The realist evaluation of educational technology

by

Melanie King

A Doctoral Thesis

Submitted in partial fulfilment of requirements

for the award of Doctor of Philosophy

of Loughborough University

Loughborough University

Department of Computer Science, 2017

The realist evaluation of educational technology

Melanie Rose Nova King B.A. (Hons)

Department of Computer Science, School of Science, Loughborough University

m.r.n.king@lboro.ac.uk

Abstract

Purpose - This thesis considers the best way to address the challenges faced by educators, institutions and funding bodies trying to not only develop and implement educational technology successfully but tackle the challenge of understanding and evidencing what works (and what does not) and why. The aim of the research was to find and validate an evaluation method that provided usable and useful evidence.

Approach - A range of evaluations were undertaken to elicit the strengths and weaknesses of different approaches, augmented by drawing upon the experiences and outcomes published by others. An analysis of the issues was made and significance of the problem established. The problem being premature timing, unsuitable models, rapid change, complex implementation chains, inconsistent terminology, ideology and marketisation. A tailored realist evaluation framework was proposed as an alternative method and it was tested to evaluate an institutional lecture capture (LC) initiative.

Findings – The theory-driven realist approach provided a level of abstraction that helped gather evidence about wider influences and theories of potential future impact of the LC programme and its linked policy. It proved valuable in generating real and practical recommendations for the institution, including what more could be done to improve uptake and support embedding in teaching and learning, from practice, policy and technological points of view. It identified some unanticipated disadvantages of LC as well determining how and when it was most effective.

Practical implications – A Realist Evaluation of Technology Initiative (RETI) framework has been produced as tool to aid the rapid adoption of the approach. Recommendations for future research and seven guiding principles have been proposed to encourage the formation of a community of realist evaluative researchers in educational technology.

Originality / value - The rigorous application of a tailored realist evaluation framework (RETI) for educational technology (including the development of two Domain Reference Models) is the primary contribution to new knowledge. This research is significant because it has potential to enable the synthesis of evaluation findings within the sector. This will enable an evidence-base of what works, for whom, in which contexts and why, ultimately benefiting policy-makers and practitioners to support better informed decision making and investment in education.

Keywords: realist evaluation, educational technology, complex implementations.

Dedication and acknowledgements

It is not the mountain we conquer but ourselves. *Edmund Hillary*

This has been hard. There is no doubt about it. Fortunately for me, stamina seems to be a trait I am blessed with. No doubt, due to the incredibly inspiring teacher I had at secondary school called Mr Dukes. He was a charismatic form tutor and a passionate believer in outdoor education. I was a regular on the Dartmoor survival expeditions and mountaineering trips that he organised, long before the days of Health & Safety and parent consent letters! He would march us across boggy moors in shocking weather. We slept rough and learned all sorts of survival tactics. I can still hear him chuckling as he rescued me, stuck waist deep in a peat bog, "Melanie, never lose your sense of humour!" It is a lesson that has stuck with me to this day. I can confess that a sense of humour has been essential while climbing my PhD mountain. It has been a truly gruelling yet epic cerebral expedition. My maternal grandmother, Marie Smeeton and my paternal granddad Reginald King have also inspired this journey. Both naval personnel, Marie was a Chief Petty Officer in the Wrens and Reginald a Chief Engineer in the Auxiliary Navy. Two of the most wise, resilient and stoic people I have ever known. They always encouraged me to 'make the most of every opportunity' and told me that I could achieve anything with grit and determination.

I dedicate this thesis to these tremendously awe-inspiring people: I reached the top of my mountain and arrived back at base camp with the rosy glow of relief, a big smile and tremendous pride. Phew!

I am so grateful of the opportunity given to me by Loughborough University to undertake a PhD by this route and waive its fee for staff. I am thankful to my wonderful supervisors Ray Dawson, Steve Rothberg and Firat Batmaz. They have pulled me out of that bog and steered me onwards and upwards, so many times over the course of writing the final few papers. I could not have done this without them! I have received such valuable advice and support along the way from Charles Oppenheim and Russell Lock, as well as colleagues in the Centre for Information Management who took me under their wing. I have also benefited from a never-ending supply of encouragement (and coffee) from some amazing colleagues, Ray Chung, Lizzie Gadd, Sue Manuel, Kay Davey and Alexandra Bowmer. All of them are superstars. I am proud and lucky enough to have worked within the super-team that was the CEDE family: Glynis Perkin and Lynda Gibbins who read and re-read many early drafts of my papers; Paul Newman for prompting so much debate, Caroline Lowery, Sarah Bamforth, Kate Everest, Glenda McMahon and Matt Mould for cheering me on and tolerating endless conversations about evaluation, even in the pub! I miss them all.

Finally, I am thankful for the support and patience of my partner Anthony, my mum and my beautiful daughter Mir. They have put up with missed shopping trips, a 'shabby chic' home and too many ready meals. Most importantly, they have helped me weather-the-storm … and never lost their sense of humour.

Table of Contents

Section A: Introduction

A 1. Aims and objectives

The aim of the research was to find and validate an evaluation method that provided usable and useful evidence of what works in the domain of educational technology. Therefore, the research had the following objectives:

1. Undertake a range of evaluations to elicit the strengths and weaknesses of different evaluation approaches. Draw upon the evaluation experiences and outcomes published by others in the field.
2. Identify and analyse the inherent issues with current evaluation approaches. Establish the extent and significance of the problem in the educational technology domain.
3. Propose an alternative evaluation approach and hypotheses.
4. Test the approach in a real-life setting.
5. Determine whether the new approach provides usable and useful evidence of what works (and does not) and why. Establish its strengths, weaknesses and limitations.
6. Convey the significance of the findings and make recommendations for the future.

A 2. Overview

A 2.1 Thesis by publication

Thesis by publication is permitted under the rules of Loughborough University[1]. This thesis is based on 11 publications. At the time of thesis submission, 10 of the publications had been published and one had been accepted for publication. The papers have been accepted by or published in journals and conference proceedings from within the disciplines of education, learning technology, systems and information technology. These publishers have different requirements; thus, the papers are not consistent with one another in terms of structure and presentation. A full list of references is provided in Annex 1.

Several of the publications included in this thesis were co-authored. Appendix 1 provides the attribution statement for each paper, summarising the contributions of each researcher. Only papers where the author took a major role in the research and preparation were included. Please note, three of the papers were published in my former married name of M. R. Bates.

A 2.2 Structure of thesis

The structure of this thesis charts a research journey that started in 2006. Section A provides the overall aims and objectives of the research, key definitions, context of the research, its positionality, paradigms and approach. Section B establishes the problem with current evaluation methods, as observed by the author within papers 1 – 7, and from the wider literature; that is, the challenges faced by educators, institutions and funding

[1] http://www.lboro.ac.uk/governance/regulations/26/current/

bodies trying to not only develop and implement educational technology successfully but tackle the challenge of understanding and evidencing what works (and what does not) and why. Section C presents an analysis that considers why there is a problem, including a synopsis of the strengths and weaknesses of traditional evaluation approaches and establishes the extent and importance of the problem (paper 8). Section D asserts the hypothesis: that is the use of a realist evaluation method, tailored specifically for technology initiatives, will overcome the barriers to effective evaluation (paper 9). Section E summarises the test of the hypothesis in a real-life setting, which is the evaluation of an institutional Lecture Capture initiative using the tailored approach (papers 10 and 11). A summary of the strengths and weaknesses are discussed, and the resultant refined Realist Evaluation of Technology Initiative (RETI) Framework is presented in Section F as the author's primary and original contribution to knowledge. Finally, the thesis concludes in Section G, by discussing the significance of the research and contribution to the domain. It also provides recommendations for future research and some guiding principles to establish a community of realist evaluative researchers in all technology domains.

A 2.3 Definitions

Educational Technology

The term educational technology is used broadly in this thesis to describe any and all technology in use in an educational context (the conditions and arrangements where teaching and learning takes place) in this case, Higher Education (HE). A useful way to conceptualise technology is offered by Lievrouw and Linvingstone's (2002) description of three distinct aspects of what technology is:

- **Artefacts and devices:** the technology itself and how it is designed and made;
- **Activities and practice:** what people do with technologies (including issues of human interaction, organising, identity, techniques and competencies);
- **Context:** social arrangements and organisational forms that surround the use of technologies (including institutions, social structures and cultures).

Realist Evaluation

Realist evaluation was originally developed in the 1990s by Pawson and Tilly (1997) to address the question 'what works, for whom, in what circumstances, and how?' in a broad range of interventions. It is a theory-based evaluation methodology and adopts the philosophy of scientific realism (Bhaskar, 2008). A full methodological review of realist evaluation is provided in paper 9 (King, Rothberg, Dawson, & Batmaz, 2016a) where an adapted realist evaluation framework is proposed and subsequently utilised within this research.

Evaluation Framework

An evaluation framework provides an overall structure to guide consistent practice and principles for evaluation activities across different programmes or different evaluations.

