via Teams). These will involve volunteers working with the remote lab interface and explaining their actions as well as discussions of the advantages and disadvantages of remote lab work in general. Ideally, we would like to video record these sessions. This will involve only a small number of students and is completely voluntary – consenting to take part in this study does not automatically make you a participant in the follow on interviews.

What are the possible benefits of taking part?

By sharing your experiences with us, you will be helping the researchers and the University to better understand how we can improve the student experience of working with remote laboratories. There are inherent limitations to attempting practical work remotely and this research aims to understand how we can improve remote labs, and the user interface through which users interact with the lab, in order to improve teaching and learning within practical STEM subjects. Your feedback will have a direct impact on future remote labs at the University of Edinburgh and, in particular, the School of Engineering and may even impact on future courses that you enrol in.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

WHAT IF I WANT TO WITHDRAW FROM THE STUDY?

Agreeing to participate in this project does not oblige you to remain in the study nor have any further obligation to this study. If, at any stage, you no longer want to be part of the study, please inform the project administrator Professor Timothy Drysdale (timothy.drysdale@ed.ac.uk) or Mr David Reid (D.P.Reid@sms.ed.ac.uk). You should note that your data may be used in the production of formal research outputs (e.g. journal articles, conference papers, theses and reports) prior to your withdrawal and so you are advised to contact the research team at the earliest opportunity should you wish to withdraw from the study. On specific request we will destroy all your identifiable answers, but we will need to use the data collected prior to your withdrawal, and to maintain our records of your consenting participation. Please note that once data has been fully anonymised it will be impossible to remove your individual data from the dataset and therefore if you do wish to be removed from the study you should request this as early as possible.

Data Protection and Confidentiality

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Unless they are anonymised in our records, your data will be referred to by a unique participant number rather than by name. If you consent to being video/audio recorded, all recordings will be destroyed once they have been transcribed. Your personal data will only be viewed by the research team. After the study, an anonymised dataset will be stored on Edinburgh DataShare and made openly

2 / 18

accessible to the public. No personally identifiable data will be stored. Data stored on Edinburgh DataShare will be available indefinitely. All electronic data will be stored on a password-protected computer file. Your consent information will be kept separately from your responses in order to minimise risk.

What will happen with the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will always be made anonymous in any formal outputs unless we have your prior and explicit written permission to attribute them to you by name. Information may also be kept for future research.

WHO CAN I CONTACT?

If you have any further questions about the study, please contact the lead researcher, Professor Timothy Drysdale (timothy.drysdale@ed.ac.uk).

If you wish to make a complaint about the study, please contact:

Professor Timothy Drysdale (timothy.drysdale@ed.ac.uk) or the Research Ethics Coordinator in the School of Engineering, Dr Timm Krueger DoR@eng.ed.ac.uk.

In your communication, please provide the study title and detail the nature of your complaint.

You can get this document on tape, in Braille, large print and various computer formats if you ask us. Please contact Mr David Reid (D.P.Reid@sms.ed.ac.uk) stating the title of this study and your request.

For general information about how we use your data go to:

https://www.ed.ac.uk/records-management/privacy-notice-research

Page 2: Participant Consent Form

I confirm that I have read and understood the Participant Information Sheet for the above study. ✱ *Required*

○ Yes

I have been given the opportunity to consider the information provided, ask questions and have had these questions answered to my satisfaction. ✱ *Required*

○ Yes

I understand that my participation is voluntary and that I can ask to withdraw at any time, until the data collection has ended and participants have been anonymised, without giving a reason and without my legal rights being affected. ✱ *Required*

○ Yes

I understand that my anonymised data will be stored for a minimum of 5 years and may be used in future ethically approved research. ✱ *Required*

○ Yes

I agree to take part in this study. ✱ *Required*

○ Yes ○ No

Full name ✱ *Required*

[ ]

4 / 18

Date * *Required*

Dates need to be in the format 'DD/MM/YYYY', for example 27/03/1980.

(dd/mm/yyyy)

Page 3: User Experience Questionnaire Section 1

This question consists of pairs of contrasting attributes that may apply to your experience with the remote lab system. The selection boxes represent gradations between the opposites. You should express your agreement with the attribute by selecting the box that most closely reflects your impression. Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression. Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the remote lab experience. Nevertheless, please select a box in every line. It is your personal experience that counts. There is no wrong or right answer. Please note that you are reflecting on your interaction with the user interface and remote lab hardware, not on the content of the practical work itself. * *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 26 answer(s).

	1	2	3	4	5	6	7	
annoying	☐	☐	☐	☐	☐	☐	☐	enjoyable
not understandable	☐	☐	☐	☐	☐	☐	☐	understandable
creative	☐	☐	☐	☐	☐	☐	☐	dull
easy to learn	☐	☐	☐	☐	☐	☐	☐	difficult to learn
valuable	☐	☐	☐	☐	☐	☐	☐	inferior
boring	☐	☐	☐	☐	☐	☐	☐	exciting
not interesting	☐	☐	☐	☐	☐	☐	☐	interesting
unpredictable	☐	☐	☐	☐	☐	☐	☐	predictable
fast	☐	☐	☐	☐	☐	☐	☐	slow
inventive	☐	☐	☐	☐	☐	☐	☐	conventional
obstructive	☐	☐	☐	☐	☐	☐	☐	supportive
good	☐	☐	☐	☐	☐	☐	☐	bad
complicated	☐	☐	☐	☐	☐	☐	☐	easy
unlikeable	☐	☐	☐	☐	☐	☐	☐	pleasing
usual	☐	☐	☐	☐	☐	☐	☐	leading edge
unpleasant	☐	☐	☐	☐	☐	☐	☐	pleasant
secure	☐	☐	☐	☐	☐	☐	☐	not secure
motivating	☐	☐	☐	☐	☐	☐	☐	demotivating
meets expectations	☐	☐	☐	☐	☐	☐	☐	does not meet expectations
inefficient	☐	☐	☐	☐	☐	☐	☐	efficient
clear	☐	☐	☐	☐	☐	☐	☐	confusing

6 / 18

impractical	☐	☐	☐	☐	☐	☐	☐	practical
organised	☐	☐	☐	☐	☐	☐	☐	cluttered
attractive	☐	☐	☐	☐	☐	☐	☐	unattractive
friendly	☐	☐	☐	☐	☐	☐	☐	unfriendly
conservative	☐	☐	☐	☐	☐	☐	☐	innovative

Page 4: User Experience Questionnaire Section 2

Having completed the practical work, what emotions are you feeling about the practical work? (Select all that apply) * *Required*

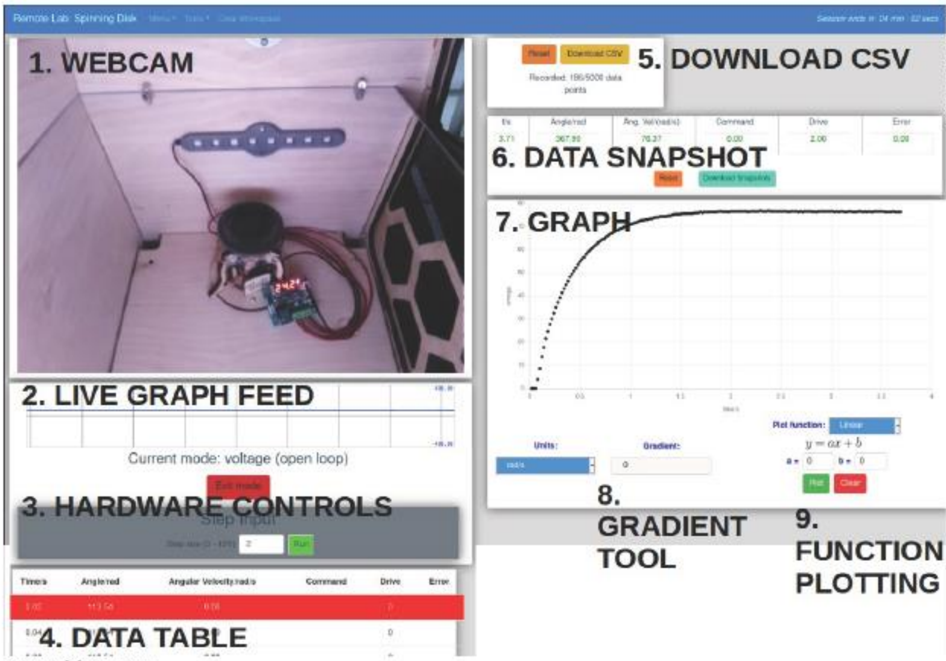
Please select at least 1 answer(s).

- ☐ surprised
- ☐ amazed
- ☐ astonished
- ☐ curious
- ☐ inquisitive
- ☐ interested
- ☐ happy
- ☐ joyful
- ☐ excited
- ☐ confused
- ☐ muddled
- ☐ puzzled
- ☐ nervous
- ☐ anxious
- ☐ worried
- ☐ frustrated
- ☐ irritated
- ☐ dissatisfied
- ☐ bored
- ☐ monotonous
- ☐ dull

Can you reflect further on why you had the above emotional responses?

Rate the following user interface components in terms of how EASY TO USE they were. If you did not use this component then select N/A * *Required*

8 / 18



Please don't select more than 1 answer(s) per row.

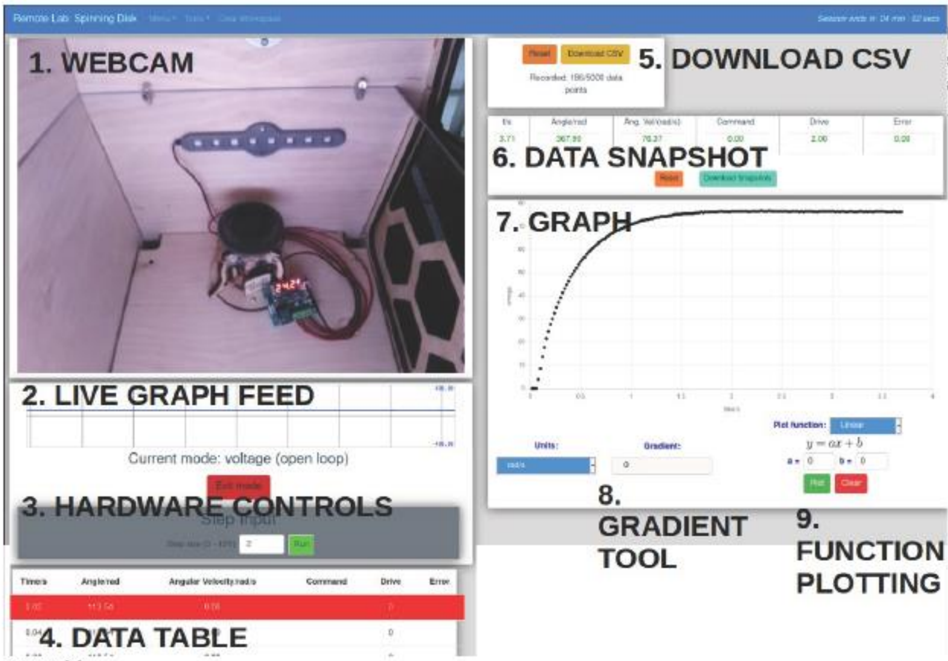
Please select at least 10 answer(s).

	N/A	Very difficult	Difficult	Neutral	Easy	Very easy
1. Webcam video feed						
2. Live graph feed						
3. Hardware controls						
4. Data table						
5. Download CSV tool						
6. Data snapshot						
7. Graph						
8. Gradient tool						
9. Function plotting						
Measuring tools: ruler and protractor						

Rate the following user interface components in terms of their USEFULNESS for completing your practical work.

Please don't select more than 1 answer(s) per row.

9 / 18



Please select at least 10 answer(s).

	Useless	Mostly useless	Neutral	Mostly useful	Vital
1. Webcam video feed					
2. Live graph feed					
3. Hardware controls					
4. Data table					
5. Download CSV tool					
6. Data snapshot					
7. Graph					
8. Gradient tool					
9. Function plotting					
Measuring tools: ruler and protractor					

If you found any component particularly easy or difficult to use can you explain why?

10 / 18

Are there any changes to the above components or additional features that would have made the user interface more useful for your practical work?

A separate simulated version of the remote lab was also accessible during your practical work. Rate how USEFUL you found this simulated environment. * *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 1 answer(s).

	Useless	Mostly useless	Neutral	Mostly useful	Vital
Simulated hardware	☐	☐	☐	☐	☐

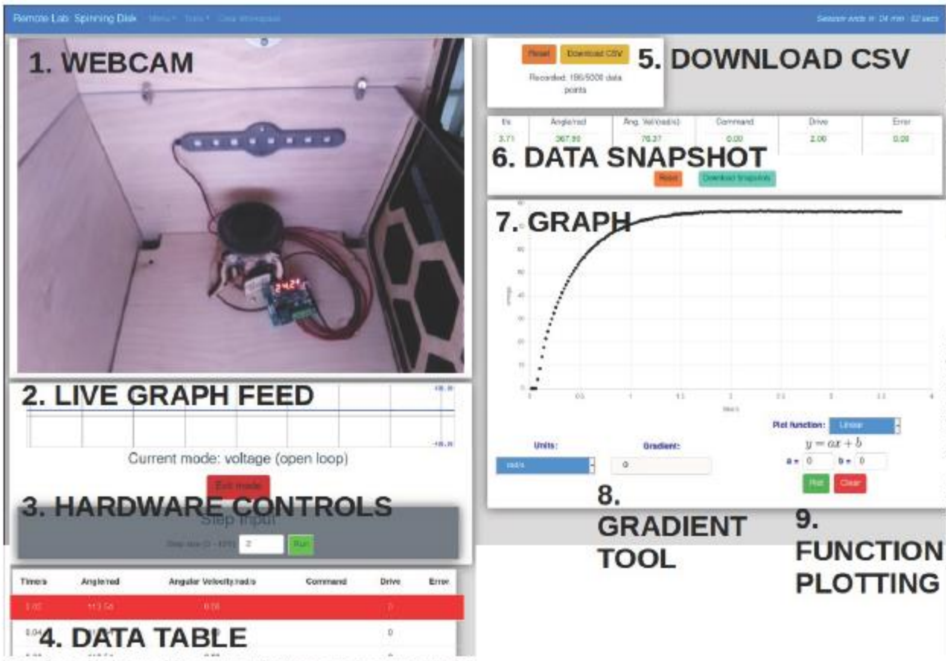
Rate the following user interface components in terms of how much each element contributed to your feeling of BEING IN CONTROL OF YOUR OWN EXPERIMENT. If you did not use the component then select N/A.

Please don't select more than 1 answer(s) per row.

Please select at least 10 answer(s).

	N/A	Completely out of my control	Mostly out of my control	Neutral	Mostly under my control	Completely under my control
1. Webcam video feed	☐	☐	☐	☐	☐	☐
2. Live graph feed	☐	☐	☐	☐	☐	☐
3. Hardware controls	☐	☐	☐	☐	☐	☐
4. Data table	☐	☐	☐	☐	☐	☐
5. Download CSV tool	☐	☐	☐	☐	☐	☐

11 / 18



Do you feel that any part of the user interface behaved in a way that was not expected?

Rate the following statements in terms of how much you agree or disagree with them. ✱ *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 19 answer(s).

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
It was easy to navigate the user interface	☐	☐	☐	☐	☐
It was easy to collect the appropriate data.	☐	☐	☐	☐	☐
The user interface allowed me to explore further than the lab sheet required.	☐	☐	☐	☐	☐

12 / 18

I feel more confident with the course content having completed this practical.	☐	☐	☐	☐	☐
I believe the results of my experiment are accurate (i.e. I calculated the true value of quantities)	☐	☐	☐	☐	☐
I believe that the results of my experiment are reliable (i.e. if I repeated the experiment I would calculate the same thing).	☐	☐	☐	☐	☐
I would prefer to be in the same location as the hardware.	☐	☐	☐	☐	☐
I felt motivated to explore my own ideas whilst completing the practical.	☐	☐	☐	☐	☐
I am satisfied that I have learned what I should have during the practical work.	☐	☐	☐	☐	☐
I enjoyed the remote lab practical work.	☐	☐	☐	☐	☐
I intend to return to the remote lab to explore further.	☐	☐	☐	☐	☐
I felt like I was in control of the hardware.	☐	☐	☐	☐	☐
I wanted to be able to do more with the hardware than the UI allowed.	☐	☐	☐	☐	☐
The user interface behaved as I expected it to.	☐	☐	☐	☐	☐
A remote lab is not real practical work.	☐	☐	☐	☐	☐
There was only one way to complete the tasks.	☐	☐	☐	☐	☐

13 / 18

I would be happy to complete more practical work as a remote lab.	☐	☐	☐	☐	☐
I would be happy to use a simulated/virtual environment for future practical work.	☐	☐	☐	☐	☐
I would be confused by the use of real and simulated hardware in the same remote lab environment.	☐	☐	☐	☐	☐

The following statements compare remote lab work to proximal (in lab where you can physically interact with the hardware) lab work. You are being asked to decide whether you believe it is easier to accomplish the following tasks with a remote or proximal system. * *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 14 answer(s).

	Much easier with REMOTE lab	Easier with REMOTE lab	No difference	Easier with PROXIMAL lab	Much easier with PROXIMAL lab
Collecting data.	☐	☐	☐	☐	☐
Managing my lab time.	☐	☐	☐	☐	☐
Comparing to theoretical models.	☐	☐	☐	☐	☐
Designing my own experiment.	☐	☐	☐	☐	☐
Analysing data to reach conclusions.	☐	☐	☐	☐	☐
Setting up equipment.	☐	☐	☐	☐	☐
Investigating unknown quantities.	☐	☐	☐	☐	☐
Exploring beyond the scope of the practical work.	☐	☐	☐	☐	☐
Working with others.	☐	☐	☐	☐	☐
Measuring accurately.	☐	☐	☐	☐	☐

14 / 18

Presenting data	☐	☐	☐	☐	☐
Controlling variables	☐	☐	☐	☐	☐
Working safely	☐	☐	☐	☐	☐
Use multiple senses to gather information.	☐	☐	☐	☐	☐

What have you found to be the main ADVANTAGES of completing practical work as a REMOTE lab?

What have you found to be the main DISADVANTAGES of completing practical work as a REMOTE lab?

Was there anything that stopped you from being able to complete any part of the practical work?

If you found that a practical result differed from that calculated from theory, which statement best represents your belief in the reason for that difference? * *Required*

- There are errors in the remote lab
- I cannot control the remote lab sufficiently
- The theory is not correct
- The theory is limited
- I must have made a mistake
- The remote lab does not measure accurately

15 / 18

- I found no difference between practical and theory
- Other

If you selected Other, please specify:

All things considered and assuming a return to normal circumstances - no pandemic restrictions - please answer the following question. * *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 1 answer(s).

	All REMOTE	Mostly REMOTE	No preference	Mostly in PHYSICAL lab	All in PHYSICAL lab
I would prefer to complete future practical work as					

Do you have any other thoughts or feelings about the remote practical work and user interface that you would like to share?

Approximately how many hours did you spend directly on the remote lab system? (i.e. not including writing up your report or any time spent outside of the remote lab interface itself) * *Required*

16 / 18

Page 5: Volunteer for follow-on interview

We are looking for volunteers to take part in semi-structured interviews to provide additional feedback on their experience with the remote lab system. The focus is entirely on how you used the remote lab and is ***not*** a test of your knowledge of the 'Controls and Instrumentation Engineering 3' course content. The feedback session will be over MS Teams and will involve you demonstrating how you used the remote lab itself whilst explaining your process. You may be asked to complete short tasks, for example, to plot the step response of the motor in speed mode whilst describing what you are doing. You may also be asked to expand on some of your opinions from this survey. The session will take around 10 minutes.

I am happy to volunteer for a follow on interview * *Required*

○ Yes ○ No

If you answered 'Yes' above then please provide your university email address so that we can contact you to organise the feedback session. Providing your email address does not guarantee that you will be contacted and you will have the opportunity to confirm your participation once emailed.

Please enter a valid email address.

17 / 18

Page 6: Final page

Thank you very much for participating in this survey. If you have any further questions about the survey or remote labs in the School of Engineering or decide that you want to remove yourself from the study then please contact either Professor Timothy Drysdale (Timothy.Drysdale@ed.ac.uk) or Mr David Reid (D.P.Reid@sms.ed.ac.uk).