It can include outline definitions, guidance on process, planning activities, data collection, and evaluator conduct. An example of an evaluation framework is one developed by the New South Wales (Australia) State Government to guide consistent and transparent evaluation of government programmes (NSW Government, 2013).

A 3. Research context

A 3.1 Positionality

The activities undertaken during this investigation are set in the context of the author working as a practitioner-researcher. Ongoing inquiry has been motivated from a background in design and experiences in roles such as learning technologist, project researcher, project manager and IT manager. In the early years, the aim was to enhance teaching and learning within the curriculum using emerging technologies: working directly with academics to adapt and adopt successful technologies into their courses. Subsequent activity centred around externally funded projects and government funded initiatives such as FDTL, TQEF, Jisc, CETL (see Appendix 3 for a full list of acronyms) where there were more formal requirements from funders to demonstrate success or potential wider impact.

The pace of change began to accelerate, both within UK HE and technologically, which was juxtaposed with government funding cuts to the sector. The necessity for the HE community to adapt helped establish the importance of the inquiry along with a growing requirement for critical reflection and continuous improvement from professional accreditation bodies such as the Higher Education Academy. There was a need to read more, learn from others, understand and utilise educational theories and models on technology adoption, and undertake more rigorous and focused investigation. There was a shared desire to make fewer mistakes; to become wiser and more creative in the design, development and embedding processes; to make better decisions about areas that would benefit from technological developments and conversely areas where it was unhelpful or damaging to students, staff and institutions.

A good proportion of the past decade of activity has been spent working as part of a vanguard of like-minded researchers and practitioners trying to innovate in the field of educational technology. Innovation, not only in new technologies or pedagogies, but particularly with this research, innovation in the way that we understand what works (and what does not) and more importantly why. This research attempts to uncover evidence that is based upon more than anecdotal accounts from an inner circle of technology evangelists. The driving force is to undertake research that is understood and received as a rigorous evidence-base upon which sound strategic decisions can be made. Particularly in the context of institutions needing to invest more of their own money in technology-enhanced teaching and learning.

The overarching professional aim was (and still is) to engage in research-informed development of educational technology that is not only usable but useful and used.

A 3.2 Research paradigm and approach

Pragmatism

The research began with pragmatism as a theoretical perspective. Pragmatism is a research paradigm that involves a mixed-model approach to applied research design, concerned with action and change and the interplay between knowledge and action (R. B. Johnson & Onwuegbuzie, 2016).

The majority of this research was carried out within projects as part of a Centre for Engineering and Design Education, working alongside academics, engineers, designers and pedagogic researchers. A variety of project methodologies were employed, epistemologically chosen based on the best way to solve problems, where change was the underlying aim. Initially the collaborative development of technologies with academics took the form of action research projects. Bryman (2015, p. 688) describes action research as "an approach in which the action researcher and a client collaborate in the diagnosis of a problem and in the development of a solution based on the diagnosis."

This interdisciplinary environment also fostered design-based research thinking within projects, a methodology commonly used in the learning sciences to bridge the gap between theoretical research and practice in formal education (Anderson & Shattuck, 2012). Design-based research "being situated in a real educational context provides a sense of validity to the research and ensures that the results can be effectively used to assess, inform, and improve practice in at least one (and likely other) contexts". Anderson and Shattuck acknowledge that design-based research builds upon action research, is an iterative process, draws upon and contributes to theory, and involves a collaborative partnership between researchers and practitioners.

Interpretivism

Although the day-to-day project approaches were rooted in pragmatism, the drive to review the work (and publish results) came from an emerging interpretivist perspective. This came from a necessity for interpretation and understanding of the new technologies, solutions and educational interventions that came from these projects and their wider impact or potential. Particularly, it became more important to understand how human factors contributed to the success or failure of educational technology developments.

Evaluation was chosen as a practical method or device to help discover and understand the interventions more broadly, including the processes, the products, human factors and any unanticipated outcomes of the projects. Various evaluation approaches were tried, each using quantitative and qualitative methods for data collection and interpretation. The context of practitioner as researcher/evaluator was also important and much of the qualitative data collection adopted a phenomenological stance. That is, it was felt important to study the experiences of all people involved (not just end-users) and questions and answers were derived from the combined experiences of both researcher and participant, exploring perceptions, emotions and volitions. Both the questions generated and evidence collected was framed and themed around a growing knowledge of

learning and technology theories. There was always an implicit desire to abstract from local experiences and contribute to a wider theoretical body of knowledge.

However, a significant driver in using evaluation as a method was that the author found it to be a very useful tool in the context of the research and project activities. The overwhelming benefit was that it had two very definite and practical purposes. Firstly, it was a tool to support the design and development of products. Secondly, it helped present a picture of impact and wider potential, as the basis of future developments or funding bids.

However, as Section D outlines, it became clear that a new evaluation approach was needed. Realism, as an alternative way of thinking, then emerged.

Realism

Realism is a philosophy of science that positions itself as a model of scientific explanation between the poles of positivism (there is a single reality that can be measured) and relativism (there is no universal objective truth). Pawson and Tilley's (Pawson & Tilley, 1997) realistic evaluation (later known as realist evaluation) is the first evaluation approach that rests on realist principles: "Realism's key feature is its stress on the mechanics of explanation, and its attempt to show that the usage of such explanatory strategies can lead to a progressive body of scientific knowledge" (Pawson & Tilley, 1997, p. 55). Their strategy was first termed 'scientific realist analysis' because their method identified with Bhaskar's scientific realism and generative view of causation (Bhaskar, 2008). A realist approach assumes that nothing works everywhere for everyone and that context really makes a difference. Their method acknowledges that all interventions or 'programmes' are in fact social systems, that are introduced into more complex social systems. Therefore, inquiry focuses on 'how' something works through a change in the reasoning and responses of people involved linked to specific contexts, that bring about a set of intended or unintended outcomes of the programme.

The action of realist theorising is based on a method of thinking called retroduction (Lewis-Beck, Bryman, & Futing Liao, 2004, p. 972) (also known as abductive reasoning), a logic of inquiry found within scientific realism. This means relying on previous expertise, experiences, hunches or imagination to generate a theory that is inspired by the evidence. Therefore, the role of the researcher-practitioner as evaluator is significant in realist evaluation. Realist evaluation is an iterative process of generating a candidate theory, testing it by gathering evidence, then refining those theories in cycles. As a realist evaluator, one is in fact a theoriser using retroduction with a combination of deduction (theory tested against evidence) and induction (theory derived through evidence) with an element of inspired and creative thinking!

As the research progressed, the emphasis of each evaluation grew from micro-level factors to macro-level concerns, such as the social, organisational, political and cultural context of interventions. Therefore, realism also provided a much-needed alternative perspective to accommodate the increasingly complex socio-technical factors encountered. An

evolving world-view also emerged. This meant that the latter evaluations were conducted with somewhat of a 'critical theory hat on', particularly when theorising significant contextual factors. Macleod (2009) succinctly compares this point of view: "Positivism wears a white coat, constructivism accepts a cup of tea, and critical theory is SUSPICIOUS." The eminent scientific philosopher, Donald Campbell, brought critical theory into scientific realism (Campbell, 1988). For Campbell, critical realism was about promoting criticism and counter criticism in the community of scientists to concentrate more closely and collectively on the quality of the reasoning in research reports rather than just the quality of the data. "Organized distrust produces trustworthy reports." (Campbell, 1988, p. 303)

A 3.3 The difference between evaluation and research

Evaluation and research are synergistic but they serve two different purposes. Stufflebeam said it best: "evaluation's most important purpose is not to prove but to improve" (Stufflebeam, 2001, p. 56). The objective of evaluation is to enable change and so evaluation activities should be action orientated. Evaluation also suggests a judgement needs to be made and therefore an evaluation is conducted when there is the practical opportunity to use the results.

Evaluation activities have been classified within four types each having their own uses, as summarized below (Stufflebeam & Shinkfield, 2007).

- **Summative** evaluations are retrospective and used to provide accountability reports when a product, project or programme of work is completed. They are useful for determining success or failure – aimed predominantly at sponsors or consumers.
- **Comparative** evaluations are to assist in dissemination, for example, to share proven practices or products to help consumers make wise adoption or purchasing decisions. For example, comparisons might be made between proprietary and open technologies with similar features (Udas & Feldstein, 2006).
- **Formative** evaluations are used to provide information to develop a service, ensuring its quality, or improving a method or approach by providing continuous feedback loops for a project. This type of evaluation is carried out before or during the implementation stage and is aimed directly at the project staff.
- Evaluations to foster **enlightenment** are conducted to bring new understanding arising from revelations. Findings from these evaluations can address particular research, theory or policy questions.