Key for selection options

26 - Approximately how many hours did you spend directly on the remote lab system? (i.e. not including writing up your report or any time spent outside of the remote lab interface itself)

< 1 hour
1-2 hours
2-3 hours
4-5 hours
> 5 hours

H. UI Research Article

Original article

IJMEE

International Journal of Mechanical Engineering Education
2022, Vol. 50(4) 828–848

DOI: 10.1177/03064190221081451
journals.sagepub.com/home/ijj
SAGE

Open-source remote laboratory experiments for controls engineering education

DP Reid[1], J Burridge[2], DB Lowe[2], and TD Drysdale[1]

Abstract

We describe and evaluate the use of a remote laboratory with approximately 250 third-year students in the 'Controls and Instrumentation Engineering 3' course in the School of Engineering at the University of Edinburgh. The remote laboratory was created using the first iteration of a new open-source infrastructure. A key component of the implementation is a secure video and data relay that permits experiments to be hosted on restricted networks and accessed from a wide range of browsers, so as to minimise the technical support burden on course staff. A heterogeneous set of twelve spinning disks driven by DC motors were provided to students via an online catalogue permitting instant and exclusive access to an individual experiment for up to 90 min a session. The student experience was evaluated via a survey, which comprised both qualitative and quantitative questions. Survey participation was optional and included some in-depth questions targeting the students' experience from a UX perspective. The link between quantitative and qualitative results allows us to draw useful conclusions about the experiences of the students participating in the survey. Our main findings are that the overall experience was positive for the survey respondents, and that they perceive a distinction between having *control over the equipment* and *control over the task*. Our work indicates there is a need to better understand the implications of students making this distinction, and the effect it will have on future designs of remote laboratory tasks, user interfaces, and evaluations.

[1]School of Engineering, The University of Edinburgh
[2]School of Computer Science, The University of Sydney

Corresponding author:
DP Reid, School of Engineering, The University of Edinburgh.
Email: s0967663@ed.ac.uk

Keywords
remote laboratory, engineering, education, student perceptions, practical work, open-source, usability

Introduction

The prospects for low-cost, open-source remote laboratory experiments have recently been enhanced by the wide availability of low-cost single-board computers capable of streaming video (such as Raspberry Pis), and low-cost electronic prototyping ecosystems with strong communities (such as Arduino microcontrollers). These devices have already fostered modest adoption of electronic hardware exercises in a number of educational settings.[1–4]

While these low-cost experiments have unequivocally lowered the barrier to using remote laboratories in more educational institutions, it is clear that additional factors must be limiting progress towards mainstream adoption, such as (1) even relatively low-cost physical equipment requires time, space and money to develop (or procure) and then host, so sharing experiments is required to maximise the value of the investment; (2) to connect experiments to users, at scale, with acceptable trade-offs between convenience, security, performance and reliability is a non-trivial task so some providers may have chosen to limit their networks to co-operative users within trusted groups to side-step some of the more difficult challenges, thus introducing a human gatekeeper into the scaling process; (3) personnel associated with the development may have been grant-funded or on other temporary posts such as fixed-term roles, internships, studentships, visiting positions, and no longer available to expand or extend either the experimental provision or the management software; (4) the architecture and implementation methods may have been chosen to allow timely delivery of a defined feature set, and present difficulties to subsequent extension and modification efforts; (5) educators who are not naturally enthusiastic about remote laboratories may remain to be convinced of their value, requiring those in the educational community who are already enthusiastic to further develop research into how student skills and cognitive developments are affected by remote (rather than proximal) interaction with physical equipment.

Together these reasons suggest that there remains a significant opportunity to investigate new means of providing remote laboratories that address barriers to adoption, long-term sustainability, and to address discussions about inclusivity, accessibility and decolonialisation of curricula (via permitting customisation to suit local needs). To a certain extent, in an ever more digital world, it can even be argued that first-hand experience of remote equipment is a directly relevant skill for graduates. Thus, the value of remote laboratories goes beyond mere replacement of proximal labs and can instead be framed as a complementary approach with its own beneficial characteristics which cannot be obtained in proximal labs (and which the community is still exploring the extent of).

We use the term "complementary" to highlight the different characteristics inherent in digital education interventions, both those already realised and others as yet to come, that go beyond merely supplementing existing face-to-face education with an online delivery mechanism. For example, the opportunity for wide-spread asynchronous interactions with automated and artificially intelligent education systems creates the opportunity for post-humanistic communities of practice,[5] which follow from new types of interactions with[6] or without[7] the involvement of artefacts, new types of automated assessment,[8] and new methods of evaluating and discovering future capabilities.[9] A fuller discussion of this next stage in the development of remote laboratories is outside the scope of this Paper.

This Paper introduces the first usage in assessed coursework of the first stage of a new remote lab software infrastructure that is ultimately intended to address many of the challenges identified above. A set of twelve spinning disk experiments were used to provide assessed remote-laboratory coursework to a cohort of 250 mechanical and electrical engineering students on a controls and instrumentation course (worth 80% of the course mark, with the course worth 10 of 120 credits for the year). We include an evaluation of the student experience via quantitative and qualitative analysis of survey data from participants. We explore the factors that affect student usability ratings of the remote laboratory system. In particular, we are interested in how the user interface (UI) design can impact student perceptions of autonomy and belief in predictable control over their practical work. We also present results on student perception of theory-measurement disparity in this remote laboratory and perceptions of the skills disparity between remote and proximal (hands-on) laboratory experiences. Since we were using a pseudonymous data collection protocol, it was out of the scope of this study to explore any potential differences in experience between the mechanical and electrical cohorts, although we note that the intended learning outcomes were the same for all students.

The Paper is structured as followed. Section Two describes the remote laboratory infrastructure, experiment design and student task; Section Three describes the evaluation methodology. Section Four presents both quantitative and qualitative results. We discuss what we learn from the survey data, its context, and limitations in Section Five.

Remote laboratory infrastructure

Remote laboratories require the connection of experiments and students, and the management of those connections. The infrastructure used to manage the connections can determine key aspects of the overall experience for both staff and students, and affect the long-term growth and sustainability of the laboratory. Over the last two decades, there has been a progression from projects being funded to explore what can be done with remote laboratories, to the creation of remote laboratories which are intended to run in a self-sustaining manner beyond the project funding date, and this is described in more detail in the following section, Section 2.1. The degree of success in sustaining long term operation of a remote laboratory is in our view tightly linked to the degree of automation in the system. Creating long-term sustainable approaches that are adopted into mainstream use remains a goal for workers in the field.

It is usually straightforward to set up a single remote experiment, for a single user. However, as the number of users and the number of experiments increase, the benefit of automation becomes apparent. The staff resource required to operate at an institutional scale (hundreds of experiments, thousands of users) is likely to be unaffordable without a significant degree of automation. The system-to-administrator ratio is a rough indicator of the efficiency achieved through automation and can range from 2:1 for bespoke, less-automated services, through to 2500:1 for internet-scale systems with high degrees of automation.[10] Note that many modern services comprise smaller components, each counting as a system that an administrator must manage. Many existing remote laboratories are likely to be categorised as bespoke systems with a low system-to-administrator ratio at present (in other words, available staff time is a strong constraint on how many experiments and users can be supported). Automation in remote laboratories offers benefits in a number of ways. For example, automating checks of the quality of results from experiments. If staff are alerted to problems as they occur, and do not need to dedicate time to manually checking the equipment, they can spend it more productively elsewhere (e.g. developing new experiments). Another example is the quality of the user experience, which is adversely affected if there are accidental or malicious interruptions to sessions. While a number of early laboratories are understood to have "made-do" with shared credentials and an honour-based booking system, automated access controls are strongly preferred because it stops well-intentioned users from inadvertently joining another's session, and reduces the opportunity for malicious interventions. There is a potential tension between introducing automation and permitting extensibility, which remains to be resolved.

Selected history of remote laboratories

Remote laboratories may not yet be mainstream education activities, but great progress is being made in this direction. The evolution of the field from exploration to sustained delivery can be visualised as shown in Figure 1, where the approximate start and end dates of several major remote laboratory projects have been plotted, based on reports on their work.[11–17] Not all remote laboratories have been included, for the sake of brevity. However, with the set chosen, there are already enough data points to see a clear pattern emerging. In the earlier years, funding was typically for research projects, while from 2010 onwards the field had matured enough that there were projects funded primarily for the realisation of laboratories for sustained usage beyond the lifetime of the initial funding. Some laboratories (not included) have not been able to sustain running beyond their funded periods. Those that do continue, are usually limited to maintaining what they have. It would be preferable to see self-sustaining growth of remote laboratories so as to achieve mainstream usage. What characteristics might such system(s) have?

A mixture of closed and open-source projects have emerged to date. A benefit of the closed-source approach is that capabilities developed remain unique to the organisation, at the cost of increasing the difficulty of collaborating or federating with other entities. Open-source projects are attractive from a community development point of

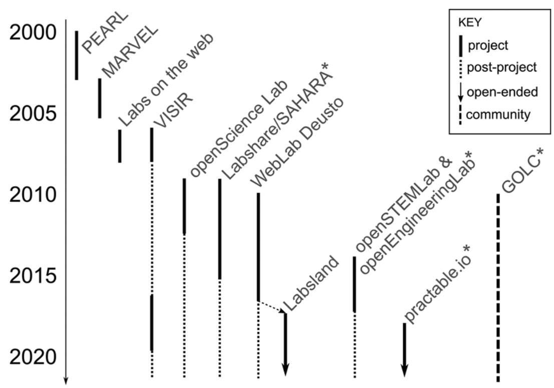


Figure 1. Approximate timelines for a selected set of remote laboratory initiatives. Asterisk denotes activities currently or previously associated with the Author(s) in some way.

view. A known difficulty in open-source projects is achieving sustainability, both in terms of development effort, community participation and longevity of project benefits. A strong community is often a key step in reaching sustainability. We have already identified in the previous section that automation is a key factor in determining the success of a remote lab in practice, so these two requirements (automation, sustainability) can lead us to infer that a successful infrastructure will likely combine elements such as:

- (S1) sufficient performance and reliability that users are happy to adopt the system for their institution;
- (S2) features and tools adequate to support the delivery of educationally useful activities, by staff who do not necessarily have specialist digital skills;
- (S3) sufficient automation and systems integration to permit sustainable operation within the limits of support staff resource available at institutions;
- (S4) a total-cost-of-ownership profile that is a good fit to on-going operating budgets such as annual laboratory maintenance funding (rather than requiring significant one-off capital funding)
- (S5) adequate opportunity to extend the system to meet emerging needs;
- (S6) the opportunity to federate with others to access a greater diversity of experiments, increase through-put for time-sensitive activities, and increase reliability by sharing risk across multiple sites;
- (S7) the opportunity for institutions to create proprietary *content to establish unique competitive positions (but encourage openness e.g. through eventual publication of designs and software, perhaps with an embargo period).*
- (S8) an ecosystem of academic and commercial service providers, to meet the operational needs of users who do not wish to run their own servers;

- (S9) an infrastructure with a decentralised architecture and copy-left open-source licensing to avoid the business risks to users of vendor lock-in, and fragmentation risks to the community;
- (S10) interoperability with other remote laboratory systems, virtual learning environments, and third-party tools such as assessment systems.

Proponents of remote laboratories have long anticipated the development of federated systems. For example, the Global Online Labs Consortium and its various Special Interest Groups. Amongst active workers in this area now are the commercial entity LabsLand, which is building a federated community using the open-source software developed by the WebLand Deusto academic activity. Another is the open-source *practable.io* system from the University of Edinburgh, which we used for the work in this Paper,[18] which is being designed from the ground-up for federation.

Given the number of ways in which the elements listed above could be achieved, we expect it is likely that a similar situation will eventually arise as we have in the more mature world of virtual learning environments: multiple complementary and distinct approaches and suppliers, ranging from turn-key commercial services to customisable services that are self-deployed, each suiting their own audiences and interoperable so far as possible with third-party extensions. We now turn to the specifics of the system we used for the work in this Paper.

Practable.io infrastructure

In this section we concentrate on a description of the requirements and implementation of the video and data relay system, and the interim booking system. The specifics of future developments such as advance booking, experiment management, federation, de-centralisation, and integrations with virtual learning environments and identity servers are not discussed because they are the subject of ongoing work.

Our main requirement is for a system that would be easy to adopt by other academics in due course, who may not have the permission or make custom network arrangements, have funding for expensive streaming computers, or have support staff available to help students get set up with custom software. We also require the relay system not to have strong opinions about the choice of audio, video or data stream formats, or how the streams are arranged, combined, and booked. These translated into some specific technical requirements for the video and data relay system, which also address aspects of the system level requirements, as follows:

- (R1) work behind restricted firewalls with multiple levels of network address translation, and that only permit outgoing connections to port 443/tcp (S2,S4)
- (R2) support switching relay servers mid-experiment so that spot-priced (i.e. cheaper) server instances can be used (S4)
- (R3) support authenticated access to secure sessions against unwanted interruptions (S1,S3)

- (R5) work with any browser a student is likely to use by default, without requiring any installations of plug-ins or extensions (S3)
- (R6) require only a headless (i.e. command-line) operating system to run, so that low-cost single-board computers can be used as the streaming devices (S1,S4)
- (R7) be able to cope with the relay-server changing mid-experiment (this translates into a requirement to be able to change the destination of the stream, without re-connecting to any USB cameras or microphones, because USB reconnections can be unreliable on some systems) (S1,S4)

A combined data and control flow diagram for the video, data and control connection between an experiment and a user is shown in Figure 2. We are using Raspberry Pi 4 (4GB) single board computers using a linux-based operating system, which uses *systemd* for managing system tasks. The single-board computers are configured either manually or via *ansible.* When located behind a firewall, a custom *shellrelay* system is used to connect securely to the experiments and install/update the configuration, scripts and applications.

Once configured, on booting the *sessionhost* application is started by the operating system. Then the *sessionrules* script is run to configure the *sessionhost* via its HTTP API. This is preferable to using configuration files because of the flexibility it provides in large scale deployments.[19] Using the API, the video relay connections can be changed to new routings or new relays, as often as every second if desired. In a stress test of the experimental side of the system, the video was directed to a new video relay location every second for 24 h, with no appreciable deterioration in the latency of the video feed transmitted from the experiment. This is quicker switching than required for spot-priced servers, which typically receive warning of a few minutes that the relay instance needs to be shut-down. The mechanism for hot-switching is not shown in Figure 2 for the sake of clarity.

Next, *systemd* starts a service which uses *ffmpeg* to stream video in MPEG-TS format to the *http://* endpoint of *sessionhost* The *sessionhost* is intended to handle frame-based streaming formats. It relies on the burst nature of streaming data at each frame to ensure data for each frame is flushed through the buffer and does not require any understanding

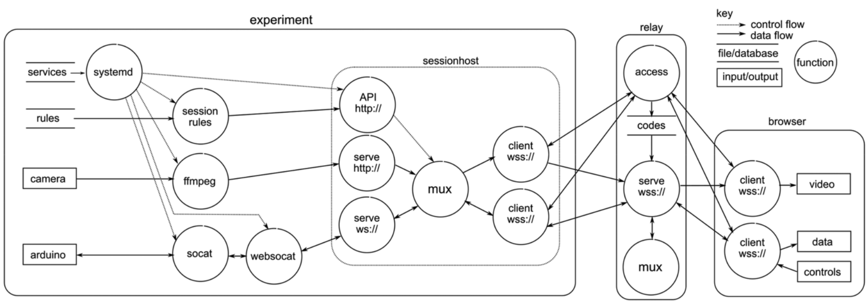


Figure 2. A Combined data and control flow diagram of a user connecting to an experiment via a relay, to obtain video, data and send control (credential flows are not shown for clarity).

or introspection of the video data itself. Audio streams can be sent by running an *ffmpeg* instance which connects to a USB microphone. MPEG-TS streams permit audio and video to be multiplexed simply by interleaving the audio and video packets. We disabled the audio for our initial usage of the system because the experiments were in a public space. Contact microphones are an attractive future possibility for capturing the sounds of the experiments but suppressing ambient conversations.

Data and control messages are passed between the *Arduino* microcontroller and *sessionhost* via scripts which run *websocat* and *socat* in a manner similar to the connection that the Arduino online-IDE uses for programming. These are routed by a separate *sessionrelay* endpoint so that there is no need for the user interface to introspect the video stream to extract data and control packets.

Connections to the relay are secured by JWT tokens. For the experiments, their connections are valid until their next scheduled physical or virtual maintenance period (perhaps a few months or a year). For users of the spinning disk experiments, the tokens last from 5–90 min depending on the length of session they chose. Longer sessions are possible, but not required by this course. Whether by an experiment or a user, their token is submitted to the relay *access* point via an HTTP PUT request The *access* application assigns a routing on a relay server instance. Any subsequent tokens submitted for the same experiment are assigned to the same relay server. This layer of indirection is included to permit spare capacity management, and transparently accommodate failure of the *serve-and-mux* services. The authentication to the *serve-and-mux* task is via a one-time code which is valid once within a 30 s window. The connection is terminated at the end of the access period specified in the original authentication token. This approach allows the booking system to pass a time-limited JWT token to the user interface, which then uses that token to connect, and reconnect as required if the network transport is interrupted e.g. by poor mobile signal, but only for the duration of the booking. This avoids reverting to the booking system for new authentication when a connection drops unexpectedly, which would have made retaining experimental state difficult. The code which permits access to the *serve-and-mux* task is passed in the response to the PUT request, and even if this is intercepted and decrypted the consequence is limited to a single-time limited connection. The *relay* also knows about *read* and *write* permissions, which are specified in the JWT tokens submitted to the relay *access* point. The booking system only gives users read access to the video and audio stream, while experiments get read and write access to both. When a group of users is using the same experiment, their read-only access prevents a malicious user from injecting unwanted video or audio frames into the stream. On the other hand, users are allowed to send control commands, and potentially conduct intra-group messaging if the experiment and user interface support it, although it is likely to be preferable for human-to-human messaging to be conducted in the standard messaging tools supported by the educational institution.

Spinning disk experiment

We wished to build a set of a spinning disk experiments that could replace a pre-existing proximal laboratory exercise. The pre-existing laboratory was developed around the Quanser Qube spinning disk experiment, and the associated coursework had been evolved from the original Quanser workbooks by the course team. Custom spinning disks of different weights had previously been manufactured in-house, and used in the in-person laboratory for a number of years, to provide different inertia systems against which students could explore and apply the principles of simple first and second order systems and their control.

For the remote version of the experiment, twelve separate "kits" were manufactured in-house. They included a DMN37BA brushed DC motor (12V), a 500CPR optical encoder, custom built weighted disks (ranging from 43–110 grams), as well as laser cut wooden stands and boxes to house the physical and electronic hardware. An off-the-shelf webcam was used for some views (see Figure 3) as well as the Raspberry Pi camera for an alternative, top-down view on the experiment. Each experiment had only one camera, although multiple camera views are possible with the system. Figure 3(a) shows the user interface students used to control and analyse the remote hardware. Figure 3(b) provides an annotated view giving the name of each component as it is referred to throughout this paper.

The hardware design files, firmware and software for the experiment, as well as the core system, are already available as open-source software at https://github.com/practable, primarily under the GNU Affero General Public License, and no permission is required to adopt the system.

The lab was performed individually, with students producing individual lab reports. The lab was done by choosing the hardware mode, setting parameters (voltage, PID parameters), and then collecting and analysing angular speed or displacement data from the hardware, so as to characterise the system and calculate appropriate coefficients for tasks such as position and velocity control of the disk. This analysis could be performed with tools available in the UI, although the raw datasets could also be downloaded for analysis offline if preferred.

Students followed a series of predefined workbook tasks that asked them to measure DC motor characteristics and then utilise these measured values to compare the theoretical response of a PID controller and plant to step and ramp inputs.

Task 1: Students were asked to confirm whether the motor obeyed the linear relationship $\omega = KV$. They used the motor in open-loop mode to apply a voltage to the motor. The angular velocity could be identified using the 'Snapshot' or 'Graph' tools, displaying the raw data in the 'Table' or by downloading and analysing in external software.

Task 2: Involved students using the DC motor in open-loop mode to apply a step voltage. They collected motor angular speed data and either used the UI tools or downloaded data for external use in order to identify the gain and time constant of this plant (DC motor + disk) for 2 different weighted disks.

Task 3: Students tuned a PID controller to apply a step position to the DC motor within set constraints.

These tasks could all be achieved using the data streaming from the hardware – angular position, angular velocity, time, motor voltage and PID error signal. They had several modes of visualising the data: UI Table, Snapshot, Live graph (plots continuous data stream) and Graph (plots recorded data) tools as well as the ability to download data for analysis externally. Additional tools, such as the ruler overlay (Measuring tools), gradient and function plotting tools could help them explore features of the motor response whilst not being necessary to complete the tasks.