In practical terms, when conducting an evaluation, it is necessary to utilise appropriate research methods (whether quantitative or qualitative) to help gather and analyse data collected within the different approaches and there is a widely-held view that research is a subset of evaluation activities. However, this thesis frames evaluation as a subset of research, as Pawson has done. Pawson describes realist evaluation as evaluative research conducted in the paradigm of scientific realism, and "wanted to emphasise that evaluation research had a different cause from other social sciences, namely to have realistic ambitions to inform real-world policy and practice." (Pawson, 2013 p.xix) An

evolution of Pawson & Tilley's early idea of cumulation (p147) is presented in Pawson's 2013 book 'The Science of Evaluation'. This realist manifesto sets out a blueprint for realist evaluation as a scientific discipline in order to move from stand-alone evaluations to a cumulative and authoritative body of knowledge. This view is supported by the author and is one of the recommendations of this thesis.

Section B: The Problem

B 1. The challenge of knowing what works: key areas

The work presented here spans a decade and, therefore, there has been ample opportunity to delve into a broad spectrum of activity across the institution and at all levels, including working with academics needing answers at a curriculum level, in discipline, institutional and sector funded initiatives, and in activities that help senior decision makers identify areas for development and investment. The technologies studied also cover a broad spectrum of types, many that were built in-house (some as part of this research), including desktop simulation software, web applications to support teaching and learning, and a repository for digital resources. Others include cheap and cheerful, off-the shelf, game-based simulations to a very expensive enterprise-level suite of 'lecture capture' technologies.

The over-arching challenge was to identify the critical factors affecting the development, implementation and use of educational technologies. The evaluations spanned activity within five key areas:

- At the **course** level: Can technology stimulate real-life learning?
- For new initiatives at **institutional** level: How do staff attitudes and current practice affect the design, implementation and perceptions of the use of new technologies?
- At the **discipline** level: What can we learn from analysing an existing model of effective practice?
- In **priority areas**: How do we decide where to make the best use of technology to improve the student experience more broadly?
- Where there is a **mixed pattern of success**: Why are some departments (or individuals) more successful at embedding educational technology into their practice than others?

B 2. The early evaluations

A total of seven evaluations (presented in papers 1 – 7, see Appendix 1 for full references and attributions) were undertaken between 2006 and 2014: two summative, one comparative and four formative. A summary of the aims and evaluation approaches used for each one is provided in this section.

B 2.1 Summative

PAPER 1
Can a virtual laboratory provide a comparable learning experience for distance learners?

A series of software virtual laboratories were developed to simulate the on-campus renewable energy laboratories for postgraduate students. The Biomass Lab was evaluated to see if it could provide a comparable laboratory experience for distance learners (Blanchard, Moron-Garcia, & Bates, 2006). A beneficiary assessment (Salmen, 2002) approach was used, to assess the value of the intervention as perceived by the intended beneficiaries (i.e. the students), thereby aiming to give a voice to their priorities and concerns. A comparison of grades achieved, with those students experiencing the real-life lab, was also made.

PAPER 7
What are the critical success factors of a discipline-based model of development?

An illustrative case study (GAO, 1990) approach was used to present evidence on the emergence and development, over 14 years, of a specialist Centre for Engineering and Design Education, as a model of success in developing engineering education (King & Willmot, 2014). Evidence assessed included funding acquired, publications produced, and learning technology embedded both within the institution and elsewhere.

B 2.2 Comparative

PAPER 5
What is the best strategy to evaluate a business simulation game that meets learning objectives but also enhances employability skills? Can we determine which game meets our needs? What are the benefits of an interdisciplinary team?

The author worked in an interdisciplinary group of experts to undertake this comparative evaluation of various business simulation games, also using the principles of utilisation-focused evaluation (Patton, 2012). This meant evaluating collaboratively with the teaching staff as the primary intended users of the results. A variety of methods were utilised, for example, a software quality assessment was made based on technical criteria derived by the author and usability testing was undertaken by a wider group, using a set of standard heuristics. A questionnaire was also used to determine the potential employability skills that would be gained by the student users. Paper 5 presents the results of this comparative evaluation (King & Newman, 2009).

B 2.3 Formative

PAPER 2
What are staff impressions of using existing repositories of teaching and learning materials? What are the main barriers and incentives in staff contributing their own materials?

PAPER 3
Can we learn from theories on technology adoption to help inform the implementation of new interventions?

PAPER 4
How would academics use the repository to collaborate with colleagues? What are the existing user processes associated with creating and sharing teaching material relating to communication, support and

technology?

The Jisc funded Rights and Rewards project carried out formative evaluation and related research activities in an attempt to establish a single blended repository to meet the teaching and research needs of Loughborough University. The results of a national survey, gathering the views of 430 people, helped to identify issues with copyright as well as the incentives and barriers for contributors when considering sharing their teaching materials via a repository (Bates, Loddington, Manuel, & Oppenheim, 2007a). It became clear that significant factors were not only the academics' attitudes to depositing and sharing their work but a desire to accommodate their existing work flow in relation to the creation and sharing of material, particularly within the Virtual Learning Environment. Interviews were conducted and staff's current practice was process mapped. This helped generate detailed user requirements for a broader repository service that comprised a complex technical architecture of distributed yet interoperable systems (King, Loddington, Manuel, & Oppenheim, 2008).

This work was followed up with a desk-based study on some potential models of technology early adoption that could be adapted for use by the project. An institutional framework for change adoption was developed, to aid in the embedding of the repository service in practice. The study also highlighted the complexity of the environment within which new innovations are situated and the challenges faced by the project (Bates, Manuel, & Oppenheim, 2007b).

PAPER 6
What are the common issues that impact negatively on a student's experience of their final year? Can a Service Design approach provide an insight into the complex interactions and relationship between students, technology and the University?

The Jisc funded Pedestal for Progression project looked at issues relating to final year progression and used formative evaluation to aid in the identification of potential enhancements (Wheeler & King, 2012). Three different evaluative approaches were taken. Firstly, Service Design was used to model the complex interactions between people and technology. "Service Design helps to innovate (create new) or improve (existing) services to make them more useful, usable, desirable for clients and efficient as well as effective for organisations." (Moritz, 2015, p. 6) For example, students were asked to recall a critical moment during their final year and produce an experience map illustrating their emotional response at certain touch-points of interactions with the university (also known as a journey map). Secondly, data mining or learning analytics techniques were used to turn raw data from multiple systems, into useful information about learners and their progression (or lack of) at various points in time for different types of learners. Finally, a variety of other methods were used to collect both primary data (interviews and focus groups) and secondary data (meeting minutes from staff-student committees, student satisfaction surveys, reports from curriculum enhancement projects) from a diagonal slice of the institution.

Section C: Analysis of the Problem

C 1. Strengths and weaknesses of the approaches taken

After five different evaluation projects, summarised in Section B, there was a need to take stock and reflect on the strengths and weaknesses of the approaches taken thus far.

The virtual Biomass Laboratory was just one in a series of virtual labs, evaluated on its ability to simulate the real-life laboratory experience for distance-learners. The results showed that it could: the distance-learners did not appear disadvantaged regarding their module marks or perceptions of their learning experience. However, though anecdotally, another virtual lab, used in a related module, did not fare so well. Although some of the original software is still in use today, the evaluation did not establish the critical factors that made the Biomass Lab successful (and others not so) and why there was no further investment made in their development. Was it purely a pocket of success or did it have wider potential? The evaluation answered the teacher's question at the course level but was not enough to establish a case for further developments, or for the institution to invest in wider adoption of this type of educational technology in other discipline areas.

There were many strengths in the approach taken in the comparative evaluation of the business simulation games. The benefit of having an interdisciplinary team was clear: diverse specialisms could contribute to an in-depth analysis of both the technology and pedagogic aspects. There was also a clear evaluation brief and the main stakeholders of the evaluation outcome were involved at every stage, that is the teachers who had to incorporate the game into their curriculum. At the course level, this type of evaluation was successful but a weakness was the omission of an analysis of students' potential response to the use of a game-based simulation. What impact would the competitive nature of the game have on their attitude and behaviours to each other and their teachers, and would this virtual environment favour particular types of learners over others?

Both of the Jisc projects used formative evaluation to support design and development activities, working at the institutional level with multiple stakeholders. Their strengths included the adoption of a variety of data collection methods and the sheer quantity of information analysed from multiple primary and secondary sources. In addition, existing theories on technology adoption underpinned the development of an institutional framework for change to aid repository adoption. However, the more data collected, the more complicated the challenge became. For the Rights and Rewards project, the attitudes of staff to an increasingly complex technology environment (including confusion over terms used and the purpose of different types of technology) was critical. It was clear that incentives would play a key role in initiating behaviour change in staff that was sustained beyond the end of the project. However, the evaluation only began to scratch the surface of the organisational and political factors inhibiting future service development, and consequently the momentum slowed and new developments fizzled out after the project ended. The same could be said for the Pedestal for Progression project,

which also revealed the complex 'teaching enhancement' environment of multiple, often competing, initiatives being trialled or implemented at any one time across the institution.