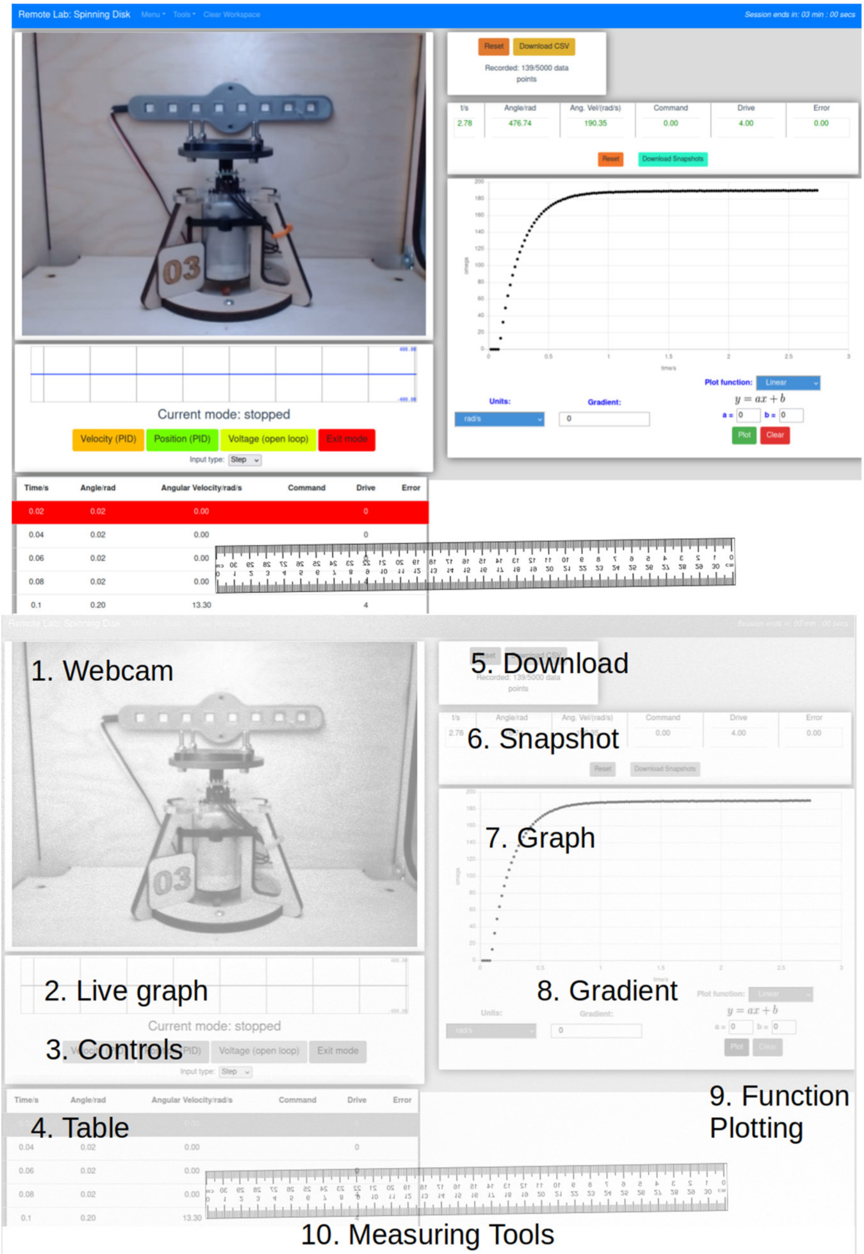


Figure 3. a) The user interface of the controls engineering remote laboratory, b) with annotated components.

Methodology

Student perceptions were collected via a voluntary, online survey accessed after the practical work was completed. The survey comprised the 'User Experience Questionnaire' (UEQ)[20] as well as custom survey questions.[21] Likert style questions have been analysed in a quantitative manner to identify statistical significance as well as correlations between ratings. Open text responses have been used to identify potential updates to the remote laboratory system as well as to perform a coded content analysis to highlight any trends in the feedback received.

The UEQ provides a quantitative measure of usability and is used to evaluate the overall user experience as it relates to its six measured attributes (see Figure 4). The custom survey provides a broader context for the usability results, helps isolate factors that may contribute to the UEQ attributes, as well as evaluating individual components of the remote lab user interface.

To test whether responses are significantly different from 'Neutral' (0), the 1-sample Wilcoxon signed-rank test is used. Correlations are calculated using the Kendall tau rank correlation coefficient. Statistical significance is set at the 5% level for tests (unless stated otherwise) and common effect size bands are used:[22] ≥0.50 strong; ≥0.3 medium; ≥0.10 weak. Effect size is calculated from the Z statistic divided by the square root of number of samples.

Of the approximately 250 students enrolled on the `Controls and Instrumentation Engineering 3' course, only $N = 24$ complete surveys were received. Three respondents were determined to have provided inconsistent responses; therefore, the following analysis is performed on $N = 21$ student surveys. With only a ≈10% response rate the data is likely affected by selection bias - students who have had particularly extreme experiences are perhaps more likely to respond.

Figure 4. Usability attributes measured by the User Experience Questionnaire (UEQ). Reproduced from https://www.ueq-online.org/.

Results

User experience and usability

Scores for each UEQ attribute can range from −3 to + 3; however, scores outside the range −2 to + 2 are unlikely according to the authors of the UEQ. Mean scores on each of the UEQ attributes between $-0.8 < \bar{x} < 0.8$ represent neutral responses.

Mean ratings for each UEQ attribute were as follows: attractiveness = 1.13, perspicuity = 1.51, efficiency = 1.08, dependability = 0.76 (1.06 with outliers removed), stimulation = 0.88 and novelty = 1.17. Students have rated the remote lab and UI positively across all attributes, with the exception of a neutral score for dependability.

Although the UEQ attempts to measure usability across six independent attributes, we found, in particular, a strong correlation ($\tau = 0.55$) between attractiveness and dependability. This hints at the potential importance of 'predictable control' over the remote lab system and will be explored further when discussing how UI components were rated in terms of making students 'believe they were in control of their experiment' (section 5.3).

Emotional response. Students were asked to report their emotional state after completing the practical work. By refactoring all the emotional options into a binary choice of 'positive' or 'negative' a Chi-squared test was used to test whether the distribution was significantly different from that expected through random selection. In this case it was found that there was a significantly positive response ($\chi^2 = 7.16$, df = 1, $p < 0.01$).

UI ratings. Table 1 provides an overview of how user interface components were rated by students in terms of 1) ease of use, 2) usefulness in completing the practical tasks and 3) how much the component made them feel in control of their experiment. All components were rated on a 5-point scale from −2 to + 2. As may be expected from the high score for

Table 1. Ratings of the user interface components in terms of how easy to use they were, their usefulness for completing tasks, and how in control of the experiment they made the user feel. Rating is the median response between + 2 and −2. Effect sizes of test of difference from 0 rating is shown: **large effect size, *medium effect size, otherwise a result is non-significant or a weak effect size.

Component	'Easy to use' rating	'Usefulness' rating	'Feeling in control' rating
Webcam	+ 2**	+ 1*	−1
Live graph	+ 1**	+ 1**	0
Controls	+ 1**	+ 2**	+ 1**
Table	+ 1**	+ 1	+ 1**
Download	+ 2**	+ 2**	+ 2**
Snapshot	+ 2**	+ 2**	+ 2**
Graph	+ 2**	+ 2**	+ 1.5**
Gradient	+ 1.5**	0	+ 1**
Function plotting	+ 1**	+ 1**	+ 1**
Measuring Tools	0	−1**	+ 0.5

Table 2. Select statements from the student survey. Rating was between Strongly agree (+2) and Strongly disagree (-2). Effect sizes of test of difference from 0 rating is shown: **large effect size, otherwise a result is non-significant or a weak effect size.

Statement	Median rating
It was easy to collect the appropriate data	+1**
I feel more confident with the course content having completed this practical	+1**
I believe the results of my experiment are accurate	+1**
I believe that the results of my experiment are reliable	0
I felt motivated to explore my own ideas whilst completing the practical	0
I am satisfied that I have learned what I should have during the practical work	+1**
I enjoyed the remote lab practical work	+1**
The user interface allowed me to explore further than the lab sheet required	0
I would prefer to be in the same location as the hardware	+1**
I felt like I was in control of the hardware	+1**
I wanted to be able to do more with the hardware than the UI allowed	0
The user interface behaved as I expected it to	+1
A remote lab is not real practical work	−1
I would be happy to complete more practical work as a remote lab	0

perspicuity, the median rating of 'ease of use' for all components was positive, with the exception of the 'Measuring tools'. Considering the user interface was designed to accomplish the intended practical tasks it is not surprising that most components have been rated positively in terms of 'usefulness'. 'Feeling in control' will be discussed further in the section 5.3.

Likert statements. Students were asked to rate their agreement with several statements regarding their overall experience of the practical work. Table 2 presents select results relevant to the discussion in this paper.

Perceived theory-measurement disparity in remote labs

It was clear in the development and deployment of the remote laboratory that there was going to be a disparity between the ideal theory of control systems that students were learning, and the reality of the low-cost, DC motor control system that they were being asked to use. This was because of significant bearing friction and a large dead-band. This was anticipated and therefore students were asked, if they found that theory and reality didn't align, what they believed the cause of this difference to be. We can see from Figure 5 that there are a range of beliefs expressed regarding the conflict between theory and reality during the practical work. The most selected option was '*the theory is limited*', although approximately half (52%) of all responses involved the lab not being sufficient in some regard. The two 'Other' responses were: '*There are losses in the practical system that are not present in the theoretical calculations*' and '*I'm unsure as to whether the theory is insufficient or the lab had problems.*'

Figure 5. Distribution of student beliefs in why theory and practical measurements differed.

Interestingly, no correlation appears to exist between a user's overall usability score (UEQ attribute average) and their belief in the reason for differences between theory and reality. Similarly, no clear correlation exists between the dependability attribute and student belief in the reasons for the difference. Additionally, there was no correlation between the number of mentions of technical issues (see Section 4.4) in the qualitative responses and the belief in the reasons for the difference.

Skills perceptions

Students were asked to rate on a scale from 'Much easier with remote lab' (+ 2) to 'Much easier with a proximal lab' (-2) the skills-based tasks presented in Table 3 (based upon Feisel and Rosa).[23] We can see from Table 3 that there is a perception amongst respondents to the survey that one or other lab mode can more easily allow the demonstration of certain skills. However, it is also important to note that there is a core of fundamental lab skills that survey respondents believe are just as easily accomplished with a remote lab as a proximal, in-lab experience.

Content analysis

In order to explore whether any themes arise in the open feedback from students a content analysis was performed. First, student responses were coded as either positive, mixed or negative based upon their response to the 'emotions felt after completing the practical work' survey

Table 3. Student perceptions of whether skills-based tasks are easier to accomplish with a remote or proximal lab experience.

Skill easier with REMOTE lab	No significant difference	Skill easier with PROXIMAL lab
Collecting data	Controlling variables	Use multiple senses to gather information
Managing my lab time	Measuring accurately	Working with others
Working safely	Comparing to theoretical models	
Setting up equipment	Designing my own experiment	
Presenting data	Analysing data to reach conclusions	
	Exploring beyond the scope of the practical work	
	Investigating unknown quantities	

question. Then, each basic unit (normally the full response to a question but potentially individual statements if separate points were being made) was coded, if appropriate, as a **technical issue** (a bug or other aspect of the lab which "broke"), a **design issue** (a limitation in the design that caused problems), or a wider **teaching issue** (something they commented on which did not have to do with the lab equipment or UI, but how it was used/integrated in teaching). Two independent coders annotated the feedback and reached a consensus.

We can see from Table 4 that users considered to have had negative experiences were more likely to provide comments than students with positive experiences. All students with a negative response gave at least one reason for their response within the feedback. The nature of the survey was heavily skewed towards identifying problems with the UI and remote laboratory, so this distribution is not unexpected. We can also see that technical issues were the most mentioned in negative respondents, suggesting that technical issues may be the most impactful on student experience. This is backed up by the negative correlation between the number of counts of technical issues mentioned and the attractiveness UEQ attribute ($\tau = -0.4$).

There was also a strong negative correlation between mentions of teaching issues and the novelty rating ($\tau = -0.51$), reminding us that the remote lab system cannot be evaluated outside of the teaching context.

Discussion

Improvements

One of the purposes of seeking student feedback is to improve the remote laboratory experience. To that end, open response questions within the student survey were used to make upgrades to the remote lab user interface. A number of comments relating to the layout of components and relative positions of buttons has led to the development of a user-controlled, reconfigurable UI. Students can now independently control the position of each component, scroll each half of the screen independently and choose from

Table 4. Average count of issues mentioned in the open text response questions. Issues are grouped into technical, design and teaching for respondents rated as having positive, mixed or negative experiences.

	Positive	Mixed	Negative
TECHNICAL	0.29	1.30	2.00
DESIGN	0.57	2.20	1.25
TEACHING	0.29	0.80	0.75

pre-set ratios for sizes of the left and right split of the screen. This may contribute to student autonomy and perceptions of control and will be explored in future studies.

Although a separate instructions page was available for student use, comments regarding misunderstandings as to how to use components have led to the development of a Toolbar for each component that includes a pop-up help dialog box and options tools for additional control over UI components.

UEQ and data collection

There are questions to be asked surrounding the appropriateness of the UEQ as a tool for measuring *learning* experiences. Usability surveys are best used to evaluate interactive tools when they are being used to get a user to an end goal as efficiently as possible. We could argue that for many practical experiences the goal is to get to a better understanding of the theory as efficiently as possible, but as we begin to explore how remote laboratories can be used for skills development it may be that we are providing inauthentic experiences by making interactions too easy or efficient (e.g. automatic data collection vs. direct measurement).

As noted by Burridge et al.[24] computer mediated remote labs are well placed for passive observation of students whilst they perform practical work. The collection of appropriate interaction data could be used to help build learner models, help in assessing students' practical skills and to further develop the causal connections between student perceptions of their remote practical experience and their interactions with the user interface. This passive data collection will also help explore how the design of user interface elements contribute to user beliefs in the control they exert over their experiment and their feelings of immersion or autonomy. Typing in text or clicking buttons may provide easy but inauthentic means of controlling remote hardware and it would be interesting to explore how varying the means of controlling the hardware and data collection/analysis affects user perceptions.

Feeling in control/autonomy

Perceived lack of autonomy is a recognised issue within remote lab research[25] and was therefore a key concern in this study. Although we have highlighted that, from a usability perspective, dependability (or predictable control) is rated lower by respondents to our survey than other usability attributes we have not yet reached a solid foundation upon which beliefs in dependability are based and suggest that it is likely to have many contributing factors. When a user considers the feeling of predictable control they have over the remote lab system they could be considering: the booking system, the user interface, the hardware, or comparing the overall remote experience to the more common proximal one. Identifying what constitutes the dependability rating is beyond the capacity of this survey; however, hints to how remote lab design could impact user beliefs are presented below.

This survey finds that dependability does not correlate with agreement with the statement *'I felt like I was in control of the hardware'* (the statement itself did receive a significant positive response). This potentially suggests that hardware control was not the key factor in students believing they were *in control* of their remote lab *experiment*. Similarly, dependability does not correlate with increasing mentions of technical issues (which includes hardware issues) within the qualitative feedback.

However, dependability was correlated with agreement with the statement *'The user interface behaved as I expected it to'*, suggesting the possibility of a stronger connection between user perceptions of predictable control over their experiment and the interactions they had with the user interface.

Table 1 reveals that most components were positively rated when students considered whether that component made them feel in control of their experiment. However, some core components such as the 'webcam' and 'live graph' tool were not similarly rated - with the webcam in fact having the opposite effect. Issues surrounding the 'backgrounding' of the webcam have been identified in the literature.[26]

It is interesting to compare components that effectively perform the same function but have different modes of interaction via the UI. The 'Table' is a component that automatically stores each data set received from the hardware and displays it on the UI. The 'Snapshot' instead displays the same data but only for the values at the instant a user clicks a button. The user has no additional control over what data is collected, how data is measured or how it is displayed, simply control over when and how many pieces of data to collect. This interaction provides a significant difference in ratings of *'feeling in control of my experiment'* between the two components ($V = 4.5$, $p < 0.05$, effect size $= 0.447$).

A similar comparison can be done between the 'Live graph' and 'Graph' components. Both are automatically displayed - the 'Live graph' continuously; the 'Graph' only when data is actively recorded and yet students reported a significant difference in ratings of *'feeling in control of my experiment'* ($V = 94$, $p < 0.05$, effect size $= 0.616$). The 'Graph' component allows users to hover over data points to show values as well as click data points to move to the corresponding data in the 'Table' component. These relatively simple differences may be contributing to significant differences in student perceptions of the control they have over their experiment.

Evidence suggests that remote laboratory design can directly affect the control that students feel when interfacing with remote hardware and that this control appears to be framed around the interaction with the user interface rather than belief in control over the hardware. Therefore, an exploration of *control of task* in contrast to *control of hardware* could be an interesting future study.

Skills and perceived theory-measurement disparity

Although a number of studies have found that conceptual learning gains using remote labs are as good, if not better, than traditional, proximal labs,[25,27–29] scientific inquiry skills are the least assessed aspect of lab work.[30] Lowe et. al.[25] identify practical skills as a possible focus of future labs.

Aside from the practical issues identified in Section 2, it could also be argued that remote laboratories have not yet made it into the mainstream of practical work, in part, due to belief that they limit the development or application of specific engineering skills during practical experiences. Student perceptions (Table 3) support the view that remote and practical laboratories have different applications. However, remote laboratories have traditionally been designed to meet the needs of specific practical tasks, often to confirm theory that students are learning in lectures. Thus, remote laboratory interfaces tend to present only the necessary tools for achieving the predefined aims of the laboratory worksheet. This was the case in this study, with the majority of tools being rated as useful for the tasks at hand (Table 1). However, if we are to further explore how student autonomy and skill development can be enhanced, a more open and exploratory user interface may be necessary. This may include the use of a bank of UI tools that must be selected from by students. It may also require a mixed-format lab that provides the opportunity for experimental design and hardware choice that a simulated environment can provide before moving students to a remote environment providing students with the rich data that real hardware can offer.

Additionally, by providing students with additional means to investigate the remote hardware – pre-recorded video, datasets, 3D models, simulated environments for "constructing" hardware – we may be able to better understand the source of perceived theory-measurement disparity in remote laboratories and encourage a design of remote systems that better support authentic practical experiences.

Conclusion

If we hope to encourage the adoption of remote laboratories into mainstream educational practice, it is important that stakeholders in practical lab work perceive the inherent value in remote systems and consider the remote laboratory a valid learning tool that is capable of achieving the skills-based and conceptual learning intentions set in the engineering lab. To this end, a study of the perceptions of students (and staff) is necessary for understanding how remote laboratories can be designed to meet these needs. Remote laboratories are not simply a reduced version of a proximal laboratory, but a different approach with alternative strengths and weaknesses, offering new opportunities not possible with traditional approaches (many of which are not yet fully identified or explored).

Although still in early development, the practable.io infrastructure, UI and remote laboratory hardware have provided 250 students on the 'Controls and Instrumentation Engineering 3' course with an effective practical experience, with results suggesting that students were satisfied with the remote practical work. This paper highlights and supports the call[31] for further educational research into the impact of remote laboratory design on student experience. Within the context of this remote laboratory and the respondents to the survey, we have identified that perceptions of 'predictable control' over experiments may be driven to a greater extent by a student's interaction with the user interface (*control of task*) rather than a belief in their *control over the hardware*. We also show that different modes of UI interaction with the same hardware and data can impact student perceptions of control.

Acknowledgements

We are grateful to the course organiser Dr Aristides Kiprakis for developing the new teaching materials and working with us to validate the experimental hardware and user interface; Andrew Brown designed the mechanical hardware and built the experiments together with Calum Melrose. Additional essential support was provided from Technical, Buildings, IT, and Professional Services staff. The work was funded by the School of Engineering.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Research data

Data collected as part of this research has been made publicly available on Edinburgh DataShare[16] at https://datashare.ed.ac.uk/handle/10283/4007.

ORCID iDs

DP Reid https://orcid.org/0000-0001-6234-1298
DB Lowe https://orcid.org/0000-0002-6777-8955

References

1. Cardoso A, Sousa V and Gil P. Demonstration of a remote control laboratory to support teaching in control engineering subjects. *IFAC-PapersOnLine* 2016; 49: 226–229.
2. Haque E, Ahmed F, Das S, et al. Implementation of remote laboratory for engineering education in the field of power electronics and telecommunications. *Proc 2015 3rd Int Conf Adv Electr Eng ICAEE* 2015; 2016: 213–216.
3. Limpraptono FY and Nurcahyo E. The development of electronics telecommunication remote laboratory architecture based on mobile devices. *Int J online Biomed Eng* 2021; 17: 26–36.
4. Sáenz J, Chacón J, De La Torre L, et al. Open and low-cost virtual and remote labs on control engineering. *IEEE Access* 2015; 3: 805–814.
5. Drysdale TD, Kelley S, Scott AM, et al. Opinion piece: non-traditional practical work for traditional campuses. *High Educ Pedagog* 2020; 5: 210–222.
6. Knox J. The 'Tweeting Book' and the question of 'non-human data'. *TechTrends* 2015; 59: 72–75.

7. Bayne S. Teacherbot: interventions in automated teaching. *Teach High Educ* 2015; 20: 455–467.
8. Gal Y, Uzan O, Belford R, et al. Making sense of students actions in an open-ended virtual laboratory environment. *J Chem Educ* 2015; 92: 610–616.
9. Ross J. Speculative method in digital education research. *Learn Media Technol* 2017; 42: 214–229.
10. Hamilton J. On designing and deploying internet-scale services. In: *Proceedings of the 21st Large Installation System Administration Conference, LISA 2007*. USENIX Association, 2007, pp. 231–242.
11. Araújo AS and Cardoso AM. Pedagogical effectiveness of a remote lab for experimentation in Industrial Electronics. In: *Proceedings - ICELIE 2009, 3rd IEEE International Conference on e-Learning in Industrial Electronics*. 2009. Epub ahead of print 2009. DOI: 10.1109/ICELIE.2009.5413201.
12. Cooper M and Ferreira JMM. Remote laboratories extending access to science and engineering curricular. *IEEE Trans Learn Technol* 2009; 2: 342–353.
13. Ferreira J and Mueller D. The MARVEL EU project: A social constructivist approach to remote experimentation. *1st Remote Eng Virtual …* 2004; 1–11.
14. Jona K, Roque R, Skolnik J, et al. Are remote labs worth the cost? Insights from a study of student perceptions of remote labs. *Int J Online Eng* 2011; 7: 48–53.
15. Lowe D, Murray S, Weber L, et al. Towards a National Approach to Laboratory Sharing. *AAEE'09* 2009. *Australas Assoc Eng Educ Conf* 2009; 458–463.
16. The openScience Laboratory. https://learn5.open.ac.uk/course/view.php?id=2 (accessed 18 October 2021).
17. VISIR Final Report. https://cupdf.com/document/erasmus-ka2-capacity-building-in-higher-education-annex-v-erasmus-ka2-capacity.html (accessed 23 September 2021).
18. practable.io. https://practable.io (accessed 18 October 2021).
19. Wurster M, Breitenbücher U, Falkenthal M, et al. Developing, deploying, and operating twelve-factor applications with TOSCA. In: *ACM International Conference Proceeding Series*. 2017. Epub ahead of print 2017. DOI: 10.1145/3151759.3151830.
20. Laugwitz B, Held T and Schrepp M. Construction and Evaluation of a User Experience Questionnaire. In: *USAB 2008*. 2008, pp. 63–76.
21. Reid D. Student feedback on 'Controls 3' remote lab, https://datashare.ed.ac.uk/handle/10283/4007 (2021, accessed 18 October 2021).
22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Routledge. Epub ahead of print 13 May 1988. DOI: 10.4324/9780203771587.
23. Feisel LD and Rosa AJ. The role of the laboratory in undergraduate engineering education. *J Eng Educ* 2005; 94: 121–130.
24. Burridge J, Lowe D, Willey K, et al. Defeating Hawthorne in tech-enabled education : Passive observation of student behaviour with a remote laboratory The Laboratory System. *AAEE2019 30th Annu Conf Australas Assoc Eng Educ*.
25. Lowe D, Newcombe P and Stumpers B. Evaluation of the Use of Remote Laboratories for Secondary School Science Education.
26. Lindsay ED, Murray S, Liu DK, et al. Establishment reality vs. maintenance reality: how real is real enough? *Eur J Eng Educ* 2009; 34: 229–234.
27. Darrah M, Humbert R, Finstein J, et al. Are virtual labs as effective as hands-on labs for undergraduate physics? A comparative study at two major universities. *J Sci Educ Technol* 2014; 23: 803–814.

28. Nickerson J V, Corter JE, Esche SK, et al. A model for evaluating the effectiveness of remote engineering laboratories and simulations in education. *Comput Educ* 2007; 49: 708–725.
29. Tsihouridis C, Vavougios D, Batsila M, et al. The timeless controversy between virtual and real laboratories in science education—"and the winner is…". *Adv Intell Syst Comput* 2019; 917: 620–631.
30. Brinson JR. Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: a review of the empirical research. *Comput Educ* 2015; 87: 218–237.
31. Post LS, Guo P, Saab N, et al. Effects of remote labs on cognitive, behavioral, and affective learning outcomes in higher education. *Comput Educ* 2019; 140. DOI: 10.1016/j.compedu.2019.103596.

I. REDACTED DATA – COLLABORATION STUDY

Context note: This data collection was emergent through a long-term and embedded collaboration taking place over 2-3 years. Formalised data collection separated from this collaboration occurred at two points during this period – one with students detailed in Appendix H, one with academics provided as the interview transcript in this appendix.

The nature of the collaboration shifted as a result of the COVID-19 pandemic. What was intended to involve significant in-person ended up operating almost entirely online.

COMMUNICATIONS SUMMARY

Participants:

- Joshua Burridge (the author)
- Professor David Lowe
- 2 academic practitioners – one senior professor, one PhD candidate

- 2019 – in-person project inception meeting, Joshua Burridge & 2 academic practitioners
- 2020-2022 – online collaborations
  - 1 per month scheduled zoom video conference discussions – laboratory design updates; research design, progress, writing; theoretical discussion
  - As required written email communication – 144 records deidentifiable excerpts included

REDACTED
PARTICIPANT CODE|DATE

1|5-1-20
My initial reaction this is v positive - it maps onto concerns I want to highlight regarding the way in which humans cement control of the narrative, and how we have elevated experiment but lost the nuance in the process - like photography, we often omit to emphasise the framing/composition/cropping (photography) or experiment design/apparatus (lab) are human decisions that control the narrative and have been "washed" by the illusion of objectivity that we place on top of "real" data. So .... simulations and real experiments both embed human bias, but student reactions to those are different, because of the cultural background that "measurement is always right". Whereas "mesaurement is always whatever that measurement was going to be, and may not be reality at all.",

Mixing in virtual, remote then it may also be that real is not on a linear axis as implied by your title, but that it is on a 2D axis, with the perception of veracity on one axis and the actual method on the other ..... we're trying to get off the diagonal line into the quadrant where elements are represented in software but are perceived as having veracity (being Real enough) - i.e. maybe students will operate under the idea there is a linear "real-enough" curve (simple mental models often used when complicated ones are not known to be required), but really if it is 2D then there are a series of pareto (equivalency) curves on that 2D space that represent perception of realness - and that gives us some wriggle room to dodge to the side by changing how we present an element of a remote/virtual/online lab, and hop over the line into a more "real" space - making the experiment that is remote/virtual be perceived to be more real. There is a wealth of social science to draw on here, and although that they may seem like a risky place to go, I think it would help open up this discussion in a way that testing A vs B never will.... Great direction to go in.

1|24/6/20
This active observer thing is interesting, and worth developing it to have virtue that goes beyond the facile "group work due to a lack of kit" - with the reduced cognitive loading and reduced emotional loading it's arguable you can get further through the subject before running out of psychological energy for the day.

Veering into anecdote of a similar-ish learning situation, so as to access a single datum this point on the emotional rationale - in police driving courses you have one instructor, one student driver and two student observers in back, and you take turns being the student in the hot seat, exercise by exercise. I found it really useful when being a student in this situation to see what the other students would mess up, and then avoid that myself, and mix it in with sometimes being the one that had the excitement/pressure of having the chance to get it right first time. Messing up by us students was often simply a case of misunderstanding expectations or what the words meant when translated into action, so sometimes it was a real ego saver not to be continually the one performing (even when one was doing well it was not obvious which exercises might or might not come easy to you).

We were all advanced drivers who'd already been assessed by police one to one, and this was a training course on fairly extreme manoevures you can't do on UK roads unless you want lots of questions, so it was a reminder to me that at any stage of the learning journey in a subject (not just the beginning) that genuinely new topics can be just as disconcerting / have hard to understand expectations / etc.

Perhaps tailored one-to-one instruction would still be superior against a more limited set of criteria, such as getting an individual to a certain standard because remedial help can be tailored for a more sustained period, but other outcomes such as understanding how other people interpret/misinterpret those instructions are valuable for laying to groundwork for the student to instruct later (and that echoes requirements/expectations we have in STEM) and with one instructor teaching three people in one day, overall, the throughout is good, so long as there is reasonable consistency in the group experience level of that particular activity, so the vicarious instruction flowing into the back seat is pitched at the right level.

Whether the police have published on the pedagogy of this method, and what where their conclusions, is another thing .... But the medical educators will likely have explored it as part of explorations into see one, do one, teach one.

2|2-19-20
PDF stitched out of sequence – "Remote Lab PID Controller: Experimental Design", matched date

2|14-9-20
PDF stitched out of sequence – "Remote Lab PID Controller: Experimental Design", matched date

2|5-10-20
PDF stitched out of sequence – "Remote Lab PID Controller: Experimental Design", matched date

1|23-4-21
Yes very good point re playing versus theory - we need to find a way to surface this, but there are indications that there are rich pickings here (e.g. we had a few people querying whether we really meant gains rather than time constants, to do with the way the ZN tables predict values that are a bit high for Ki and Kd compared to what actually works .... if we could dig into that a bit more it would be potentially interesting. It's possibly even on the boundary of asking does the kit work well enough, to withstand the

kind of scrutiny it is getting, and if it doesn't, what does that say about the way in which we present nice clean theories that slightly disagree with the real world).

2|2-9-21
PDF stitched out of sequence – “Remote Lab upgrades and triggers”

1,2|24-9-21
Interview transcript

<preamble redacted>

<obtaining named consent redacted>

Okay, so with that, out of the way, quite a bit of the initial script doesn't make any sense as we've spoken together before. You have a good understanding of what I'm about and what the research is about. So I'll dive right into the part where I ask you about your laboratory. So what I would like to talk about is the equipment and activities that the lab involves. Now I'm assuming that you're going to be talking about the control systems lab but if there is another one that you prefer then we can talk about that as well. But it's about how the laboratory fits into the unit of study. Now I know that you're not the sole educator using that lab but your understanding of how it does. What you believe the purpose of the laboratory to be and how well it ran or sorry, not expected because it has run, but how well it went. So, yeah, they suppose to start with the equipment and activities.

So, yeah, like an artificial contrast originally it was using Kwanzaar cube servo motors and the studies effect that the practicals effectively about students measuring the characteristics of DC motors, the DC motor plant with a weighted disc on it, figuring out the gain of that motor, the time constant of the motor, and then using those. Measure values to get a theoretical idea of what the response to a step input would be for position control and then tuning a PID controller in order to control that motor position within set constraints, not going over 20% you know, beyond the step point and doing so in a certain time. So the equipment we used is a remote lab. So they, we had relatively cheap DC motors bought in we had in house constructed.

Well motors with stands and the discs on top all built in the <redacted>. And then they were connected up with Arduino so an Arduino ran the firmware, the PID controller itself. And then that streamed information from the DC motor, the position time, etc to Raspberry Pi, which then streamed it to servers. And then students interact via user interface in order to control that. So, and the UI provided kind of basic means of setting the modes of the motor itself, and then quite a few different data analysis tools, so the ability to display all the data streaming from the DC motor, but also graphing tools and tools for selectively picking out certain snapshot data points, things like that. The, what was the next part.

So the equipment, how the laboratory fits into the unit of study.

Ultimately, the, in my mind, the practical was really about confirming theory. That was the real focus in it. And to be honest, I don't think there was a huge priority given to skills development or anything like that.

Even before it was a remote lab. There was no real sense of student choice in terms of how they construct the practical. There was no real sense of your investigating a variable and how it would affect the outcome it was effectively that traditional follow the practical book in order to look at confirming the theory you've been learning on PID controllers. So it was pretty much kind of, you know, set them all run the controller, run the motor, collect the data and then use that data to identify the characteristics of the motor. And then use your theoretical models and to get prediction and then see what the motor actually did and how far off it was from that prediction. So the real focus of the lab really was was that was, can you confirm the theory you've been learning. I would say. Okay. Yeah, and a third part.

So the third part you've sort of covered already so that is what the purpose of the laboratory was. So suppose all that's left is how well the laboratory ran.

Yeah, so there was 250 engineering students. So it was a relatively big cohort having to get them through in a couple of months. We had 15 remote labs available, which seemed like a good number if students were to can uniformly book the system out. But so we have few issues around the booking system just because of that idea that the peaks and troughs at different times. So running up to the deadline, there was large numbers trying to access it. So, you know, in terms of actually, there was a bit of a bottleneck and getting them on there, which was a slight issue. But I think when students got on and were using it, it run, it worked well. I think perhaps the disparity between the theory they were learning and the practical nature of the hardware box a little bit bigger than expectations were originally.

But that was the nature of kind of the relatively cheap DC motor and things we were using in there. But students were able to accomplish all the tasks that were set. So I think from that perspective, it ran really well. And marks are high. So I think students have done pretty well in terms of being able to answer the course workbook and be able to collect the data. So, yeah, I can't say definitively, but I think it ran well if we're judging it on did they get reports written? Well, because that was all marked.

Marked. And certainly, you know, upgrades and improvements to be made for the next run. I think things like booking systems and maybe a number of pieces of kit available, but I think that will be lessened when there's an advanced booking system in place. And perhaps a change of motor as well, just to ensure if the intentions of the lab are to confirm theory and that is the main intention. And I think we have to make sure the lab confirms theory a bit closer than we might have been expecting in the first place. But if the lab instead was to change it in my mind, I would have a slight shift in it towards a huge few you've been learning, but can you investigate his real system and understand why there's differences? Then I think it'd be a bit more valid in that case. But you know, students got through this and, you know, from student feedback, they seem to enjoy it. If that's important.

That's a discussion to have, I guess. And they felt like they were in control of the hardware. So I think from a kind of usability perspective, it ran quite well as well.

Okay. That's great. Now, I know we've discussed this already, but on the record, as it were, you mentioned a couple of other

things. So I think that's the nature of being a developer of a system, right, that you learn very quickly or, you know, you force yourself quickly to know exactly how to use that system. You can't help it. You designed it that way. So you use it that way. And then we've talked about the way that you use it. So I think when you get feedback from people who have never seen that before and their instant response to it can be a bit different.

So I think in the feedback, there's a, and most of the feedbacks say they work fine and great and they could do it well, but there was a little bit in terms of just the relative positioning of elements on the screen. And then quite a few students found relatively tricky. There were moments in the practical where you didn't necessarily have to, but I understand why students wanted to kind of quite quickly start and then stop the motor effectively to get as little time, but only the necessary time on the graph is possible and things like that. So they were talking about having difficulties, say starting with a button position down in the bottom left of the screen and then getting it stopped with a bottom position up in the right. So there was a few issues surrounding kind of the relative placements of things. So because of that, and I've tried to redesign it in order to make it as kind of customizable as possible. So now each of the different UI elements, the graph or the controls hardware can be shifted around the screen. So students can move an element to, you know, next to another one that they know they're going to be using in order to make that much quicker to jump between the two of them. I've also split the left and right hand side of the screen so that you can align something that you've got positioned on the left hand with something at the same height or something on the right hand. And it will just make sure that buttons you're using or you know, graphs you're looking at can always be positioned on the screen in a way that's going to be comfortable for you to use and to make things visible as well. There was also, again, in using it myself, I was just doing a test run so I was collecting, you know, maybe a couple of hundred data points or something like that as I was running it. And then viewing the table was fine, but students were running it and collecting the full 5000 data points and then trying to view that table on the screen was causing it to scroll massively down the screen and they'd have to scroll all the way down to get to the bottom. So I've done things like limit the size of those components there as well.

When you say, sorry, is that just from that feedback questionnaire that was at the end or did you have other avenues of feedback from them?

That is 99% from that student feedback questionnaire. So the vast majority became from that there were a couple of changes that were actually made during the course that kind of effectively became necessary. Again, in testing it, we were not hammering the system right you were collecting the data need to be collected viewing it looking at the idea this all worked. And just because of time, you know, we were rushed to get this whole thing developed. The testing process probably wasn't as good as it could have been. But when students were then using it, they were collecting tens of thousands of data points all at one go and that was trying to display on graphs here and there all over the UI and it was slowing things down. And so they were getting a little bit freezing and things are juttering on the UI. So during the course I had to go in and limit the number of data points that can actually be collected and displayed on the screen. So that was that was feedback got during the actual run quite early on maybe just a week or two in.

So I don't think many students have actually used it at that point. But that was the one thing that we did have to change. And I want that feedback you got at that point. Otherwise it's from the online survey that we gave out at the end of the course. So after they'd all run through.

Yeah. So what I'm going to do now is share just a quick slide. Okay, let me know if you can see that. Yep. It's a very basic black and white slide. So this is intended as a bit of a thought process intervention in the way that people think about labs, the way that they model them in their heads and the way that they design them. A lot of the description of the lab that you've gone through already fits onto this model itself.

So we've got things like the purpose that we've discussed already. There is you as a designer. There is also the <redacted> and obviously <redacted> as well.

But also stakeholders. And what I'd like to do for this second half of the interview is to first give you my interpretation of what bits of the discussion map onto what bits and what bits I might still have questions about. And then to see what this idea of modeling the lab in this way and the different questions that it asks how that might measure map differently to what you see as lab design and what questions it raises or what possibilities it might be.

raise. So this model starts with the purpose. It's based on a purpose first approach to lab design.

And the PIPOV that labels it is purpose inputs, processes, outputs and validation, which is effectively an idea for a pipeline of how a student is going to get into and through a laboratory teaching and learning experience. So the, as I said, the designers and stakeholders that I know about from this discussion and previous discussions are you, and <redacted>. Would you identify any other stakeholders in the lab?

Yeah, I mean, stakeholders or designers, sorry. Stakeholders. I mean, well, students are stakeholders in the lab as well, I'd say. And even I think in this, in this case, and maybe not necessarily the case in all practical labs, but or even all remote labs, the actual mechanical designers off the hardware as well, I think are part of this and have been a big part of this since the beginning.

So there are a couple of mechanical engineering technicians at <redacted>, they've been quite heavily involved and will continue to be involved in the production of the hardware side of our remote lab system. And so they have been certainly, you know, actually involved more than I might have expected in a really positive way, as in they are interested in the outcomes of these remote labs. They are interested in how we can iterate the designs of them to improve them.

And they've been putting a huge amount of time into developing better and better systems as we go along with many, many different iterations as we say of the designs. So I think there's stakeholders in here as well. But then on a larger scale, the, the, you know, management and lecturers and teachers as part of the engineering faculty as well are definitely stakeholders.

They may not be that heavily involved at the moment. But if the purpose and outcomes of these remote labs continues to be seen more clearly and to, and we really get good outcomes from these labs, then I think they will start to be more involved as time goes on. And we started with this one lab just now. This was our first run that we have subsequently attracted the attention of maybe two or three other courses, one which I've built and designed and is ready to run this semester, and others that are going to start in semester two.

So it's because of, you know, the people that are involved in this, this first run and the word of mouth and thing that's gone on because of this first run that other people are now getting involved to. So I think it runs through the full range down from from students up to the kind of upper echelons of the hierarchy in the engineering department and potentially even up into the university wide as if engineering can show that this system works and gets good outcomes than other other departments might take on as well.

One of the, I mean, obviously the central aspect of this model being the purpose. And we've discussed that already in terms of the validation of theory side of things. Quite often, what I've found when people are working with designing laboratories is that they have that sort of purpose which some people can and some people can't point out and articulate. But there are other non learning outcome purposes that people will know about and will incorporate into the design of the lab, but won't necessarily consider when they're asked about the purpose of the lab. So there are two aspects of the laboratory that we've discussed that actors constraints that I would probably include in a statement of purpose to do with time and cost in particular.