Finally, the illustrative case study gathered together a mass of evidence to demonstrate success factors of a discipline-based model of development. The strength of this was the articulation of the variety of activities and qualities (such as operations, location, innovation culture, in-house technology development, funding and research, links with national networks) that had been sustained over a prolonged period. However, most of this information was gathered together for the purposes of evidencing success retrospectively for the funders, predominantly the government funders, so 'what didn't work' was never published. Another weakness was that it did not examine which of these factors were 'critical' and why. Particularly, it did not explore fully the relationship between the Centre and the host institution. For example, were the institution's critical success factors different to that of engineering academics, the funders or the national engineering education community? Could this be the reason the Centre was closed in 2016? Did rival institutional initiatives play a factor, was off-the shelf educational technology better or were institutional costs simply too high amidst a challenging funding environment where economic benefits were not evidenced?

The challenge of doing meaningful and impactful evaluation was becoming ever more apparent. The approaches and methods used up to this point were simply not giving the right answers to help decision making above the course level. They were inadequate in addressing the complexity of the combination of contextual factors from the macro perspective.

C 2. Is this a widespread and important problem to solve?

PAPER 8
The need for innovation in the evaluation of educational technology.

A review was undertaken to establish the barriers to effective evaluation of educational technology and to see if this was a widespread problem (King, Dawson, Batmaz, & Rothberg, 2014).

Five factors were identified:

1. **Premature timing**: Summative evaluations never provide a full understanding of potential influence and impact.
2. **Unsuitable models**: Existing technology acceptance and maturity models are inappropriate for dealing with complex contextual factors.
3. **Rapid change**: HE is in a rapid state of political and corporate change often requiring evidence of efficiency savings.
4. **Complex implementation chains**: The iterative nature of agile development means what is being evaluated is always in a state of change.
5. **Inconsistent terminology**: often locally adapted or country specific is a barrier to synthesis of findings across studies.

The evaluation imperative for educational technology in HE is described in section 1.1 of Paper 9 (King et al., 2016a). In summary, financial pressures in the HE sector are driving institutional efficiency programmes requiring difficult decisions on whether to invest or cut resources for educational technology developments, particularly bespoke software development. This, coupled with the migration of technology development skills away from HE to the commercial sector, brings the challenge of providing academics with specialist support. The prominence of digital systems in all aspects of HE also makes for an increasingly complex and problematic landscape of data structures and work processes across all boundaries of operations, teaching and research (Selwyn, 2014). Concerns over the social, political and cultural roles of these systems can be seen in the dispirited accounts from interviews with academics in Australia (Hill, 2012). The use of digital technologies in general featured prominently as exemplifying the worst aspects of working in modern universities, perhaps posing a risk of losing research expertise from academia to industry too.

This reality is juxtaposed with the celebratory terms that educational technology is often described in and "the tendency for the majority of people to unthinkingly assume that educational technology to be inherently beneficial" (Selwyn, 2013, p. 9). Selwyn prompts us to understand educational technology as an ideology: "Consistent throughout this history of digital hysteria has been a belief that new technologies herald substantial educational change, renewal and [..] 'disruption'" (Selwyn, 2014, p. 9). The polarized opinions of the excessively optimistic 'boosters' against the small but significant group of 'doomsters' (Bigum & Kenway, 1998) are problematic for evaluators. These unconscious biases are perhaps present to some extent in all stakeholders of evaluations: staff, students, evaluators, management, reviewers and publishers "preferring concepts or facts one wishes to be true, rather than concepts or facts known to be true" (Spring, 2012, p. 30). This may be the reason why many evaluation sponsors can, and often do, prohibit distribution and publication of evaluation findings (Henry, 2009), making learning from others' experiences difficult.

As the marketisation of UK HE grows (Molesworth, Nixon, & Scullion, 2009) so too does the competition amongst educational providers to improve their offerings, reduce costs and carefully manage their outward reputation. Institutions are forced to consider the management of their reputation, as university image influences student satisfaction and loyalty formation process (Alves & Raposo, 2010). In this context, therefore, institutions may retain a legitimate reason to keep the outcome of unfavourable evaluations internal to the organisation.

The marketisation of higher education has also introduced the concept of student as consumer "where students seek to 'have a degree' rather than 'be learners'" with the market addressing consumer needs rather than learner needs (Molesworth et al., 2009, p. 278). "The maxim of getting good 'value for money' effectively becomes a guiding principle in how higher education's core activities are appraised" (Tomlinson, 2015, p. 452). The institutionalisation of students' contributions to evaluation of their experiences, namely the student voice, is seen to have wide ranging influence: "student evaluation of

higher education in the past was seen to be for a few interested students, the formalisation of student voice activity over the last decade has provided all students with new forms of power within the sector" (Freeman, 2016, p. 860).

Therefore, a further two factors appear as challenges for the effective evaluation of educational technology:

6. **Ideology**: evaluators are faced with the general orthodoxy of educational technology as a positive force for good.
7. **Marketisation**: Reputation management inhibits the sharing of evaluation outcomes and students' changing relationship with the university inculcates a consumerist context to evaluations of educational experiences.

The growing scholarly critique surfacing the distrust of educational technology and the need to take a more critical perspective on the use of technology in education (Bigum, Bulfin, & Johnson, 2015; Selwyn, 2013; 2016b; 2016a) is a call to arms for evaluators. There has never been a more crucial time to make sure that evaluation approaches and methods are grounded in theory, take a critical perspective and address the complex technical, political, economic and social contexts. "There is a need to be relentlessly realistic as well as occasionally optimistic about the relationship between education and technology". (Selwyn, 2016a, p. 184)

Section D: Hypothesis

D 1. A proposed alternative

PAPER 9
A realist evaluation framework refined for complex technology initiatives

In adopting realism as an alternative research paradigm, a methodological review of realist evaluation and realist synthesis was undertaken to establish its potential in addressing the evaluation problem for educational technology. Realist evaluation is a type of theory-driven evaluation to foster enlightenment (see section A 2.3) and tests the theory underpinning the initiative, in other words, the reason it is expected to work.

A realist evaluation framework was developed not only to illustrate the key steps of the method (and demonstrate its iterative nature) but to tailor it specifically to aid the evaluation of complex technology initiatives in education. The author's refinement included two domain reference models for the classification of technology and associated staff/student roles in HE (King et al., 2016a) to help map the complex technical environment and to address the problem of inconsistent terminology identified as a barrier to the synthesis of evaluation findings.

The hypothesis is that a realist approach with a tailored framework can address many of the factors identified as barriers to effective evaluation and provide real and practical recommendations for users, from individuals to institutions.

D 2. Expected results

Realist evaluation is distinct from traditional forms of technology evaluation, in that it goes beyond assessing the quality, usability, or utility of a particular digital tool (or project) at a single point in time. Its emphasis is on evaluating the social system within and around an intervention, and it is expected that this will mitigate against the narrow conventional techniques of evaluating solely the software or specific technology-based models to understand usage and adoption. The realist's response to mapping the initiative before the evaluation (with refinement during) takes into account the complexity inherent in people, organisations, rival interventions and rapid change, as well as the complex development and implementation chains of in-house educational technology development. The technique is expected is to provide evidence of what works, for whom, in which contexts and most importantly why. The realist approach to evaluation, however, does not aim to provide a full and complete picture of the initiative under investigation. It is expected to surface the significant factors that influence observed outcomes and provide evidence, in the form of a testable theory, of a 'slice of the pie' upon which future evaluations can build.

Section E: Testing the hypothesis

PAPER 10
What are the behaviours and attitudes of staff that make their students use lecture capture more?

PAPER 11
Does a realist approach (and the tailored framework) address the factors identified as barriers to effective evaluation? Can a theory-driven approach provide real and practical recommendations for organisations? Does this approach provide greater insight into the political, economic, cultural and social complexity of technology implementations in education?

An evaluation of a Lecture Capture (LC) initiative was conducted, as a test of hypothesis (King, Dawson, Rothberg, & Batmaz, 2017). LC was chosen as it is technically complex, it is a widely used and established technology and there is limited rigorous research of its impact on education. Realist evaluation was an appropriate approach in this instance because there was ad hoc usage of the system at department level, that anecdotally seemed to work for some, but real evidence of what was working (or not), for whom, in which circumstances and why still needed to be understood.

E 1. Methodology

During a preliminary mapping of the initiative (including categorising the people and technologies involved using the reference models) the evaluation questions were established and two programme theories were generated based on the underlying assumptions of why it was expected to work. Four cycles of increasingly focussed investigation then tested these initial candidate theories. These investigated:

Cycle 1. The *outcomes* relating to the *context* of department and prevalence of local champions.

Cycle 2. Lecturer *contexts* and *mechanisms* linked to positive viewing *outcomes* (a more in-depth presentation of findings from this cycle is provided in paper 11 (King, Dawson, Batmaz, & Rothberg, 2016b)).
Cycle 3. LC use and attendance *outcomes.*
Cycle 4. Student *contexts* and *mechanisms* linked to negative viewing *outcomes.*

A post-evaluation review (also known as meta-evaluation) was undertaken to help assess the overall quality of the evaluation and to determine whether the hypothesis proposed in this thesis is a valid one.

Section F: Findings

F 1. Evaluation findings

The realist approach enabled consideration of the wider strategic and organizational settings of the LC initiative. This provided a level of abstraction that helped gather evidence about wider influences and theories of potential future impact of the programme and its linked policy. It proved valuable in generating real and practical recommendations for the institution, including what more could be done to improve uptake and support embedding in teaching and learning, from practice, policy and technological points of view. It identified some unanticipated disadvantages of LC as well determining how and when it was most effective.

This approach has provided a greater insight into the complex factors at play with regards to the adoption, use and adaptation to specific educational technologies, in this case LC. The evaluation identified significant political, economic and cultural contextual factors that would warrant future research as both challenges for evaluation (see section C 2) and to provide a starting point in further realist theory generation, these being marketisation of HE and responding to the student voice. In addition, the impact of global marketing and multimillion-pound investment by corporate educational technology companies provides a significant economic contextual factor.

F 2. Strengths and weaknesses

The realist evaluation method took a while to grasp but the realist community is pro-active in support of new researchers and the author benefited from workshops provided and advice from the community's mailing list. The LC evaluation took longer and was more labour intensive than expected and this is a potential weakness of the realist evaluation approach. However, a large proportion of effort was manually extracting detailed system usage data from each captured session and then merging this with other data sources to find cohorts and contexts for further analysis. This is a fundamental weakness of the technological environment rather than with the evaluation technique. The realist theory generation (that drew upon existing educational, behavioural and technological theories) required either an in-depth knowledge of the literature or time to

review the literature. This was also the case when validating the evaluation results with the current LC literature.

Iterative refinement of the programme theory using existing theoretical concepts, evidence from existing literature and analysis of the data collected was time-consuming as a novice realist evaluator. However, ultimately this is the strength of this approach. One option during the early stages could be Rapid Realist Review, which "has been developed as a tool for applying a realist approach to a knowledge synthesis process in order to produce a product that is useful to policy makers in responding to time-sensitive and/or emerging issues, while preserving the core elements of realist methodology" (Saul, Willis, Bitz, & Best, 2013). An interdisciplinary team-based approach would also be advantageous.

One area for improvement was identified as the need to incorporate stakeholder communication and feedback at the end of each evaluation cycle. The inclusion of this step would help commissioners of the evaluation better understand the theory-driven approach and, therefore, the rationale for practical recommendations that had been derived from the evidence collected within each cycle.

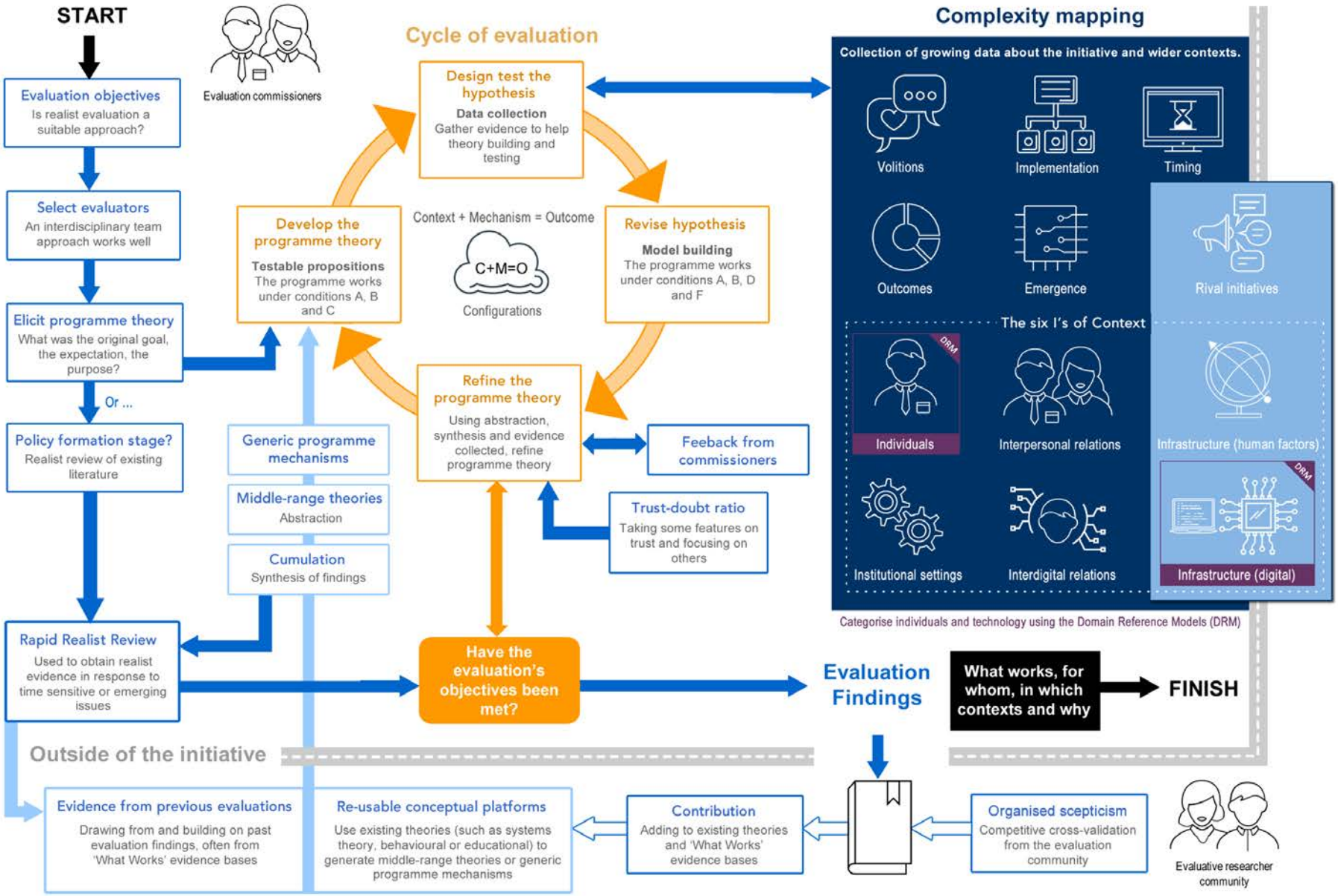
Antonenko (2014) has demonstrated the instrumental value of conceptual frameworks in educational technology research in connecting the context of practice and theory and this research not only concurs with that but its strength is that it provides a practical tool to do it. Not only was theory used for generating testable hypotheses, but the method can also contribute to theory, in this case Venkatesh et al.'s (2003) unified theory of acceptance and use of technology.

Similarly, the strength of the framework was in the categorization of the technology itself and the roles of the people involved. These reference models provide an indication of the generic capabilities of the technology and the functional roles that people inhabit in relation to it, providing the evaluator with a technological context as a starting point to gather evidence of technological outcomes. For example, is the technology providing the expected capability, is it being used by people in ways unexpected ways or is a crucial technical role missing from the initiative in a particular context?

F 3. The Realist Evaluation of Technology Initiatives (RETI) Framework

Based on the outcome of the LC evaluation, further refinements have been incorporated to create a tested Realist Evaluation of Technology Initiatives (RETI) Framework (Figure 1) - the author's significant original contribution to new knowledge.

Figure 1 The Realist Evaluation of Technology Initiatives (RETI) Framework



Icons: KOLO Design and Elizabeth Arostegui used under a Creative Commons license (Attribution 3.0 Unreported).

The author's RETI Framework (figure 1) succinctly illustrates the key steps of Pawson's realist evaluation method (2013). One begins by eliciting the programme theory, then the cycle of evaluation commences; building testable propositions, gathering the evidence and refining the theories. Within each cycle, the evaluator builds a growing map of complexity of the programme (or initiative) using Pawson's VICTORE checklist (the Volitions of people involved, the Implementation itself, the Context, Timing, Outcomes, Rival initiatives and Emergence) (Pawson, 2013, p. 33). Throughout the evaluation, generic programme mechanisms, middle-range theories, synthesis from existing literature and re-usable conceptual platforms are drawn from outside the initiative to help in theory building and testing. Pawson's trust-doubt ratio means taking some features on trust while focusing on others (Pawson, 2013, p. 86). When the evaluation has met its objectives, the findings contribute to a growing evidence base of what works, for whom, in which contexts and why. Organised scepticism is a key principle of Pawson's principles of evaluation as a science (Pawson, 2013, p. 86), which relies on close scrutiny of each other's work.

The RETI Framework incorporates the author's additional steps specifically for the realist evaluation of educational technology. Additional steps include considering Rapid Realist Review (Saul et al., 2013) as an alternative to full evaluation, often used in response to time sensitive or emerging issues. Ensuring a dialogue with evaluation commissioners, to gain feedback after each iteration, is also suggested. Most significantly, the inclusion of Domain Reference Models to categorise people and technology, to aid in the synthesis of findings. The author has also proposed two more additional digital-based contexts as part of the complexity mapping phase, explained in detail in section G 1.1 as another contribution to new knowledge.