So obviously you were rushed to the deadline, as you mentioned of this getting it up and running for a semester. And you mentioned earlier with them, the discussion of <redacted> how the low cost of these labs is distinct from from other labs. So one thing that this model tries to capture is the idea that all of those different constraints exist within the purpose of the lab, which is important when we move into the inputs of the lab and what you choose for those, because there's going to be certain inputs that you might like to have. But due to some of those constraints that are not learning outcomes, obviously you couldn't have either in this run nor in any run. So the inputs being the particular physical written things that are provided. So the laboratory equipment, any lab guides that were provided.

So they're the obvious ones. I'd also like you to think about the way that the students access it. So whether that is an individual student doing individual work for a report, or if it's grouped, or whatever the circumstances there are, what timing that the lab has. So obviously this comes right into play with a being a remote lab, we have to ask more questions than normal. There's no scheduled lab time that they have to access it at. There might be differences when they have a lab demonstrator or someone to help them available in a class, if they're able to access the lab at that time. So is there anything more that you can tell me about the inputs to the lab? So what the students were given, what the context were?

Yeah, so again, this all has to be prefaced with the fact that it may be that I'm not correct in what I'm saying, but this is my best guess, let's say. The hope of <redacted>, I think was to basically maintain what he had done as in person lab. So the they were provided with the same lab workbook, modified just for how to access the labs, etc.

and updated a little bit. But basically the tasks and questions things were all the same. So that was their that was effectively the main input, I think, into the process was, here's your lab workbook. They again, originally had had two 90 minute sessions. So the intention of this remote one was still for them to be able to achieve the outcomes of the lab in two 90 minute sessions. I think it's clear, though, that a lot of students spent a lot longer on it than that, though.

And that contributed to kind of the bottlenecking of getting it booked out. The <redacted>, obviously been doing this in class in groups of four, I think, originally, but that shifted to them doing it completely independently, well, but with no constraints, really assigned. So there was nothing stopping them from meeting up together and running the lab. There was nothing stopping them from having each other on zoom and having a chat as they were going along.

And so there was no limitations there on who they could work with and how they could work in order to get the data. But I think most of them will have gone ahead working on their own in front of the computer. The, the word, therefore, I don't think any kind of demonstrators available, you know, at set times for them to chat to now I could be wrong about that. But I know that they were in communication with the course organizer, fairly frequently when they were stuck. So they're able to email him. And I think he had some drop in sessions where they could speak to him. And they were also given a contact for, for <redacted>, and so that if they were having issues with it from more of technical kind of UI perspective, they could get in touch with us and we could provide some feedback. So that happens a little bit, especially to begin with.

And then when students were struggling to get actually booked on the system, and they were contacting them as well. The I also had produced a little website that gave more information on the user interface and how to work and how to run it and what things should do and, and, and that kind of thing. So there was that as an input as well on top of the lab book. There was this additional and resource that was intended, I guess, to try and replicate being a, you know, demonstrator a little bit, that hopefully it could answer some of the more basic technical questions as to how to get this data on a graph or how to use this component that a demonstrator might have helped with if it was in person. Anything else? Not I can think of right now.

So you've touched on one of the big splits. Any feedback from somewhere? So you touched on one of the big splits between desired and actual in what you were talking about there with the no, they weren't expected to do group work, but they could have.

So that's a big part of this model as well as this recognition that you can lead a horse to water, but he might go on to side to a grass if he wants to. Did you have any way of observing what it was that they were doing, whether it was individual or group or is that not something that was measured?

No, I'd, I'd them anyway. So there was I'm not trying to think. No, I originally, I actually originally had in the survey a question about whether it worked as a group or not. And somehow it disappeared from the survey.

Or did it have to check that? No, basically no, I can't really tell from survey responses. And that's really the only way I've got, I've got input into how they were working was in how they responded to the survey. I can't tell whether they worked individually or as groups or contacted each other or, you know, got help from each other. Really, no. Okay.

Did <redacted> have an input on individual or group design, you know, so was going to individual something that they're after or something that they just didn't mention?

There was no discussion between me and <redacted> as to how the, how they intended, you know, that element of work to go, whether they would be allowing them to work as groups or whether they're encouraging to work on their own or anything like that. Or even if they were providing, you know, I don't know if <redacted>, you know, identify the possibility of working in different ways to them. So, unfortunately, I can't really give an input there. Okay.

Is that the sort of thing that maybe it was discussed with <redacted>? It just wasn't a factor.

I think probably again, my best guess would be the intention from the off was that they would just do this independently.

Okay. So, I mean, the main reason that I'm querying that is just that idea of trying to nail down unarticulated purposes to labs quite often that idea of team psychomotor is something that comes up where people have it but might not write it down as the purpose and so on. So, the next step being the processes. Now, part of that we've gone through with what you've observed with the queuing and the conversations you would have had about troubleshooting and so on. Recognizing that there is that planned process that they're going to go through.

So, I won't cover too much of that again, except to say that the processes that they go through obviously are going to be driven by the inputs that they are given. So, not one to one. It's not a case of the workbook says do this, therefore they're going to.

But it's going to be made more likely. The part of processes that will probably be more of a conversation with the course organizer, but you may have some insight anyway, is the cognitive processes that the students are going through, not the physical things that they're doing, the buttons they're clicking and so on, but the thought processes that they're having. So, you may have observed some of that from the feedback to the survey or from the questions that they're asking in the troubleshooting. But did you get any insight into what the students were thinking about while doing the lab, what their focus was on, how they were interpreting what they were seeing?

Hmm. Yeah, now that you've highlighted, it was effectively something I was hoping to do more of as part of this study. I was wanting to have some post lab feedback sessions with individual students and I was wanting to then sit with them and go through some of the tasks and ask them to actually discuss and explain what they were thinking about. So, I was hoping to get some feedback from the students and about as they were trying to achieve these tasks.

Now, COVID and exams came up and the lab was delayed by good few weeks and so that pushed into their exam time. Those interviews just never happened. So, that aspect of trying to get feedback more on exactly what they were thinking was effectively lost because although the survey asked them to reflect on the experience, it doesn't really ask them to reflect on how they got for me to be and what they were thinking in order to do that. So, unfortunately, there's probably not much I can actually give you in terms of what they were thinking as they were doing it. There's a little bit in terms of, I think, a slight barrier and we see this in some of the responses to the we asked them the question, why if you saw a difference between theory and reality in your practical, why did you believe that that difference existed? And you can see a range of thoughts there, but quite a few of them referred to them not feeling like the lab could measure as accurately wanted it to or there were some issues with the lab. So, by making it a remote lab, I think there was kind of an abstraction or a layer of difficulty between them and getting to the answer. And I think of quite a few times, and we found this in some emails we got from them early in the early days, there was some element of they thought they were fighting with the system when really a lot of the times they were getting the appropriate data, but they had difficulty processing that data in a certain way, because they felt that they were it was a UI or it was a remote lab system problem, rather than a problem with their understanding of the theory or of what the response should be like. So, I think that didn't necessarily and that's part of what I'm wanting to look at and study more remote labs as well, by giving them a remote lab by putting an interface between them and the hardware, they start to think and get worried more about what they're seeing and how they're interacting with the interface, rather than what results are coming from the hardware itself or what they should really be thinking about. Yeah, otherwise it's difficult, I think it's difficult for me to say necessarily what they were thinking about or were worried about in there.

There was a few mentions in the survey as well about the lack of a demonstrator and the lack of team mates to speak to and things of that as well. So, I think that was something they were concerned about as they were working through the lab. But it obviously is not necessarily something they're considering as they're trying to solve. Well, it's something they're considering to try and solve a problem, because I think a lot of times they were thinking, well, I could have called someone over to speak to me to really help me through this, whereas instead I'm sad on my own and unable to solve this problem for now. So, it was a concern of theirs as well, not being able to communicate with people as easily throughout the process. But otherwise, in terms of kind of their thought process and getting from what they're seeing on the screen to the theory they're supposed to be comparing it to, I don't have too much of an insight into that side of things.

Okay, that's been really helpful. The last step before we look back, I already know quite a bit about both from this discussion and previous ones. So, the outcomes of their marks that you indicated were high for the report. There's the post lab survey that they've undertaken. So, you have an idea of the actual outcomes that you've observed in both of those.

So, I'm assuming for the mark side of things, <redacted> obviously intended for that to be the case and therefore that's a match between desired and actual outcomes. Was there anything in the observed outcomes in the user survey that was unexpected or a surprise based on what you had thought the processes were going to lead them to? Let me think.

Maybe if I actually have more survey information up. Well, one thing definitely that I wasn't really aware of but then have no thought is quite an interesting aspect of the lab was this sense of where the nature of a remote lab means they're not touching and manipulating and playing with hardware and I think that's the thing that traditionally has been seen as a real weakness of remote labs. Students can't get their hands on it. They can't change it.

They can't do this. But I was interested to see that students actually felt what seemingly seemed and again, we've got small numbers really in the samples. We can't necessarily generalize to the whole group. But at least the people answering the survey seem to feel that they were in control of the hardware.

They didn't really have an issue with that. So I thought that would be something that would come up. Instead, what came up was this idea of, well, we don't necessarily feel in control of the interaction we're having with the user interface.

And that was interesting as well. Because I think again, traditionally, the interface has just been viewed as a means of really being in control of the hardware. So as long as you can get that control of the hardware aspect figured out and they're happy with that, then you're fine. But actually, the interface is removing from the students a lot of the aspect of measurement, a lot of the aspect of data analysis is doing things automatically a lot for them. So that was an interesting kind of unexpected outcome, I would say that understanding that we have to really make sure the UI is designed in a way to encourage the skills that we're trying to develop. Now, in this case, as I said, the main kind of conceptual purpose of or learning attention of the lab really was to confirm theory. So by having automatic data measurement, etc, etc, that's not really limiting the intended outcome of learning, you know, learning the theory and comparing it to that. But perhaps it's something maybe I should have mentioned earlier in terms of kind of my purpose for <redacted>, I want to look a lot more carefully at how can we redesign labs to encourage certain skills development.

So that was missing from the intentions for lab. And that's fair enough. That's not really a problem. But that is within my purposes. And I think I could also see within the data that came out that there was a disparity between the skills you might want a student to learn in a lab, and the skills that students at least perceived or thought that they could achieve with that with a remote lab.

So I think that was an interesting and not not necessarily intended, but an outcome that came out for me. This idea that I've mentioned already, the design of the UI is going to be an important feature of remote labs and particularly an important feature if we're thinking about skills and other aspects of lab and work that's going on. Otherwise, otherwise, nothing, that's, nothing, that's it.

So you've sort of given me a nice segue for the, what is the final part, but the largest part of the model itself, which is there, what sort of benefits is it intended to provide to people that are using it for labs. And by identifying that your purposes are different to <redacted> purposes is in that sort of mindset, where recognizing that each stakeholder is going to arrive to the lab with a different set of purposes that they intend for the lab, whether that is the students who expect a certain thing and will react negatively if not that's not there, even though that might not have been the educational purpose of the lab, that you as a designer have a particular purpose that <redacted> has a particular purpose.

The university as a whole probably want something out of it industry might want something different out of it and so on and so on. Where that so just by recognizing that and seeing that there's going to be different purposes that flows on through the inputs that should be prioritized what processes you're going to be seeking and so on. The part that I believe has the most value though is the recognition that outcomes are not the only thing that can be or should be measured. If we're trying to identify the purpose, sorry, if we're trying to identify whether the purpose has been met or not. So you mentioned that you wanted to but didn't have the ability to have that post lab chat with students about their cognitive processes. Do you know of or can you think of a way to try and measure that not as an outcome, but as a process during the lab?

Well, one thing that I'm interested in you've written about this in the past as well. I think I've referenced you in <redacted>. In fact, is we're using computers in order to perform these these labs. And we're one thing that is a vast improvement in there and that we can't really do necessarily in an inface lab is actually collect the data as they're running through these things. What are the interacting with on the screen?

How long are they taking to get from point A to point B? And have they are they actually accessing tasks in the order that is laid out in the practical workbook and those kind of things. So one thing I'm interested in upgrading these labs with is some element of data tracking in the background as students are actually running it. And I think I think that combined with a discussion afterwards would be useful not only in trying to confirm, you know, what students actually did or why they actually feel that way, but also looking at the contrast that will probably exist between how they actually try to process things during their lab and how they felt about that afterwards as well. And the hope or the intention is that that kind of low level data tracking of their interactions might reveal things about their beliefs in the control they have over the experiment or in terms of other usability factors, but also in terms of will do the process. Are they processing the task in the way we thought he would? Are they approaching the tasks in the way we designed the lab to do? And that might be quite a rich data source for for them redesigning things to ensure that either they can run more efficiently or we can, you know, have tasks in a different order or whatever else. So I think that's going to be a useful way of doing that. You know, which is not really possible, not in a remote lab environment.

Okay, thanks. I'm conscious of the time. So if you've got to run, let me know, I just have the one final question for the end. And that is whether you've used frameworks, maybe not like this, but frameworks to model what you're doing in the teaching and learning activity before, and how you believe this compares or contrast to those.

Yeah, I mean, in terms of <redacted>, no, definitely not, not used anything like this. And I say because, well, you know, as the designer, I'm a huge part of that that processing, you know, one key element along the rest of them. So I feel absolutely that this should be more of a process that people are running through as they are actually in development of the lab. That I think it takes a bit more communication between the designers and other stakeholders off the lab for for these things to happen than happened in our case. And that's for many reasons, you know, COVID, blah, blah, blah, etc. etc. And so I feel like as a designer, along with the other designers, we've had our purposes and our desired inputs, but none of them have really been communicated to each other.

And none of them have necessarily been thought about as how they would work together, etc. etc. So I think it's really important that we do these things to a lesser extent as a <redacted>, which I was in a previous career, a simplified version of this would be would be run through on the larger scale in terms of planning whole schemes of work, you know, for a year group. But in that case, you know, well, in that case, there are there are kind of desired and unintended outcomes in there as well, obviously. But also then on a smaller scale in terms of individual lessons as well. Where, you know, we were constantly thinking about, well, how did that go? And then how could we then feed that back into how we redesign a lesson in the future? So, yeah, I don't even as a teacher, not kind of as clear cut.

But I think this this is a really good model. And I think it's important. And just discussing it with you now, you know, has made it obvious that there were massive gaps in terms of our thoughts in running it. And even if I was to be the sole designer and have come up with a lab, etc, etc, would I have thought about it in this detail? Probably not. And I think that, you know, is an issue, actually. And something we need to think about more.

Okay. Yeah, that's fantastic. Thank you very much. Cool. I might kill the recording.

Collab|8-9-22
What does 'collaboration' look like?

Axes of consideration:

- Where in the process
  - Design of laboratory
  - Design of experiment
  - Control of equipment
  - Collection of data
  - Analysis of data

- Write-up
- Follow-up

- Depth (? Nature?) of interaction
  - Necessary shared work (2 people lifting a heavy thing)
  - Independent shared work (2 people working in parallel on tasks that could be solo-sequential)
  - Enhanced shared work (2 people doing the ‘same’ task, but richer/faster/more reliably, e.g. pair programming)
  - Timeshare (2 people repeating the same task, swapping)
  - Driver & Observer (1 person conducting work, 1 person observing – can blend into timeshare with swapping, or enhanced if there are active observer tasks)
  - Divide & Conquer (2 people take independent tasks and work on them individually. Can be sequential where 1 task is an input to the next, or collating where the tasks are done in parallel and then combined)
- Proximity (student-student, not student-equipment)
  - Proximal
  - Remote
  - ...simulated? Could be fun and novel, but unlikely to be practical
- Synchronicity
  - Sync, both people actively ‘in’ the same space at the same time
  - Async, students collaborating over a time gap

What considerations for a remote system supporting collaboration:

- What drivers for collaboration have been removed in a remote setting?
  - Constant, computer-aided data collection (measurement becomes passive rather than active) – no need for enhanced shared work
  - No equipment set up requirements – no need for extra hands to carry the bits and pieces, hold the heavy thing while I screw in the fiddly bit, etc.
  - No/reduced time pressures – squeeze a 3-hour traditional lab into a 2-hour timetable slot, needs more people doing tasks in parallel. Remote can just take 3 hours.
- What dangers might exist?
  - Roadblocks seen as annoyances, ‘childish’ gamifications – need for perception of intrinsic benefit to collaboration
  - Could ‘spend’ benefits of remote labs seeking collaborative benefits of traditional labs
  - No strong evidence that there is a **need** for collaboration in labs (particularly the control of equipment/collection of data steps that nearly all pubs appear to focus on) – strong assumption, yes, but perhaps it can come in after the fact with a group written report?
- What opportunities exist?
  - Certain types of collaboration are enabled by remote/computer-mediated labs – particularly async
  - Better group interaction observation and management
  - Perhaps more ‘meat’ to the collaboration, if busywork is allowed to stay computer-managed and more activities can be brought in for different students to conduct
  - More authentic collaboration (presumably industry collaboration is not nearly so focused on shared work as eng-ed lab research is – if divide & conquer gets the work done, then great)

Controls Engineering Remote Lab: Experimental Design Overview

[redacted]

August 12, 2020

1 Overview of potential research questions

Although having in no way exhausted the literature, it seems relatively clear that most of the focus on remote lab research so far has been either a) to demonstrate a particular system and explain the technical aspects, b) to discover whether students find remote labs useful or c) to compare learning between a traditional lab and a remote one. The overriding view appears to be that remote labs can do just as good a job as traditional ones.

However, two key points stand out to me amongst the literature:

- Student *perceptions* of remote labs are quite different from traditional ones.
- Much of that could be due to the lack of direct involvement in the *construction* of a remote lab.

Students interact with a remote lab via the interface, not the equipment itself, therefore generating a sense of control over the experiment could be supported by giving students more control over the interface.

What also stands out in the literature is that few studies have considered how individual elements of the interface could affect learning or perceptions. Most studies consider the remote lab as a monolithic, unchangeable entity that is to be compared to a similarly invariable traditional lab. A major benefit of remote labs could be the ease with which interfaces are tailored to student needs. Users should be at the heart of the design process (see section 2).

1.1 Phase 1: PID controller remote lab

The following is a relatively vague list of general exploratory research questions. An attempt to make them more specific can be found in section 4.

These questions are potential considerations for my work this academic year, including in the PID Controller remote lab.

- Can the use of contextual media beyond that possible in a traditional lab help improve student perceptions of a remote lab?
- Many studies find that students (unsurprisingly) feel isolated from the lab itself. How can student perceptions of personal involvement be improved in remote labs?
- Studies seem to suggest students "believe" in the results from remote labs (eg [Lowe et al., 2013]). What is the source of student belief in the veracity of results? If real and simulated data were substituted would students be convinced?
- Can sensory input to and output from the user help improve feelings of immersion and presence?
  - Does something as simple as a microphone in the remote lab improve physical "connection" with the lab?
- How does student perception of the "remoteness" of a lab change when they are given more interaction with the lab interface?
- How can the interactivity available in remote labs be improved?
  - Students making choices in terms of interface modules to include in their experiment.
  - Ability for students to select different experiments to complete with the remote lab.

1.2 Phase 2+: Beyond controls

The following is a list of general research questions that could be considered beyond the first year of my PhD.

- What components of a remote lab interface most encourage student exploration of a practical lab?
- How can data analytics be used to improve the evaluation of remote labs and students using the labs?
- How can collaborative learning be promoted within remote labs?
  - Inclusion of a live messaging feature
  - Data/results sharing
  - Labs requiring simultaneous input from multiple users to accomplish tasks?
- How can we replicate some of the communication that students would receive in a traditional lab?
- Inclusion of student learning models to help evaluate student interaction with the remote lab and their learning.
- What study designs will best help evaluate the direct impact of individual elements of a remote lab and/or help quantitatively identify correlations between learning outcomes and the remote lab/interface?
- How can remote labs help improve students' sense of attempting "real" scientific enquiry; as opposed to following set tasks with known results?
- Can 'gamification' help improve student involvement in remote labs? The inclusion of 'accomplishments', for example.