The biggest strength of the framework is that it has been rigourously tested in the evaluation of a complex educational technology initiative and appears to address the problems identified with previous evaluation methods, therefore validating the hypothesis: realist evaluation can provide real and practical recommendations while addressing the complexity of the combination of contextual factors from the macro perspective.

Section G: Discussion, recommendations and conclusion

G 1. Discussion

G 1.1 Embracing contextual complexity

The author's research has found that the way realist evaluation embraces complexity addresses many of the barriers for effective evaluation identified earlier. However, the author proposes that the most significant of the checklist elements is the mapping of 'context', which is fundamentally missed in many evaluations, including the early evaluations presented in this thesis (Section B 2.). After adopting realist evaluation, it is now apparent that documenting and understanding context within a snapshot in time is vital. Interventions are never implemented or embedded in the same way twice and,

therefore, without knowing the conditions within which the evaluation can be considered valid, it would be difficult to know where and how to apply the knowledge gained elsewhere.

Table 1 builds upon and adapts Pawson's four I's (I-IV) (Pawson, 2013, p. 37), which describe the layers of the concept of 'context', and adds two more linked to technology (V – VI). Context I (Individuals) is adapted to use the author's Domain Reference Model for people (King et al., 2016a). The first additional context is to acknowledge the digital infrastructure of system architectures and information architectures as well as the wider digital context (V). The second is the context of interdigital relations, which is the relationships that people have with the technologies and the digital environment provided by (or the wider context of) the initiative (VI). These additional contexts have been added to the RETI Framework (figure 1).

Table 1. The six I's listing the contextual layers of technology-based initiatives

I	*Individuals*	The characteristics and capacities of the various stakeholders ascertainable using the domain reference model (King et al., 2016a).
II	*Interpersonal relations*	The stakeholder relationships that carry the initiative.
III	*Institutional settings*	The rules, norms and customs local to the initiative.
IV	*Infrastructure (human factors)*	The wider social, economic and cultural setting of the initiative.
V	*Infrastructure (digital)*	The local technological context and the wider digital setting of the initiative.
VI	*Interdigital relations*	The relationships that stakeholders have with the digital environment provided by (or in use within) the initiative.

The addition of these two digital layers is significant for two reasons. The first is the sheer change in the technological environment that has advanced so rapidly, even since the beginning of this research journey. For example, elements of the digital context might be the maturity of the technology on Gartner's Hype Cycle and how that influences the questions asked and methods used (O'Leary, 2008). It could also include an assessment of the maturity of the institution and its readiness for the technological initiative, or its capability to deliver rapid technical change, for example, by mapping its agile maturity (Benefield, 2010). The following three examples demonstrate what constitutes the context of 'Interdigital relations'. Firstly, people's own digital ideologies provide a context for the mechanisms and outcomes observed. Secondly, individuals relate to technologies in ways that augment with themselves either virtually or physically, for example visualisation and skills development in complex medical training (Kamphuis, Barsom, Schijven, & Christoph, 2014). Finally, interdigital relations could also describe relationships that stakeholders have with robots and machines (or artificial intelligence in software) that provide a human-like role in the initiative. For example, students at Deakin University now ask their support questions to Watson (IBM's cognitive computing technology platform) which gives them a personal response. It is continually learning,

adapting and therefore changing. Inevitably the stakeholders' relationship with it will change over time too as it becomes more supportive. (Deakin University, 2015).

The digital environment therefore constitutes its own 'system of systems' as a significant layer of technological complexity. As such it would be beneficial to draw upon systems concepts such as tipping points and non-linear effects in relation to these digitally adaptive environments during the mapping of 'Emergence' (the E in VICTORE complexity checklist), which is in fact another systems concept.

It appears that the dominant theories heralded in the drive to link theory to practice are educational ones (Gunn & Steel, 2017; Hannon & Al-Mahmood, 2014). However, this thesis argues the importance of 'technology' in 'educational technology' evaluation. Mapping the complexity of digital system of systems contexts alongside interdigital relations should be an essential component of any evaluation. It is also important to draw upon technological theories, for example systems theory, actor-network theory, or technological determinism.

G 1.2 Evaluation as a fundamental component of new initiatives

It is recommended that evaluation (using the RETI framework) should be a fundamental component considered at the start of any new initiative. Early consideration will prompt the organisation and policy makers to think about the theory underpinning the initiative: why is it expected to work? This establishes a clear brief to help design the intervention explicitly for evaluation. If it is a complex implementation chain, it would be worthwhile to articulate a Theory of Change too (Taplin, Clark, Collins, & Colby, 2013), as a comprehensive description of how and why a desired change is expected to happen in stages leading to its desired goal. This provides a structure for iterative formative evaluation, taking stock of the unfolding implementation and checking the outcomes at each stage against the theory. Of course, this applies to agile and iterative technology development and pilot technology implementations too.

The RETI framework was originally developed to evaluate an existing initiative. However, it could also be used to evaluate the policy itself: testing the assumption before any implementation had taken place. Rapid Realist Review (Saul et al., 2013) could be used for knowledge synthesis from other evaluations and even the activities associated with policy formation could also investigated. For example, perhaps the policy was created before a real understanding of the underlying problem was explored fully. Selwyn and Facer (2013) believe that, "a critical study of educational technology necessarily begins with a critical reflection upon the definition of the educational 'problem' at hand". Therefore, one of the significant 'contexts' in an evaluation might be the process of policy formation: what were the breadth, depth and approaches used to understand the problem before a policy and potential solution was defined?

G 2. Recommendations

G 2.1 Next steps and recommendations for future research

The following recommendations and guiding principles (section G 2.2) are outlined to demonstrate the future direction (and wider potential) of this research.

Researchers should:

- Use and test the RETI framework (figure 1) in other settings to validate the results of this thesis and to refine the domain reference models.
- If feasible carry out evaluations within an interdisciplinary team (using a participatory approach where possible) to mitigate against the weaknesses identified (Section F 2.).
- Contribute to mapping the complexity of educational technology initiatives by further investigating significant 'contexts', such as ideology and marketisation, significant 'volitions' such as the role of academic attitudes and behaviours, and rival initiatives.
- Use mixed-methods for data collection appropriate to the context or mechanism under investigation and reflect on their suitability for realist evaluation.
- Draw upon not only learning theories as conceptual platforms, but from behavioural, organisational and technological ones too. For example, theories on software and information systems (requirements engineering, development processes, implementation and adoption).
- Seek out and publish candidate realist theories on any emergent effects of the initiative that future researchers can investigate.
- Critique methods and results of others, "the objectivity of physical science does not come from turning over the running of experiments to people who could not care less about the outcome, [..] it comes from a social process that can be called competitive cross-validation." (Campbell, 1988, p. 324).
- Adopt the proposed guiding principles to help establish a community of evaluative researchers in educational technology.
- Contribute to the establishment of a community 'what works' evidence base to inform better policy and practice.

G 2.2 Guiding principles: towards a community of evaluative researchers

These guiding principles (table 2) are rooted in realism, underpinned by the organizing principles of evaluation science, as set out in Pawson's realist manifesto 'The Science of Evaluation' (2013, p. 86). The author has contributed principles I – IV (and borrowed V-VII from Pawson), however, the author has summarised what all of these principles mean for the educational technology community, both in terms of the actions required and the rationale behind them.

Table 2. Seven guiding principles for the educational technology evaluation community

	Principle	*Community Action*	*Rationale*
I	Realist evaluation and realist synthesis (using the RETI Framework).	Adopt realism as an approach for all evaluation requirements. Ideas for new initiatives, process evaluation (formative), at the end of implementation (summative) and when trying to understand what has gone on before (synthesis).	All evaluations will contribute evidence to inform a wider understanding of what works, for whom, in which contexts and why. Thereby creating a robust evidence base for future policy and funding decisions.
II	A community of expert evaluators	Develop the skills, knowledge and expertise of the realist educational technology researcher. Build a Community of Inquiry around the application of the framework.	Evaluators need to be recognised as expert practitioners that play an important role in designing the realist evaluation, building theories and models, and drawing upon conceptual platforms from education and software development.
III	A common language	Create, maintain, utilise and refine a shared terminology (Domain Reference Model) to describe the digital environment and related roles within higher education.	By using a consistent terminology, findings from disparate evaluations can be more easily synthesised.
IV	Contribute to the map of complexity	Map the complexity landscape of every initiative using the realist approach and communicate this in findings.	If every researcher mapped different elements of the complex terrain of technology and organisations, it will provide important context clues for future researchers to build upon.
V	Abstraction	Provide more data using evidence collected within evaluations to refine existing middle range theories and re-usable conceptual platforms.	Collecting evidence about educational technology use and adoption in a realist format will help to contribute to general theories or models about technology adoption, for example.
VI	Cumulation	Draw upon and build on evaluations that have been carried out before, even in a different policy domain. Publish findings using realist standards (Wong et al., 2016).	No evaluation should start from scratch; there have been evaluations of the role of incentives for example in different programme domains.
VII	Organised skepticism	Be constructively critical of realist evaluations to maintain the standards of realist research and learn from each other.	No evaluation can fully understand the complexity of the initiative. Evaluations may have lost focus in their method or presentation of their findings. It is up to the community to support each other in the adoption of the method and be skeptical of findings.