2 Impact of Design Research on remote lab design

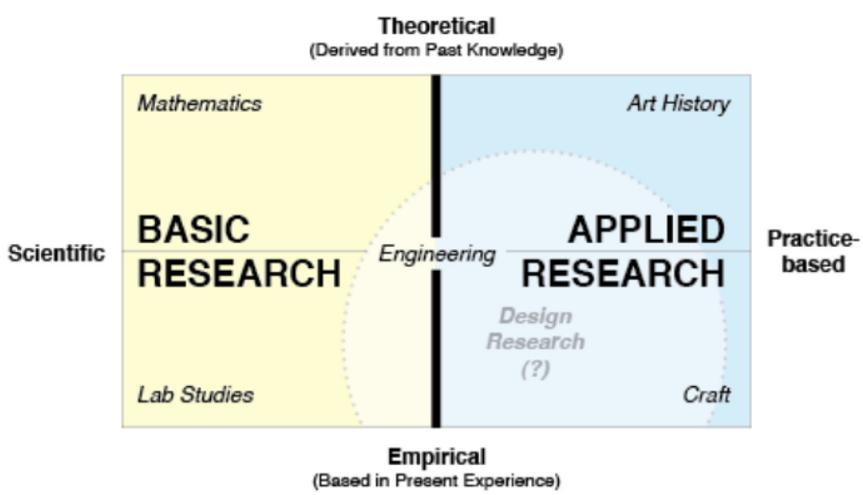


Figure 1: Reproduced from [Faste and Faste, 2012].

Design research, rather than being a series of specific methodologies on how to "do design" is instead a loose philosophy that puts importance on evaluating the design *process* rather than focusing solely on the end *product*. [Faste and Faste, 2012] highlight the fact that design research "has come to mean wildly differing things to different people" so they attempt to break down design research into more concrete categories, of which the most important for remote lab design could be *research through design: a research process that culminates in an artifact that summarises the body of knowledge gained.* In this regard, a remote lab should not be built as a static artifact that can then be used in a study to decide whether it works or not; but iterated over through a series of exploratory studies until the artifact (remote lab and interface) match what the research says is necessary/useful/preferred etc. Key aspects of attempting design research include (from [Faste and Faste, 2012]):

- Qualitative research observing people in context.
- Prototyping of forms to evaluate usability using field trials.
- Exploratory designs and tests to help inform future paths.

The creation of prototypes is essential to design research, as is the use of prototypes in the field to garner feedback which informs new iterations of the design. There is no assumption of an ultimate, best design, research can only help you tend towards more *useable* designs that meet user needs better. There can also be no assumption of the theoretical user, user research is integral to good design and user feedback is necessary at all stages.

Perhaps, overall, the evaluation of remote labs in light of design research suggests that how we assess the impact of remote labs could be framed differently. Rather than the quest for scientific knowledge (e.g. students achieve a greater increase in learning from pre to post test using remote labs compared to traditional labs); we should be framing the research in terms of user experience and component design (e.g. students believed that the interface was comparably complex to the real experience; students found graphing to be an easier process in the remote lab in comparison to the traditional lab; how are data collection times affected by varying interface design? etc. etc.)

3 Examples of current practice

3.1 Exploratory study of the acceptance of two individual practical classes with remote labs

[Tirado-Morueta et al., 2018] use 98 engineering students in an exploratory, pseudo-experimental study of two remote labs. There is no control or comparison group completing a traditional lab. The study is between-subjects, with individual students only exposed to one remote lab. Results are gathered through the use of a questionnaire, reproduced in figure 2. Students responded on a 5 point Likert scale to questions on usefullness (Q1 and 2), usability (Q3,4 and 5) and student perception of "proximity" to the lab (Q6 and 7).

Table 1. Questionnaire, descriptive analysis, and ANOVA.

		EMRL		PSRL		ANOVA
		Mean	SD	Mean	SD	*F*
Q1	This lab work allows strengthening the theoretical concepts?	4.00	0.87	4.15	0.69	0.72
Q2	The remote lab work makes easier the theoretical-practice understanding?	4.11	0.68	3.71	0.64	7.95**
Q3	The exercise can be carried out without the professor supervision?	4.11	0.80	4.54	0.59	7.65**
Q4	The access to the remote lab through the LMS is easy?	4.18	0.68	4.59	0.67	7.77**
Q5	The time available to complete the tests is enough?	4.00	0.82	4.76	0.53	24.78***
Q6	The use of the interface is easy?	4.07	0.89	4.02	0.79	0.05
Q7	The interface allows carrying out the exercise in the same way as in the lab classroom?	3.89	0.95	4.37	0.53	7.89**

$^{**}p < .01$, $^{***}p < .001$.

Figure 2: Questionnaire and student responses, reproduced from [Tirado-Morueta et al., 2018].

3.2 Pedagogical evaluation of remote laboratories in eMerge project

[Lang et al., 2007] again use engineering undergraduates undertaking a practical course in engineering. They randomly assigned participants (all from one single population of engineering students on the same course) to one of two groups: traditional practical work (control) and remote lab (experimental). Students were not allowed to select their group in order to remove selection bias. The study proposes a number of hypotheses including: *"do students find experiments carried out on the internet useful?"* and *"are the experiments considered positive and helpful to the learning process?"*. A 6 point Likert scale was used for student questionnaire responses. Results are given in figure 3.

Table 5. Mean values and standard deviations on items answered by students according to acceptance, usability and usefulness.

Scale	Item[a]	M[b]	SD
Acceptance	I would wish that other practical courses were put online.	3.77	1.17
	I would do this course again.	3.67	1.12
	I prefer learning in the eLab compared to learning with textbooks.	3.67	1.49
	If there is a choice between web-based learning and learning with textbooks I will prefer the book. (−)	3.53	1.72
	I enjoyed using the eLab.	3.97	1.03
	It's much better to see the experiments in reality than in the eLab. (−)	2.07	1.08
	I would recommend a fellow student to take this course, too.	3.87	0.97
	The eLab motivated me to learn more about the topic.	3.07	0.91
	The fact of being in a PC room instead of a physical lab disturbs me. (−)	4.29	1.46
Usability	Much effort is required to learn how to use the eLab. (−)	4.77	1.36
	I would like to have a preliminary course for using each instrument. (−)	3.73	1.28
	The instruction to make use of the eLab is sufficient.	4.17	1.29
	The help functions are sufficient.	3.83	1.12
	To get along with the notebook-tool is difficult. (−)	4.10	0.84
	The use of the textbook tool is complex. (−)	4.27	0.91
Usefulness	The eLab is a good preparation for my exams.	2.53	1.04
	For my studies it was helpful to take this course.	3.10	1.24
	The eLab is a good preparation for my future job.	3.07	1.26
	Being in the eLab was a waste of time. (−)	4.79	1.21
	Taking the course was the right step to achieve my professional goals.	3.24	0.95

[a](−): polarity reversed for this item.
[b]answer format from 1 = 'I totally refuse' to 6 = 'I completely agree'.
M: mean.
SD: standard deviation.

Figure 3: Questionnaire results on acceptance, usefulness and usability, reproduced from [Lang et al., 2007].

4 PID Controller remote lab

The remote lab is being designed with a number of possible benefits in mind: a) to replicate a teaching resource for a course with increasing student numbers; b) to act as a test case for the continued introduction of remote labs at Edinburgh; c) to allow for my training through the lifecycle of design, construction, deployment and evaluation of a remote lab; and d) as an exploratory study to provide jumping off points for later study.

The 'Control and Instrumentation Engineering 3' course at Edinburgh has a practical lab element that already utilises the Quanser QUBE and a LabVIEW interface for PID controller design (see figures 4 and 5).

Figure 4: The Quanser QUBE unit.

The lab involves students tuning a PID controller using the Ziegler- Nichols method and to tune for maximum overshoot of less than 10%. It then asks students to use these tuned parameters to setup a steering control system on a simulated vehicle. The remote lab will be created in order to perform the same tasks - details in the controls 3 course practical workbook.

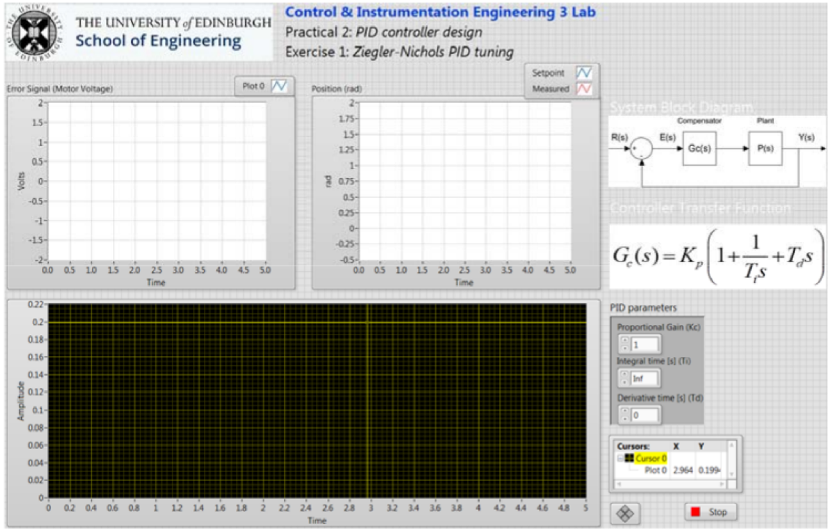


Figure 5: LabVIEW interface for the PID Controller practical laboratory in the 'Controls 3' engineering course at Edinburgh.

4.1 Hypotheses and independent variables

In my, albeit limited, review of the literature the consideration of the remote lab ***user interface*** (UI) as a potential contributor to student perceptions of the remote lab is a little lacking. I did not come across any studies within the remote lab literature of UI studies - A/B studies presenting variations of interfaces for example - although this is a standard research method within HCI research. I would propose therefore that we look at how elements of the user interface itself can help promote student perceptions of remote labs.

There are a number of hypotheses/questions that we could attempt to answer using the PID controller remote lab. I have highlighted a few below for discussion:

- Students are more likely to believe that they have more personal involvement in a practical task when they have increased control of the user interface on a remote lab.
  - Different groups are presented with either: pre-designed static interfaces, limited choice in variables etc. or full control of interface elements in use.
- A PID controller remote lab (with varying levels of UI interaction) provides the same perception of usefulness and usability as the traditional hands-on lab.
- Student feelings of "immersion" within remote labs education improves with the use of contextual media within the remote lab.
  - PID controller with and without additional vehicle control interface.
- Student perceptions of having completed a 'ready-to-go' experiment are reduced when they are given choice over interface elements to include in their practical.
- Do student perceptions of the 'ease of gathering data' and 'reliability of results' reduce when the level of customisation of the interface is changed?

4.2 Methodology and experimental design

This will be an "in the field" intervention, with students completing the same lecture courses and labs prior to the PID controller lab in which students will then be presented with alternative interfaces.

With a design research perspective, the experimental techniques employed could follow similar methods to UX/HCI research. A brief list of possible approaches is below:

- User feedback: utilising a number of exploratory studies, prior to the controls course, to gather feedback on prototype variations for the interface and lab itself.

- A/B testing: different groups of students will be exposed to different interfaces.
- Interaction analysis: user data is tracked as they use the remote lab: time taken, elements opened, options chosen etc.
- Field testing: students are observed while using the remote lab and asked to feedback on what they are doing or struggling with.
- Usability data: are students able to complete the workbook tasks? How long does it take them? Is there a difference in accuracy with each interface?
- User attitudinal questionnaire: modelled on [Lowe et al., 2013].
- User emotional response questionnaire: as in [Drysdale and Scoles, 2019].
- Pre and post test data: on either related engineering questions, questions from the lab workbook or experimental skills in general.

With the intention of maintaining as natural a setting as possible, a between-subjects design is necessary where groups of students are only exposed to one condition. This requires a larger number of participants for statistical significance than within-subjects, but this shouldn't be an issue if the number of students in the class is approximately 150-200. A between-subjects design removes the learning effect and reduces user fatigue having to complete the same task on multiple systems. However, it will reduce the ability of students to compare their experiences with different interfaces. Ultimately, we may need to provide students with opportunities beyond the study to access the alternative methods (open up the remote lab to all and/or give students additional lab time).

Students all come from the target population as they are all engineering students taking part in engineering practicals. They will be randomly assigned to one of the test conditions.

Practical work will follow the current syllabus and all students will be following the 'Controls and Instrumentation Engineering 3' practical workbook (practical 2 PID Controller). This will hopefully reduce any confounding variables involved in different preparation materials etc. Data for evaluation of the remote lab could include marked practical workbooks.

4.3 Variations on interface design

If we are studying how variations of interface customisation/interactivity affect student perceptions, then multiple variations of the PID Controller interface will need to be designed. Figure 5 shows the current interface used by students in this course. A similarly designed interface would be created as the static/non-customisable version of the remote lab interface. Something nearing that has already been created for a pure simulation version of a vehicle steering PID control, see figure 6. As a step from no user control to complete user control, a second version could include additional vehicle setup parameters (max speed, max steering angle, mass etc.) and the ability to vary what is plotted (distance error, angle error, speed, steering angle etc.). The third variation, with most user control, could effectively start with a blank screen and provide options for interface elements to reveal, including additional elements that are not necessarily useful for this practical.

References

[Drysdale and Scoles, 2019] Drysdale, T. and Scoles, J. (2019). Performing engineering digital literacies in context. In *Submitted to the 47th SEFI Annual Conference, Budapest.*

[Faste and Faste, 2012] Faste, T. and Faste, H. (2012). Demystifying "design research": Design is not research, research is design. In *IDSA Education Symposium 2012.*

[Lang et al., 2007] Lang, D., Mengelkamp, C., Jäger, R. S., Geoffroy, D., Billaud, M., and Zimmer, T. (2007). Pedagogical evaluation of remote laboratories in emerge project. *European Journal of Engineering Education.*

[Lowe et al., 2013] Lowe, D., Newcombe, P., and Stumpers, B. (2013). Evaluation of the use of remote laboratories for secondary school science education.

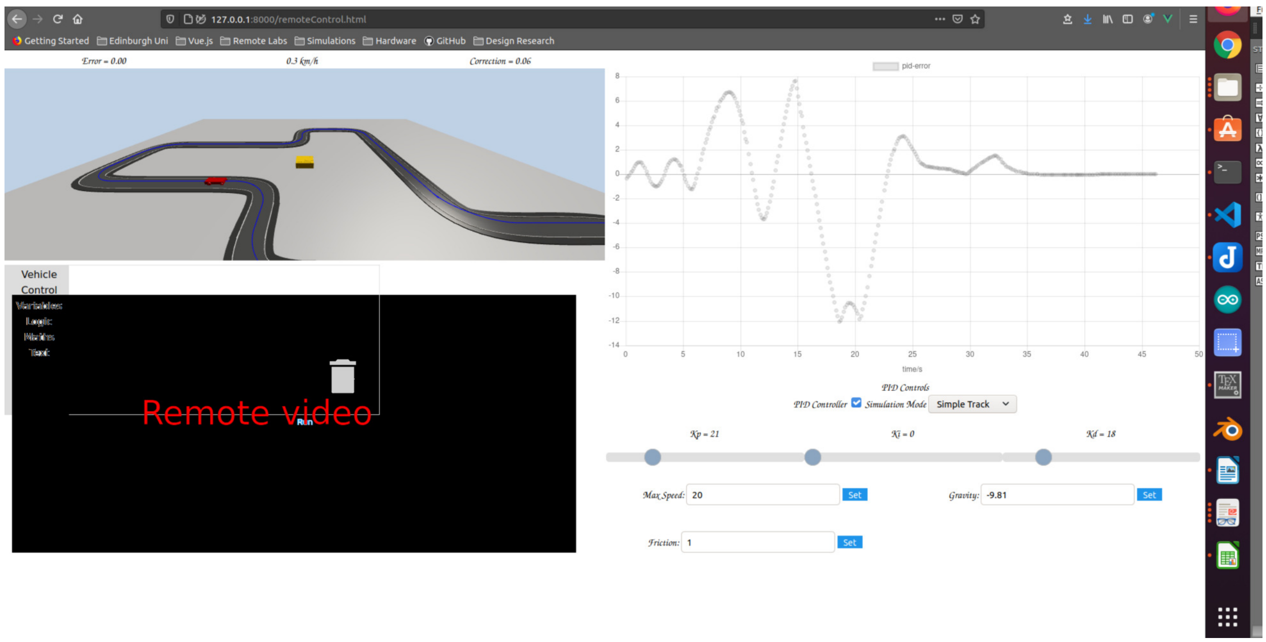


Figure 6: Basic prototype of the PID Controller interface showing ammo.js car on track, error output graph, PID parameter sliders and temporary, example settings.

[Tirado-Morueta et al., 2018] Tirado-Morueta, R., Sánchez-Herrera, R., Márquez-Sánchez, M. A., Mejías-Borrero, A., and Andujar-Márquez, J. M. (2018). Exploratory study of the acceptance of two individual practical classes with remote labs. *EUROPEAN JOURNAL OF ENGINEERING EDUCATION*.

Remote Lab PID Controller: Experimental Design

[redacted]

September 2, 2020

1 Research question

The general research question is: *how do student perceptions of remote labs vary with increasing levels of user interface interactability.*

Perceptions include student belief that they are completing a "proper" experiment; that they are collecting reliable results; that they are personally involved with the experiment; and how easily they can gathering data. Where possible, quantitative measurements of trackable interactions (task time etc.) could help support survey responses on student perceptions.

2 Participants

Approximately 250 3rd year engineering students taking part in the 'Control and Instrumentation Engineering 3' course at Edinburgh University. These students will be randomly assigned to one of three test conditions in a between-subjects design. The literature commonly has examples of studies with between roughly 10-60 participants per condition (although a small minority have over 100).

3 Hardware and software

Remote lab hardware of DC motor and spinning disk with encoder for measuring position and angular velocity. Interaction with the remote lab is through an online web app. All users will require a large screen and mouse to ensure that differences due to varying interaction technique or visibility of UI elements is controlled. No mobile phones or tablets to be used for this study.

4 Tasks

Students are following the 'Controls and Instrumentation Engineering 3' course and the remote lab is substituting for the original Quanser/LabView practical. As such, the intention is for students to complete all tasks from that original practical. A brief outline of each task is given below:

Task 1: Static characterisation of a DC motor. Plot angular velocity against input voltage.

Task 2: 1st order step response of motor with rotating mass. Vary transfer function coefficients to match real response of system for 2 different rotating disks. Calculate gain and time constant from output graph.

Task 3: 2nd order step and ramp response. Input step and ramp signals to set position of disk. Calculate peak time and percentage overshoot for step reponse. Measure steady state error for ramp response.

Task 4: PID Controller. Use the 'Ultimate gain' Ziegler-Nichols method to tune the PID controller parameters.

Task 5: Car speed control: Exploring a "real" PI controller.

Task 6: Car steering control: Exploring a "real" PD controller.

4.1 Possible issues

The current practical utilises 6 different LabView interfaces to complete the full practical. This is potentially a lot of new interfaces to develop, particularly if comparing multiple versions of each.

The remote lab will need to be designed with two separate motors as the original lab asks students to use two different disks.

Lab material provided to the students will need to be re-written and ensure it is not biased towards any particular interface. General tasks need to be given, rather than specifics on how to perform the task with the interface. The same lab material needs to be provided to each student no matter the interface used.

4.2 Completing tasks

As this is an 'in the wild' evaluation of the remote lab, complete control over when and where students complete tasks is not possible. They will be given access to the web app at the appropriate time(s) during their lecture course. They could be given a time frame within which the lab work must be completed and appropriate data submitted. They will be asked to complete tasks on a desktop PC (any OS) with keyboard and mouse. Hardware used as well as time and location (home, university etc.) could be logged.

5 Variables

5.1 Independent variables

Factor: user interface. ***Levels***: Basic, customisable, exploratory.

Description of levels: Basic - a static page (as with the LabView interface) with only the necessary tools, inputs and outputs for PID tuning. Customisable - interaction elements can be switched on/off, only necessary elements exist. Exploratory - additional elements exist, also connected to data collection/input/output, but are additional tools to help exploration of data.

5.2 Dependent variables and collecting data

Student perceptions will be collected through a questionnaire(s). Student interviews could also be undertaken with a random sub-sample of students. Field testing can also be performed, asking students to provide live feedback on what they are doing with the interface as they use it.

User Experience Questionnaire (UEQ) measures across 6 scales: attractiveness, perspicuity, efficiency, dependability, stimulation and novelty.[Laugwitz et al., 2008]

A student survey of opinions after using the remote lab, similar in design to [Lowe et al., 2013].

Pre and post assessments on "emotional response" as in [Drysdale and Scoles, 2019].

Interaction analysis: data can be tracked on a) task completion time b) number of elements interacted with (i.e. do students with the more exploratory interfaces actually explore the interface?) c) accuracy of calculated coefficients during tasks.

5.3 Control variables

Students may use personal devices as well as school computers, but need to ensure reasonable size screen and use of mouse.

Same lecture series attended and same lab material provided for completing tasks.

Random assignment of students to UI conditions.

5.4 Confounding variables

Do students have prior experience with LabView? Do lab materials benefit one interface over another? Has location or hardware used affected perceptions?

References

[Drysdale and Scoles, 2019] Drysdale, T. and Scoles, J. (2019). Performing engineering digital literacies in context. In *Submitted to the 47th SEFI Annual Conference, Budapest*.

[Laugwitz et al., 2008] Laugwitz, B., Schrepp, M., and Held, T. (2008). Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work*.

[Lowe et al., 2013] Lowe, D., Newcombe, P., and Stumpers, B. (2013). Evaluation of the use of remote laboratories for secondary school science education.

Remote Lab PID Controller: Experimental Design

[redacted]

September 14, 2020

1 Research question

The general research question is: *how do student perceptions of remote labs vary with increasing levels of user interface interactability.*

Perceptions include student belief that they are completing a "proper" experiment; that they are collecting reliable results; that they are personally involved with the experiment; and how easily they can gathering data. Where possible, quantitative measurements of trackable interactions (task time etc.) could help support survey responses on student perceptions.

2 Participants

Approximately 250 3rd year engineering students taking part in the 'Control and Instrumentation Engineering 3' course at Edinburgh University. These students will be randomly assigned to one of three test conditions in a between-subjects design. The literature commonly has examples of studies with between roughly 10-60 participants per condition (although a small minority have over 100).

3 Hardware and software

Remote lab hardware of DC motor and encoder for measuring position and angular velocity. Interaction with the remote lab is through an online web app. The web app should be responsive to screen size and designed with both desktop and mobile use in mind. However, a keyboard and mouse is desirable for more fluent manipulation of the interface. In order to control for differences in interaction technique, it will be important to log users screen sizes when they are using the web app. They will also be encouraged to use a large screen, keyboard and mouse.

The current intention is to have two variations on the PID remote lab to cover the tasks. Tasks 1-3 will use a variable governor in which the position of two rotating masses can be changed to vary the moment of inertia of the system. This hardware covers the 1st order angular velocity problems. The second version, a variable length robotic arm, will be used for the second order PID position controller tasks (4-6 below).

A virtual model (virtual twins) of the system(s) will be developed (in three.js and ammo.js) for prototyping purposes. This will allow interface design to begin before all the hardware is completed. It may also offer students a system to explore outside of any time/scheduling contraints of the remote lab.

4 Tasks

Students are following the 'Controls and Instrumentation Engineering 3' course and the remote lab is substituting for the original Quanser/LabView practical. As such, the intention is for students to complete all tasks from that original practical. A brief outline of each task is given below:

Task 1: Static characterisation of a DC motor. Plot angular velocity against input voltage.

Task 2: 1st order step response of motor with rotating mass. Vary transfer function coefficients to match real response of system for 2 different rotating disks. Calculate gain and time constant from output graph.

Task 3: 2nd order step and ramp response. Input step and ramp signals to set position of disk. Calculate peak time and percentage overshoot for step reponse. Measure steady state error for ramp response.

Task 4: PID Controller. Use the 'Ultimate gain' Ziegler-Nichols method to tune the PID controller parameters.

Task 5: Car speed control: Exploring a "real" PI controller.

Task 6: Car steering control: Exploring a "real" PD controller.

4.1 Possible issues

The current practical utilises 6 different LabView interfaces to complete the full practical. This is potentially a lot of new interfaces to develop, particularly if comparing multiple versions of each.

The remote lab will need to be designed with two separate motors as the original lab asks students to use two different disks.

Lab material provided to the students will need to be re-written and ensure it is not biased towards any particular interface. General tasks need to be given, rather than specifics on how to perform the task with the interface. The same lab material needs to be provided to each student no matter the interface used.

4.2 Completing tasks

As this is an 'in the wild' evaluation of the remote lab, complete control over when and where students complete tasks is not possible. They will be given access to the web app at the appropriate time(s) during their lecture course. They could be given a time frame within which the lab work must be completed and appropriate data submitted. They will be asked to complete tasks on a desktop PC (any OS) with keyboard and mouse. Hardware used as well as time and location (home, university etc.) could be logged.

5 Variables

5.1 Independent variables

5.1.1 Iteration 1:

Factor: user interface. ***Levels***: Limited choice vs full range of choice.

The main vehicle for student exploration within the remote lab is the moment of inertia of the spinning governor/robot arm. One version of the UI could include only 2 or 3 pre-set positions for the governor accessible via a simple button/dropdown selection. A second version could expose the full range of governor positions, with students in complete control of where they want to position it.

5.1.2 Iteration 2:

Factor: user interface. ***Levels***: Students given measurements vs students perform measurements.

In the traditional lab, students are required to make measurements of the disk dimensions for calculation of the moment of inertia. Hands on measurement with the equipment could be a major contributing factor to students' perceptions of having undertaken a "real" experiment. In a remote lab this capacity to perform physical measurements is often not present. One UI could simply provide the students with the physical dimensions of the governor/robot arm. A second variation could provide on screen tools for making the measurements themselves (see Fig 2).

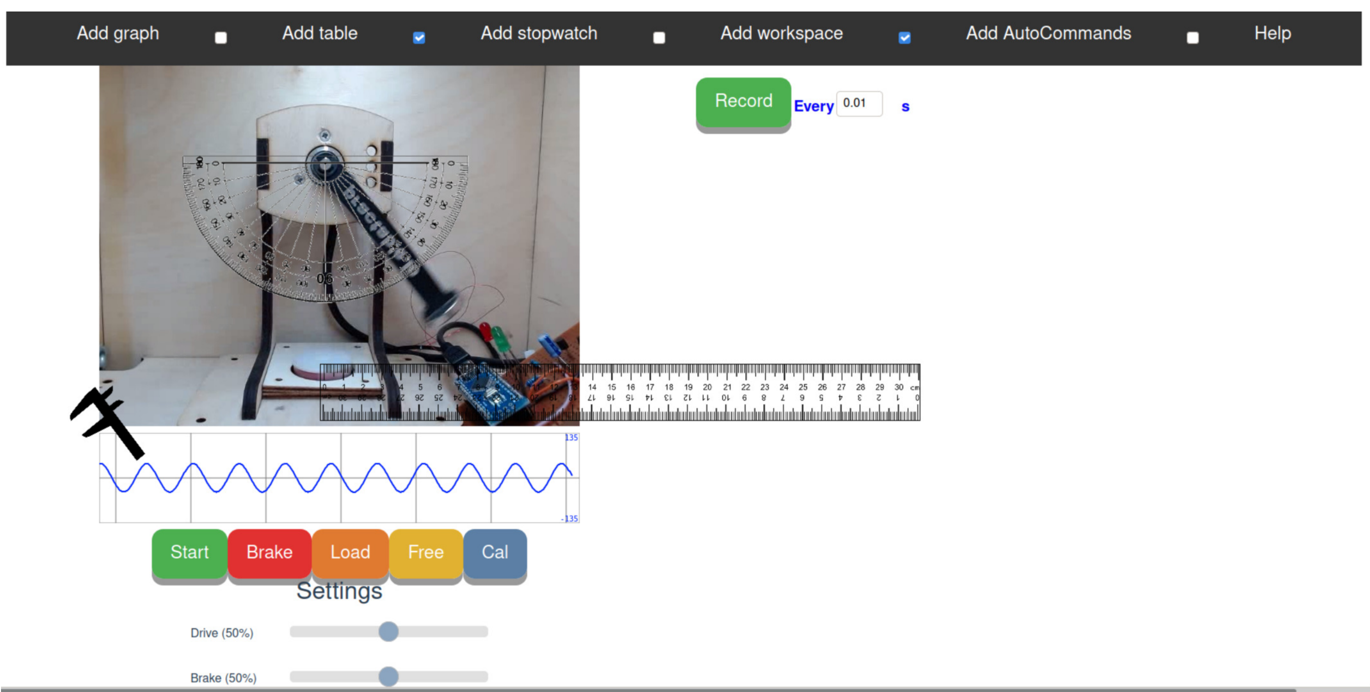


Figure 1: Screen grab of the 'pendvueino' remote lab interface with protractor and ruler measuring tools.

5.1.3 Iteration 3:

Factor: user interface. ***Levels***: Single graphical output of key variable (angle or angular velocity) vs choice of variable to plot and/or other graphing tools (displaying tabular data, gradient calculator, function plotter and exporting to CSV file).

A potential major aspect of student belief in the reliability of an experiment is in terms of the data collection and analysis. Many remote labs currently display data automatically and provide the basic tools (and nothing more) for processing the data.

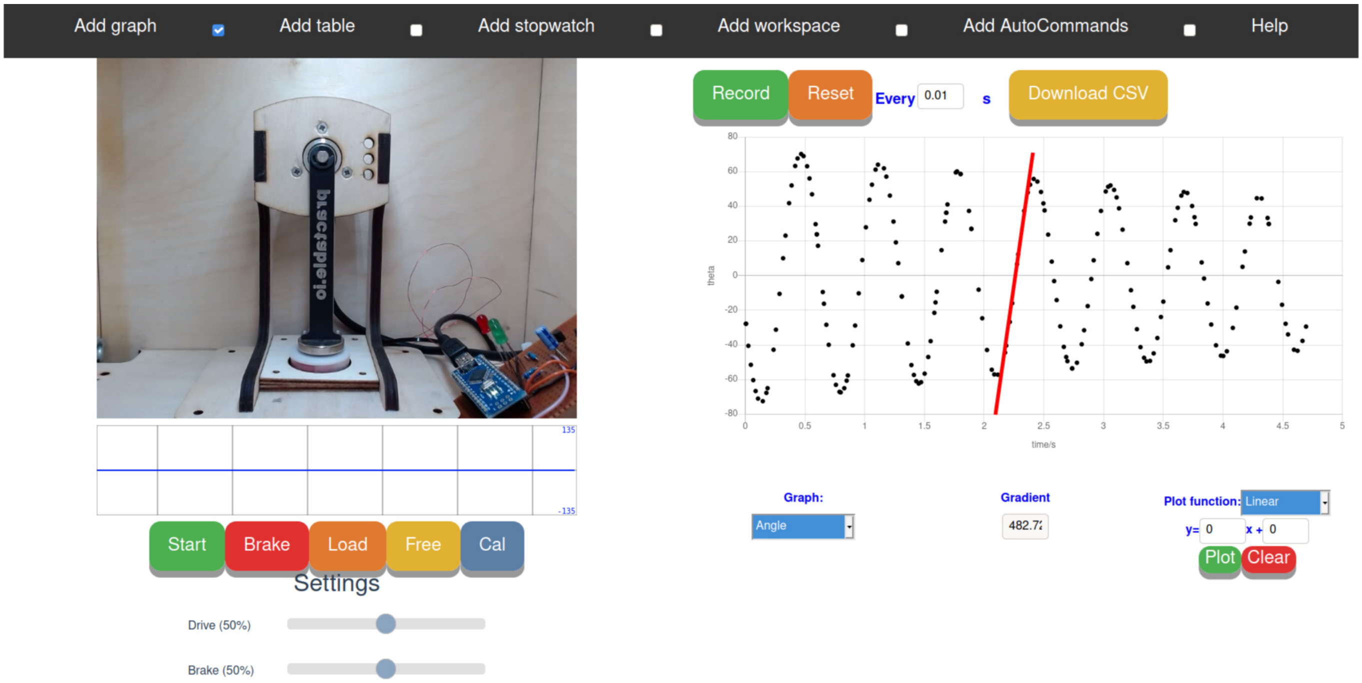


Figure 2: Screen grab of the 'pendvueino' remote lab interface with graphing tools including graphing different quantities, gradient calculation tool and function plotter.

5.2 Dependent variables and collecting data

Student perceptions will be collected through a questionnaire(s). Student interviews could also be undertaken with a random sub-sample of students. Field testing can also be performed, asking students to provide live feedback on what they are doing with the interface as they use it.

User Experience Questionnaire (UEQ) measures across 6 scales: attractiveness, perspicuity, efficiency, dependability, stimulation and novelty.[Laugwitz et al., 2008]

A student survey of opinions after using the remote lab, similar in design to [Lowe et al., 2013].

Pre and post assessments on "emotional response" as in [Drysdale and Scoles, 2019].

Interaction analysis: data can be tracked on a) task completion time b) number of elements interacted with (i.e. do students with the more exploratory interfaces actually explore the interface?) c) accuracy of calculated coefficients during tasks.

5.3 Control variables

Students may use personal devices as well as school computers, but need to log the interaction method used (keyboard + touchpad, keyboard and mouse, touchscreen, 17 inch desktop monitor, 13 inch laptop....)

Same lecture series attended and same lab material provided for completing tasks.

Random assignment of students to UI conditions.

5.4 Confounding variables

Do students have prior experience with LabView? Do lab materials benefit one interface over another? Has location or hardware used affected perceptions?

References

[Drysdale and Scoles, 2019] Drysdale, T. and Scoles, J. (2019). Performing engineering digital literacies in context. In *Submitted to the 47th SEFI Annual Conference, Budapest*.

[Laugwitz et al., 2008] Laugwitz, B., Schrepp, M., and Held, T. (2008). Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work*.

[Lowe et al., 2013] Lowe, D., Newcombe, P., and Stumpers, B. (2013). Evaluation of the use of remote laboratories for secondary school science education.

Remote Lab PID Controller: Experimental Design

[redacted]

October 5, 2020

1 Research question

The general research question is: *how do student learning and perceptions of remote labs vary with increasing levels of user interface control?*

Students will complete the same tasks on the same remote hardware, but differences will exist in the level of control that users have over a) setup of the remote apparatus, b) measurements being made and/or c) data analysis and variables.

Perceptions include student belief that they are completing a "proper" experiment; that they are collecting reliable results; that they are personally involved with the experiment; and how easily they can gathering data. Where possible, quantitative measurements of trackable interactions (task time etc.) could help support survey responses on student perceptions.

Possible relationships that could be explored with the data collected (section 5.2) include:

- Is the level of control a student has over the UI correlated with a belief in the reliability of data?
- Is the level of control a student has over the UI correlated with student belief in having completed an experiment with unknown results?
- Do students with more controllable UIs have higher levels of satisfaction when completing the practical?
- Do students with more controllable UIs have greater learning gains as measured by the control theory question/lab report grade/submitted task answers?
- Do students with more controllable UIs rate the usability lower or higher than those with less controllable UIs?
- Is the usability of the UI correlated with learning gains?
- Is usability of the UI correlated with satisfaction?
- Does feeling of immersion have a negative correlation with usability?
- Which UI features do students use most in completing the practical tasks?
- Which UI features to students believe had most impact on their learning?

2 Participants

Approximately 250 3rd year engineering students taking part in the 'Control and Instrumentation Engineering 3' course at Edinburgh University. These students will be randomly assigned to one of three test conditions in a between-subjects design. The literature commonly has examples of studies with between roughly 10-60 participants per condition (although a small minority have over 100).

3 Hardware and software

Remote lab hardware of DC motor and encoder for measuring position and angular velocity. Interaction with the remote lab is through an online web app. The web app should be responsive to screen size and designed with both desktop and mobile use in mind. However, a keyboard and mouse is desirable for more fluent manipulation of the interface. In order to control for differences in interaction technique, it will be important to log users screen sizes when they are using the web app. They will also be encouraged to use a large screen, keyboard and mouse.

The current intention is to have two variations on the PID remote lab to cover the tasks. Tasks 1-3 will use a variable governor in which the position of two rotating masses can be changed to vary the moment of inertia of the system. This hardware covers the 1st order angular velocity problems. The second version, a variable length robotic arm, will be used for the second order PID position controller tasks (4-6 below).

A virtual model (virtual twins) of the system(s) will be developed (in three.js and ammo.js) for prototyping purposes. This will allow interface design to begin before all the hardware is completed. It may also offer students a system to explore outside of any time/scheduling contraints of the remote lab.

The interfaces will be designed using Vue.js.

4 Tasks

Students are following the 'Controls and Instrumentation Engineering 3' course and the remote lab is substituting for the original Quanser/LabView practical. As such, the intention is for students to complete all tasks from that original practical. A brief outline of each task is given below:

Task 1: Static characterisation of a DC motor. Plot angular velocity against input voltage.

Task 2: 1st order step response of motor with rotating mass. Vary transfer function coefficients to match real response of system for 2 different rotating disks. Calculate gain and time constant from output graph.

Task 3: 2nd order step and ramp response. Input step and ramp signals to set position of disk. Calculate peak time and percentage overshoot for step reponse. Measure steady state error for ramp response.

Task 4: PID Controller. Use the 'Ultimate gain' Ziegler-Nichols method to tune the PID controller parameters.

Task 5: Car speed control: Exploring a "real" PI controller.

Task 6: Car steering control: Exploring a "real" PD controller.

The practical workbook will need to be adapted for use with the remote lab. An additional, final task could be added, asking students to plan and run a brief experiment of their own. Does the more exploratory interface encourage more exploration?

4.1 Possible issues

The current practical utilises 6 different LabView interfaces to complete the full practical. This is potentially a lot of new interfaces to develop, particularly if comparing multiple versions of each.

The remote lab will need to be designed with two separate motors as the original lab asks students to use two different disks.

Lab material provided to the students will need to be re-written and ensure it is not biased towards any particular interface. General tasks need to be given, rather than specifics on how to perform the task with the interface. The same lab material needs to be provided to each student no matter the interface used.

4.2 Completing tasks

As this is an 'in the wild' evaluation of the remote lab, complete control over when and where students complete tasks is not possible. They will be given access to the web app at the appropriate time(s) during their lecture course. They could be given a time frame within which the lab work must be completed and appropriate data submitted. They will be asked to complete tasks on a desktop PC (any OS) with keyboard and mouse. Hardware used as well as time and location (home, university etc.) could be logged.

5 Variables

5.1 Independent variables

The following sections highlight three possible variations on the UI study. Each considers a relatively small element of the overarching idea of *interactability*. The first considers increased student choice on the experimental parameters; the second changes the level of "physical" measurements that can be done with the interface; and the third deals with the level of interaction possible with displayed data.

5.1.1 Iteration 1:

Factor: user interface. ***Levels***: Limited choice vs full range of choice.

The main vehicle for student exploration within the remote lab is the moment of inertia of the spinning governor/robot arm. One version of the UI could include only 2 or 3 pre-set positions for the governor accessible via a simple button/dropdown selection. A second version could expose the full range of governor positions, with students in complete control of where they want to position it.

5.1.2 Iteration 2:

Factor: user interface. ***Levels***: Students given measurements vs students perform measurements.

In the traditional lab, students are required to make measurements of the disk dimensions for calculation of the moment of inertia. Hands on measurement with the equipment could be a major contributing factor to students' perceptions of having undertaken a "real" experiment. In a remote lab this capacity to perform physical measurements is often not present. One

UI could simply provide the students with the physical dimensions of the governor/robot arm. A second variation could provide on screen tools for making the measurements themselves (see Fig 2).