G 2.3 Wider research potential

The RETI Framework has been born out of research of technology in educational settings and the Domain Reference Models were specifically developed for this context. However, the use of the models is not core to the use of the framework, which can be applied in any

setting. Consequently, there is potential to develop it further with different domains contributing their own reference models for the technologies and roles people play (as well as core conceptual theories linked to particular domains). It has the potential to become an Evaluation Architecture Standard, much like the Open Group's TOGAF® Enterprise Architecture standard (Keller, 2012). This would provide organisations with principles and practices for developing and using evaluation as an integral component of organisational development.

G 3. Conclusion

The aim of this research was to find and validate an evaluation method that provided usable and useful evidence of what works in the domain of educational technology. A range of evaluations were undertaken at course, discipline and institutional level, to elicit the strengths and weaknesses of different approaches. Seven barriers were identified by analysing the inherent issues with current approaches, drawing upon the evaluation experiences of others. These barriers are: premature timing, unsuitable models, rapid change, complex implementation chains, inconsistent terminology, ideology and marketisation. The extent and significance of the problem in the educational technology domain was established. A methodological review of realist evaluation was undertaken and a tailored framework developed to test the hypothesis that this would overcome the barriers to effective evaluation. The framework was used successfully to evaluate an institutional Lecture Capture initiative. The research found that this approach provides usable and useful evidence of what works (and does not) and why. Its strengths and weaknesses are discussed and, based on this, an enhanced Realist Evaluation for Technology Initiatives (RETI) Framework has been produced as the significant output of the research (figure 1). Recommendations for future researchers and guiding principles for the community have been proposed.

The identification of the barriers to effective evaluation and the rigorous application of a tailored realist evaluation framework (including the development of two domain reference models) are the main contributions to new knowledge. This research is significant because it has potential to enable the synthesis of evaluation findings within the sector. To enable an evidence-base of what works, for whom, in which contexts and why. Ultimately, this brings benefit to policy-makers and practitioners by supporting better decision making on investments in education.

References

Alves, H., & Raposo, M. (2010). The influence of university image on student behaviour. *International Journal of Educational Management*, *24*(1), 73–85. http://doi.org/10.1108/09513541011013060

Anderson, T., & Shattuck, J. (2012). Design-Based Research. *Educational Researcher*, *41*(1), 16–25. http://doi.org/10.3102/0013189X11428813

Antonenko, P. D. (2014). The instrumental value of conceptual frameworks in educational technology research. *Educational Technology Research and Development*, *63*(1), 53–71. http://doi.org/10.1007/s11423-014-9363-4

Bates, M., Loddington, S., Manuel, S., & Oppenheim, C. (2007a). Attitudes to the rights and rewards for author contributions to repositories for teaching and learning. *Research in Learning Technology*, *15*(1), 67–82. http://doi.org/10.1080/09687760600837066

Bates, M., Manuel, S., & Oppenheim, C. (2007b). Models of Early Adoption of ICT Innovations in Higher Education. *Ariadne*, (50).

Benefield, R. (2010). Seven dimensions of agile maturity in the global enterprise: A case study (pp. 1–7). Presented at the Proceedings of the Annual Hawaii International Conference on System Sciences, IEEE. http://doi.org/10.1109/HICSS.2010.337

Bhaskar, R. (2008). A Realist Theory of Science (2nd ed.). Routledge.

Bigum, C., & Kenway, J. (1998). New Information Technologies and the Ambiguous Future of Schooling — Some Possible Scenarios. In *International Handbook of Educational Change* (pp. 375–395). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-011-4944-0_19

Bigum, C., Bulfin, S., & Johnson, N. F. (2015). Critical Is Something Others (Don't) Do: Mapping the Imaginative of Educational Technology. In *Critical Perspectives on Technology and Education* (1st ed., Vol. 43, pp. 1–13). New York: Palgrave Macmillan US. http://doi.org/10.1057/9781137385451_1

Blanchard, R. E., Moron-Garcia, S. D., & Bates, M. (2006). Converting the physical to the virtual: providing a laboratory experience for distance learners in engineering. In A. Méndez-Vilas, A. Solano Martin, J. Mesa González, & J. A. Mesa González (Eds.), *Current Developments in Technology-Assisted Education* (Vol. 2, pp. 1208–1213). Badajoz, Spain: British Computer Society.

Bryman, A. (2015). Social Research Methods (5 ed.). Oxford University Press.

Campbell, D. T. (1988). Methodology and Epistemology for Social Sciences. (E. Overman, Ed.). University of Chicago Press.

Deakin University. (2015, November 25). IBM Watson helps Deakin drive the digital frontier. Retrieved August 8, 2017, from http://www.deakin.edu.au/about-deakin/media-releases/articles/ibm-watson-helps-deakin-drive-the-digital-frontier

Freeman, R. (2016). Is student voice necessarily empowering? Problematising student voice as a form of higher education governance. *Higher Education Research & Development*, *35*(4), 859–862. http://doi.org/10.1080/07294360.2016.1172764

GAO. (1990). Case study evaluations. US General Accounting Office.

Gunn, C., & Steel, C. (2017). Linking theory to practice in learning technology research. *Research in Learning Technology*, *20*(2), 16148. http://doi.org/10.3402/rlt.v20i0.16148

Hannon, J., & Al-Mahmood, R. (2014). The place of theory in educational technology research. *Rhetoric and Reality: Critical …*.

Henry, G. T. (2009). When Getting It Right Matters: The Case for High-Quality Policy and Program Impact Evaluations. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 32–50). Thousand Oaks, CA:

SAGE Publications, Inc. http://doi.org/10.4135/9781412995634.d9

Hill, R. (2012). Whackademia. UNSW Press. http://doi.org/10.2307/j.ctt5vkpq8

Johnson, R. B., & Onwuegbuzie, A. J. (2016). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26. http://doi.org/10.3102/0013189X033007014

Kamphuis, C., Barsom, E., Schijven, M., & Christoph, N. (2014). Augmented reality in medical education? *Perspectives on Medical Education*, *3*(4), 300–311. http://doi.org/10.1007/s40037-013-0107-7

Keller, W. (2012). TOGAF 9.1 Quick start guide for IT enterprise architects. *Version 09 A*.

King, M. R. N., Dawson, R. J., Rothberg, S. J., & Batmaz, F. (2017). Utilizing a realist evaluative research approach to investigate complex technology implementations: an e-learning lecture capture exemplar. *Journal of Systems and Information Technology*, *19*(1/2), 22–41. http://doi.org/10.1108/JSIT-04-2017-0027

King, M. R. N., Rothberg, S. J., Dawson, R. J., & Batmaz, F. (2016a). Bridging the edtech evidence gap: A realist evaluation framework refined for complex technology initiatives. *Journal of Systems and Information Technology*, *18*(1), 18–40. http://doi.org/10.1108/JSIT-06-2015-0059

King, M., & Newman, R. (2009). Evaluating business simulation software: approach, tools and pedagogy. *On the Horizon*, *17*(4), 368–377. http://doi.org/10.1108/10748120910998399

King, M., & Willmot, P. (2014). A blueprint for success: a model for developing engineering education in the UK. *International Journal of Engineering Pedagogy (iJEP)*, *4*(2), 18–22. http://doi.org/0.1109/ICL.2013.6644583

King, M., Dawson, R., Batmaz, F., & Rothberg, S. (2014). The need for evidence innovation in educational technology evaluation. In J. Uhomoibi (Ed.), *Global Issues in IT Education* (pp. 9–23). British Computer Society.

King, M., Dawson, R., Batmaz, F., & Rothberg, S. (2016b). What are the behaviours and attitudes of staff that make their students use lecture capture more? In *ICERI2016 Proceedings* (Vol. 1, pp. 2350–2357). IATED. http://doi.org/10.21125/iceri.2016.1514

King, M., Loddington, S., Manuel, S., & Oppenheim, C. (2008). Analysis of academic attitudes and existing processes to inform the design of teaching and learning material repositories. *Active Learning in Higher Education*, *9*(2), 103–121. http://doi.org/10.1177/1469787408090838

Lewis-Beck, M., Bryman, A., & Futing Liao, T. (2004). The SAGE Encyclopedia of Social Science Research Methods. (M. Lewis-Beck, A. Bryman, & T. F. Liao, Eds.) (Vol. 3). Thousand Oaks, CA: SAGE. http://doi.org/10.4135/9781412950589.n865

Lievrouw, L. A., & Livingstone, S. (2002). Handbook of New Media: Social Shaping and Consequences of ICTs. 1 Oliver's Yard, 55 City Road London EC1Y 1SP : SAGE Publications, Ltd. http://doi.org/10.4135/9781446206904

Macleod, A. (2009, April 21). Paradigms in research; or, how your worldview shapes your methodology. Retrieved June 11, 2017, from https://mackle.wordpress.com/2009/04/21/paradigms-in-research-or-how-your-worldview-shapes-your-methodology/

Molesworth, M., Nixon, E., & Scullion, R. (2009). Having, being and higher education: The marketisation of the university and the transformation of the student into consumer. *Teaching in Higher Education*, *14*(3), 277–287. http://doi.org/10.1080/13562510902898841

Moritz, S. (2015). Service Design. Retrieved from http://www.lulu.com/shop/stefan-moritz/service-design/paperback/product-22151053.html

NSW Government. (2013). NSW Government Evaluation Framework. NSW Government. Retrieved from http://arp.nsw.gov.au/sites/default/files/NSW_Government_Evaluation_Framework_Aug_2013.p

df

O'Leary, D. E. (2008). Gartner's hype cycle and information system research issues. *International Journal of Accounting Information Systems*, *9*(4), 240–252. http://doi.org/10.1016/j.accinf.2008.09.001

Patton, M. Q. (2012). Essentials of Utilization-Focused Evaluation. SAGE.