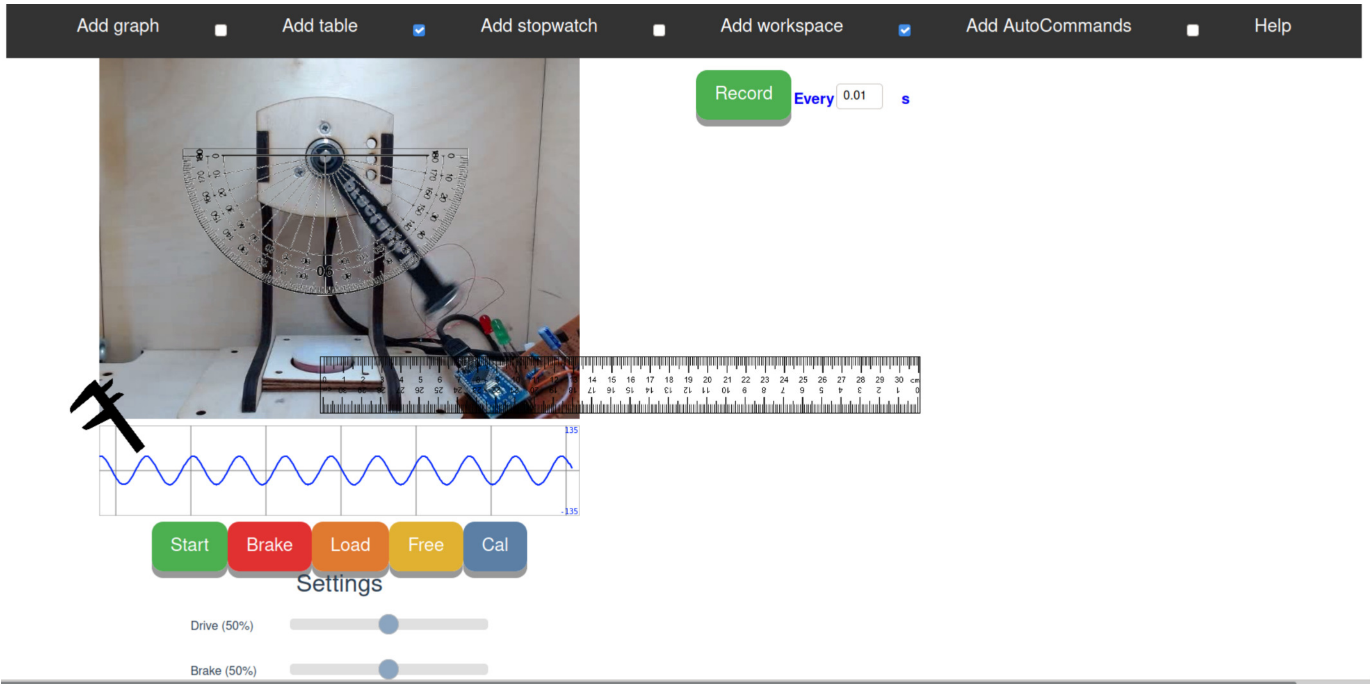


Figure 1: Screen grab of the 'pendvueino' remote lab interface with protractor and ruler measuring tools.

5.1.3 Iteration 3:

Factor: user interface. ***Levels***: Single graphical output of key variable (angle or angular velocity) vs choice of variable to plot and/or other graphing tools (displaying tabular data, gradient calculator, function plotter and exporting to CSV file).

A potential major aspect of student belief in the reliability of an experiment is in terms of the data collection and analysis. Many remote labs currently display data automatically and provide the basic tools (and nothing more) for processing the data.

The UI can be built with all these elements and then hiding or exposing them for specific conditions is straightforward.

5.2 Dependent variables and collecting data

Background data on students:

- Year
- Gender
- Engineering school: mechanical, electrical etc.
- Student academic achievement will be collected: previous module results or previous lab results? This is a measure of student prior achievement before the lab.

Student emotional responses and perceptions: A student survey will cover student perceptions, beliefs and emotional responses to the remote lab work. Example survey questions given in section 6. A pre and post survey should be given to understand how perceptions change. Survey will be similar in design to [Lowe et al., 2013] and [Drysdale and Scoles, 2019].

Student learning during remote lab:

- A short control theory question related to the practical work should be completed as a measure of learning during the remote lab. Questions should cover the range of learning intentions and necessary skills to complete the lab tasks. Perhaps this could be delayed to a short time after completing the lab tasks to measure learning retention.
- The overall result of their lab report.
- Questions in the practical workbook can be coded as demonstrating knowledge, understanding, measurement, analysis, investigation etc. Differences in UI can then be correlated with the types of question.
- Submission of calculated values during tasks. The task can be selected via a dropdown and an answer submitted.

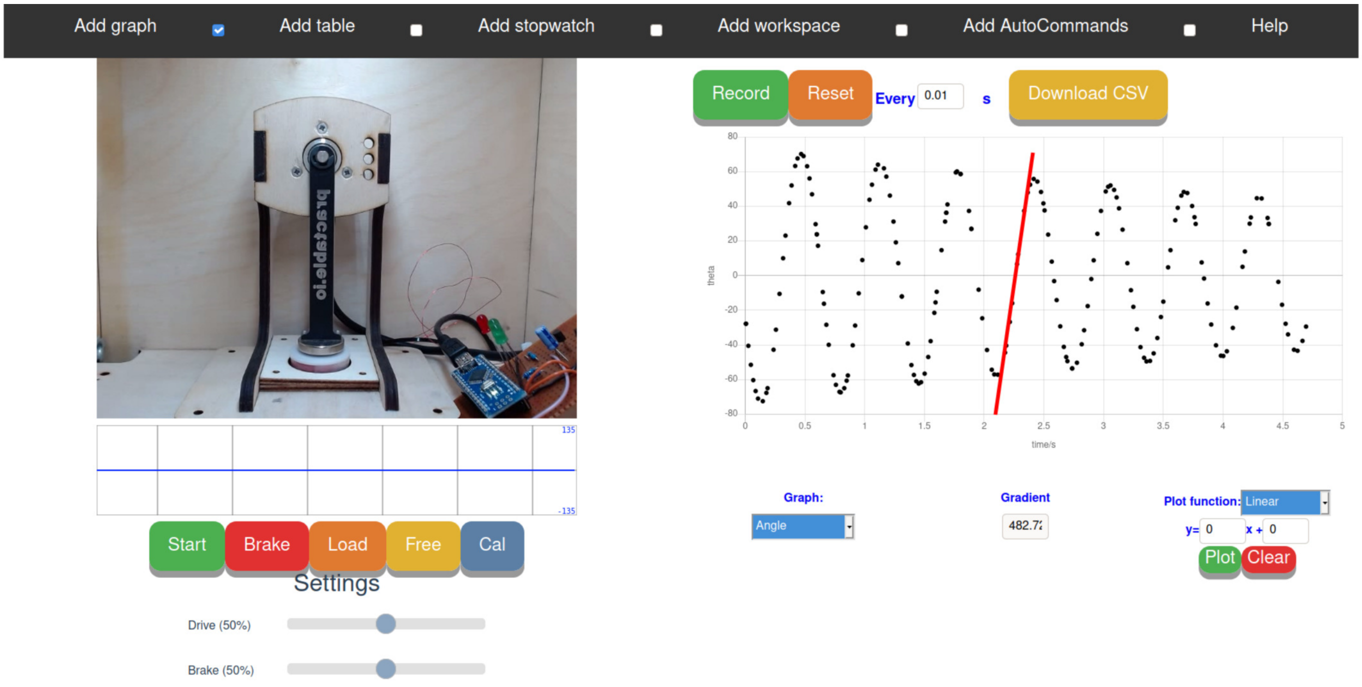


Figure 2: Screen grab of the 'pendvueino' remote lab interface with graphing tools including graphing different quantities, gradient calculation tool and function plotter.

- An additional task to investigate their own relationship between variables. An open tasks to assess whether certain UI configurations encourage or allow better exploration of the data.

Student exploration of the UI: Interaction data can be automatically logged during use of the remote lab interface. This can help understand the extent to which students actually explore the interface; the UI features that students focus on using; the time taken for students to complete tasks; and identifying whether students select the wrong tools for a task or attempt different methods when provided with additional tools? Collect data on:

- The task that a student is currently working on, via a dropdown selection box or that automatically changes once an answer is submitted.
- The UI elements hovered over or clicked with a timestamp and associated task.
- The interface features that are optionally added to the interface e.g. do students access the table feature whilst completing a task or not? Did they use the ruler tool? Did they download CSV data?
- Submission of calculated values for tasks. Are there differences in the range of submitted answers when students have more exploratory interfaces? Or are asked to make measurements on screen?

Student interviews: The potential exists for running smaller tests with a few students during the development process and completing short field trials with them, asking them to explain how they are using the interface to complete short tasks.

5.3 Control variables

Students may use personal devices as well as school computers, but need to log the interaction method used (keyboard + touchpad, keyboard and mouse, touchscreen, 17 inch desktop monitor, 13 inch laptop....)

Same lecture series attended and same lab material provided for completing tasks.

Random assignment of students to UI conditions.

5.4 Confounding variables

Do lab materials benefit one interface over another? Has location or hardware used affected perceptions?

6 Survey questions

Survey questions should cover a range of student perceptions and beliefs about their learning. 5 or 7 point Likert scales will be used where appropriate.

6.1 Rating UI features

Rate the following UI features on their usefulness in completing the practical tasks (1 - Did not use, 2 - Not useful 3 - Useful 4 - Very useful 5 - Vital

- Video feed
- Data graphing tool
- Data table tool
- Data download tool
- Inertia control tools
- Measuring tools
- etc

6.2 Student emotions

Before the remote lab practical session what emotions are you feeling about the session: [tick boxes] surprised, curious, inquisitive, interested, happy, bored, puzzled etc etc.

After the remote lab session........

6.3 Student perceptions

Rate the following statements on how much you agree with it. (1 - Strongly disagree, 2 - Disagree, 3 - Slightly disagree, 4 - Neither agree nor disagree, 5 - Slightly agree, 6 - Agree, 7 - Strongly agree.

I found it easy to navigate the interface.

I found it easy to collect the appropriate data for the tasks.

During the practical, the ***practical workbook*** encouraged me to explore the topic rather than being limited to specific tasks.

During the practical, the ***user interface*** design encouraged me to explore the topic rather than limiting me to the tasks.

I think there was only one way to complete tasks with the web app.

I believe that I completed an experiment with unknown results rather than finding out already known values.

I discovered something surprising during the practical work.

I feel more confident with the content of the lectures after having completed the practical work.

I believe that the results of my experiment are accurate (i.e. I have calculated the true value of quantities.)

I believe that the results of my experiment are reliable (i.e. if I repeated the experiment or someone else attempted the same practical they would measure the same as me.)

I would prefer to be in the same location as the physical hardware.

I felt motivated to explore my own ideas whilst using the remote lab.

I am normally enthusiastic about practical lab work.

I was enthusiastic about completing the remote lab practical work.

I am satisfied that I have learnt what I needed to during the practical work.

I enjoyed the remote lab practical work.

I prefer thought experiments compared to real experiments.

6.4 General/Open questions

What size of group did you work in for the remote lab practical? Individually, 2-3, 4+.

Did you return to the remote lab to help with your report write up? Yes/No

Are there any additional features that you would like to have had to help you complete the tasks?

Was there anything that limited you from performing your own investigation during the practical work?

References

[Drysdale and Scoles, 2019] Drysdale, T. and Scoles, J. (2019). Performing engineering digital literacies in context. In *Submitted to the 47th SEFI Annual Conference, Budapest*.

[Laugwitz et al., 2008] Laugwitz, B., Schrepp, M., and Held, T. (2008). Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work*.

[Lowe et al., 2013] Lowe, D., Newcombe, P., and Stumpers, B. (2013). Evaluation of the use of remote laboratories for secondary school science education.

J. Redacted data – Field validations

Study 1

Context note: This data was compiled collaboratively through an initial in-person workshop/review, with follow-up zoom sessions, from April 2022 through November 2022

COMMUNICATIONS SUMMARY

Participants:

- Joshua Burridge (the author)
- In person: 3 academic practitioners – 2 senior teaching academics, 1 junior teaching academic
- Online: An additional 2 teaching academics

REDACTED

PIPOV notes

PIPOV CHEATSHEET

PURPOSE: WHY ARE YOU DELIVERING THIS LABORATORY ACTIVITY?

- For students to learn:
  - instrumentation
    Students learning components of real industrial processes
  - models
    Matches theory across multiple subjects
    Intending to develop digital twins (note – not that sort of model)
  - ~~experimental design~~
    Experimental equipment setup provided and carefully managed – processes used to be developed by students, but these are largely managerial/safety processes. Future use of the equipment to be student designed, but the equipment is fixed and not at all easily moved.
  - data analysis
    Combination of in person observations and digitised data
  - ~~component/system design~~
    Not a feature, system is designed and built in-place
  - from failure
    Several guards against failure of equipment, but students can get up to those barriers. Lab technicians will 'catch errors' in between the students making them and the equipment exploding. Good to give them a safety net and fail here first.
  - ~~creativity~~
    Limited in the labs, time pressures are significant and they are in early years for the two current use cases. The student club use of the equipment is where the creativity comes in.
  - psychomotor (expertise with tools/equipment)
    A mix of physical use and digital use – none of the major components can be operated by hand by students. Measurements/readings can fit here though.
  - Safety
    A big component, the equipment is significant and there are risks. Students need to

prove to the techs that their intended uses will work safely. Need a lot of induction into spaces.
  - communication
    Not linked to the equipment, will be part of the units using. Reports like in an industrial context.
  - teamwork
    2 levels here, students working in groups in the lab, and students working in the student club together.
  - ethics
    Not sure it's given explicit treatment, comes up under compliance with legal and safety maybe? Environment and sustainability?
  - sensory awareness
    Not too sure what this is referring to? They make recordings of data in the lab.
- Engagement/Satisfaction
  This is probably the biggest motivation for the creation of the lab – that and longitudinal use to link units together.
- Purposes other than student outcomes
  Costs will go up, so it's not a cost or time saver. Visibility or student recruitment might be a valid reason? Probably engagement is the main one though.

INPUT: WHAT IS GOING INTO THE ACTIVITY?

- Experiment
  - Discipline
    Chemical engineering
  - Phenomenon under study
    A pilot plant for industrial processes that can be configured to run different water processes. Case study example being filtration.
  - Equipment
    - Degrees of freedom
      Loads of configuration options, but the equipment setup is fixed – pipes fixed to walls, specific modules in the process. Configuration is through opening valves to branch in different ways between modules.
    - Data source
      Combination of in-person readings and digital sensors – intending to develop a digital twin.
  - Nature of activities/instructions (Didactic <-> Inquiry)
    Used differently across year levels – didactic for early years, opening up as their understanding of the equipment and theory increases. They should become more familiar with the lab as they use it year by year.
- Coordination
  - With Other Material
    Still pending – the classes will be developed later.
  - With Other Students
    Groups will go into the lab together.
- Interface
  - With the equipment
    Combination of physical and digital (remote). Could it be a remote lab? Probably, and for overseas students this is an interesting question.

  - With other students
    This will most likely just be in person. Possibly some local/overseas interaction, but probably not.
  - With educators
    Local will be in person, overseas will have somebody local and the technicians obviously here with the equipment.
- Synchronicity
  - With equipment
    Haven't considered 'batch' control – maybe. It would probably help with anything that needed to sit in a module for a while. Mostly synchronous, though it's chemical processes, they take time anyway.
  - With other students
    The student groups will probably self manage, they are in class and they'll have their own chats.
  - With educators
    In person with the lab, but they can contact asynchronously as well

  Side note –sync questions are a bit hard to think through for an in-person lab.
- Student Attributes
  - Prior knowledge
    The students will have different theory understanding – the lab will run over a few weeks. Order will be different in unit 1 and unit 2. Need to consider how to manage that order split.
  - Learning preferences, habits, ability
    Not sure here – students have asked for more labs, they seem to want that hands-on. COVID has obviously been a problem.
  - Demographics
    Mix of local and international students, and a local and international campus too.
- Locations/Proximity
  - Student to student
    Proximal, they can be remote if they choose to outside of class
  - Student to equipment
    Proximal and remote mixed – overseas only remote
  - Student to educators
    Proximal for the in person use, remote access available (esp. for overseas)
- Extraneous
  - Institutional context
    Institution wants more in person experiences and better student affective survey results. Funding and industry connection appetite is definitely there.
  - Geographical context
    Need to service both students here in Australia and overseas.
  - Chronological context
    Going to stretch across multiple years and they repeatedly use the lab.

PROCESS: WHAT SHOULD OCCUR DURING THE ACTIVITY?

- Physical actions
  - Equipment manipulation
    They can't move things around, cause you can't do that with this sort of equipment. They can open this valve, or that valve, and it flows from this piece to this piece or the other way. They need to put in the right instructions for the filtration process to work the way

it's supposed to with the right valves and things opening at the right times, and the equipment set to the right processes.
  - Observation recording
    They will be able to see the outputs on the equipment, and they'll need the digital sensor data as well. They'll need to use those to evaluate how the process has worked.
  - Report writing
    An 'industry simulated' report with exec summary etc.
- Discussion actions
  - With other students
    This isn't being specified, they will work together how they decide to. They might divide and conquer, but there's one process and one report, so they should all get a feel for the different parts. They have an exam on the unit later, so they should all be motivated to participate.
  - With educators
    This is an interesting question. They will engage with the technicians with their plan, and if their plan isn't going to work or is unsafe the technician will talk them through why. That's the main pre-planned discussion, but they can always talk to their tutors or lecturer.
- Mental actions
  - Combining with theory
    They need to take the different pieces of theory and build it up into a real process end to end. Students struggle with this, they nail the little bits but connecting them is harder.

OUTCOME: WHAT ABOUT THE STUDENTS SHOULD BE DIFFERENT AFTER THE ACTIVITY?

- Knowledge/Understanding of
  - Theory
    I think this one will be the same understanding, but a better understanding? How they fit together and how they work in the real world will be better. The equipment won't match the theory 100%.
  - Equipment
    This is the big one, they have so little exposure to industrial equipment otherwise, this lab involves real equipment they might use in their careers.
  - Analysis
    The students need to understand what they've done, so they can't just report the data. They will need to analyse that and fit it to their theoretical process. So they should be developing skills here.
  - Inquiry/Design
    They need to design the process, within the pre-specified equipment. So part of this? They design what flows into what, how long, what settings.
- Affective
  - Motivation/Satisfaction
    They keep asking for more hands-on, even before COVID – so this should be an easy win. The engagement with the student club has been fantastic, they should hopefully bounce off that and have a great relationship with the lab over a few years.

VALIDATION: HOW CAN YOU OBSERVE WHETHER THE INPUTS/PROCESSES/OUTPUTS GIVEN ABOVE ACTUALLY OCCURRED? HAS THE LABORATORY DELIVERED ITS PURPOSE?

- Lab report or written assignment
  They have a report on their process, that's the main assessment.
- Quiz or exam
  They do have an exam, but that covers the whole unit.
- Survey or questionnaire
- Direct observation
- Interview

  Haven't decided on any specific research validations yet. The learning analytics dashboard idea sounds interesting. We're already planning on having the sensors, so we can definitely see what their configurations are.

Internal notes:

- 3rd year unit using the lab
- Assessment driven - not just a lab report
- EA Accreditation reqs eng competencies - alignment and purpose from these
    - Assessments demonstrate skills - not in a lab report
- Link between theory and practice
- "Practice is not authentic on single equipment"
- Process control vs Heat & Mass, two different units
- Explorative vs linked theory by experience
- First virtual, 2 hands-on (split by team?)
- Only 1 pilot plant equipment
- Around 70-80 students for control
- 120 students - 68 local, some <redacted> for heat & mass
- What 1st year labs - still open
- AR module benefits
- Whiteboard activity at start - prelab activity
- In 2 weeks pre-lab, team meeting AR lab, address questions
- AR for safety induction
    - Interim oral examination for safety, gated
- Reverse
    - Cookbook process to set up
    - Reverse being investigate existing lab
- 1 page - concise, clarity | memo to boss
- ESO failure? Australian. Honeywell? - drives relevance
- Groups of 2, all in 3 weeks. 4 per week, 4 apparatus
- 1 lab 3-4 weeks
- Practice run in first run

STUDY 2

Context note: This data was collected through an in-person workshop/review, November 2022

- COMMUNICATIONS SUMMARY

Participants:

- Joshua Burridge (the author)
- 2 academic practitioners – one experienced with remote laboratories, one novice

• REDACTED

<redacted – laboratory design specifications and student guides>

- Electrical circuit analysis
    - Prac comp. - Simulation
    - Using <redacted> (limited current)
        - ? number of labs. 3 <redacted> pracs – maybe more on different equipment
        - 1 formative
        - 2 big pracs, 2 big lab reports (measured vs calced)
    - 20ish students - first run online
    - <redacted> unit code
        - 2nd year, Not first sem
            - Has pre-reqs
            - Core unit for electrical engineering
            - Part elective
    - <redacted> coordinator
    - F2F (50ish students) as well - using <redacted> as well, same design
        - Replacing sim with netlab - 1, perhaps 2
        - 5 practical
            - 2 net, 3 sim
        - 1 hands-on using pre-built (probes, wires, using oscilloscopes etc.) First.
        - Pre-covid, still no F2F labs, all <redacted>
    - March 2023
- Purpose:
    - Messiness
    - Sim there to practice analysing a circuit.
    - Prebuilt circuit, implement it, measure elec params.
    - Getting a sense of how the circuit works
    - Limited experience – fleshing out
    - No room for improvement aspects targeted
    - Need to provide the opportunity for things to go wrong
    - Error-finding and handling
        - Another dimension
- Identified issues
    - 1 ohm resistor overvoltage not seen by students
    - Misconceptions and lack of surrounding knowledge
    - Learning from mistakes