Pawson, R. (2013). The Science of Evaluation: A Realist Manifesto. London : SAGE Publications Ltd. http://doi.org/10.4135/9781473913820

Pawson, R., & Tilley, N. (1997). Realistic Evaluation. London: Sage.

Salmen, L. F. (2002). *Beneficiary assessment: An approach described*. *Social Development Paper*.

Saul, J. E., Willis, C. D., Bitz, J., & Best, A. (2013). A time-responsive tool for informing policy making: rapid realist review. *Implementation Science*, *8*(1), 35. http://doi.org/10.1186/1748-5908-8-103

Selwyn, N. (2013). Distrusting educational technology: Critical questions for changing times. *Distrusting Educational Technology: Critical Questions for Changing Times* (pp. 1–192). Routledge. http://doi.org/10.4324/9781315886350

Selwyn, N. (2014). Digital technology and the contemporary university. *Digital Technology and the Contemporary University: Degrees of Digitization* (pp. 1–156). Routledge. http://doi.org/10.4324/9781315768656

Selwyn, N. (2016a). Education and Technology (2nd ed.). Bloomsbury Publishing.

Selwyn, N. (2016b). Is Technology Good for Education? John Wiley & Sons.

Spring, J. (2012). Education Networks. Routledge.

Stufflebeam, D. L. (2001). Evaluation Models. Jossey-Bass.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). Evaluation Theory, Models, and Applications. Jossey-Bass.

Taplin, D. H., Clark, H., Collins, I., & Colby, D. C. (2013). Theory of Change Technical Papers: A series of papers to support development of Theories of Change based on practice in the field. New York: ActKnowledge.

Tomlinson, M. (2015). Student perceptions of themselves as "consumers" of higher education. *British Journal of Sociology of Education*, *38*(4), 450–467. http://doi.org/10.1080/01425692.2015.1113856

Udas, K., & Feldstein, M. (2006). *Apples to Apples: Guidelines for Comparative Evaluation of Proprietary and Open Educational Technology Systems*. London: Observatory on Borderless Higher Education.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. http://doi.org/10.2307/30036540

Wheeler, A., & King, M. (2012). Exploring the balance between automation and human intervention in improving final year university student non-completion. In *INTED Proceedings* (6 ed., pp. 2358–2363). Valencia, Spain.

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J., & Greenhalgh, T. (2016). RAMESES II reporting standards for realist evaluations. *BMC Medicine*, *14*(1), 96. http://doi.org/10.1186/s12916-016-0643-1

Appendices

Appendix 1 – List of papers and attributions

The following 11 papers (8 journal articles and 3 conference papers) have received a total of 113 citations a breakdown of which can be seen on Google Scholar[2]. Please note, that Papers 1- 3 were published in my former married name of Melanie Bates.

Paper 1

Blanchard, R. E., Moron-Garcia, S. D., & Bates, M. (2006). Converting the physical to the virtual: providing a laboratory experience for distance learners in engineering. Current Developments in Technology-Assisted Education, 2, 1208–1213.

This conference paper was an outcome of an action-research project. Blanchard was the academic lead and Moron-Garcia was a co-researcher on the project. Bates' contribution to the research was the development of the simulation software, based on translation of requirements from the real-life laboratory. Subsequent developments were made after usability testing and focus groups with both staff and students. An analysis of student marks was undertaken by Blanchard and Moron-Garcia. The paper was jointly prepared, Bates' contribution being 'Section 2 Converting the course' and Section 3 – Designing the labs'.

Paper 2

Bates, M., Loddington, S., Manuel, S., & Oppenheim, C. (2007). Attitudes to the rights and rewards for author contributions to repositories for teaching and learning. Research in Learning Technology, 15(1), 67–82.

Paper 3

Bates, M., Manuel, S., & Oppenheim, C. (2007). Models of early adoption of ICT innovations in higher education. Ariadne.

Paper 4

King, M. R. N., Loddington, S., Manuel, S., & Oppenheim, C. (2008). Analysis of academic attitudes and existing processes to inform the design of teaching and learning material repositories. Active Learning in Higher Education, 9(2), 103–121.

[2] https://scholar.google.co.uk/citations?user=K1oALAkAAAAJ&hl=en

Papers 2, 3 and 4 are publications generated from Jisc funded 'Rights and Rewards' research and development project, which aimed to establish a single blended repository to meet the teaching and research needs of Loughborough University. The project was led by the Department of Information Science with the Engineering Centre for Excellence in Teaching and Learning (engCETL) and the Library as partners. Oppenheim was the Principal Investigator, while Manuel and Loddington were research assistants who focused on copyright issues. King was the technical lead on the development of the repository and undertook the associated research investigating incentives for adoption and tactics for embedding in practice.

Paper 2 reports on a national survey that was developed and piloted by King. Analysis of the quantitative data was carried out by King who also contributed to the thematic analysis of free text responses in conjunction with Manuel and Loddington. The statistical analysis was done by the Library and Information Statistics Unit (LISU) at Loughborough University. The paper was jointly prepared by King, Manuel and Loddington under the supervision of Oppenheim.

Paper 3 presents on the findings of the desk-based background research jointly undertaken by Bates and Manuel. The research resulted in the co-production of an institutional framework for change adoption by Bates and Manuel. The conclusion section was drawn from both the investigation and the experiences of the project and was written under the supervision of Oppenheim. The paper was jointly prepared by Bates and Manuel.

Paper 4 presents the results of a series of research activities including; a desk-based review of repository work-flows and business processes, ten staff interviews, the development of a proposed repository service architecture, and four proposed scenarios to aid future design and development. Loddington and Manuel undertook the literature review of repository workflows. King proposed the user-centred design approach and designed the interview questions. Loddington and Manuel carried out the interviews and wrote up the transcripts. King, Loddington and Manual shared the thematic analysis of the interview transcripts. King generated scenario one and four with Loddington and Manuel generating two and three. A proposed repository service architecture (figure 2) was developed by King. Conclusions and recommendations were generated by all authors and the paper was jointly prepared by King, Loddington and Manuel under the supervision of Oppenheim.

Paper 5

King, M., & Newman, R. (2009). Evaluating business simulation software: approach, tools and pedagogy. On the Horizon, 17(4), 368–37.

This journal article reports on the outcome of an interdisciplinary action-based research project. The authors King and Newman worked within a project team that included Professor Rob Thring (the academic in charge of the module and who proposed the project) and Rob Dover (a business consultant from Accenture who co-taught on the module). King

lead on the research aspects of the project including proposing the model to evaluate the software for teaching and learning. King and Newman jointly undertook the scoping of potential simulations to evaluate. Participants were recruited by King and the evaluation structure (e.g. usability criteria and reflective questions) was developed by King for all pilot participants to use. The technical evaluation was undertaken solely by King, while Newman generated the skills questionnaire element of the evaluation. All project team members contributed to the analysis of the results and therefore the choice of simulation to use within the module. Recommendations for enhancements were generated jointly by the project team: Thring and Dover contributed the pedagogic recommendations and King the technical ones. King wrote the article, with Newman contributing to the text on skills development, the figures and final edits.

Paper 6

Wheeler, A., & King, M. (2012). Exploring the balance between automation and human intervention in improving final year university student non-completion. In INTED Proceedings (6 ed., pp. 2358–2363). Valencia, Spain.

This conference paper examines the research methods used in the Jisc funded 'Pedestal for Progression' project. King was the Principal Investigator for the project and Wheeler the research associate. The project's research plan and methods were chosen by King. Most of the student data were collected by Wheeler through student workshops and interviews, however King collated data from student reflective journals and analysed these texts. The paper was jointly prepared.

Paper 7

King, M. R. N., & Willmot, P. (2014). A blueprint for success: a model for developing engineering education in the UK. International Journal of Engineering Pedagogy (iJEP), 4(2), 18-22.

This paper presents a case-study developed and written by King. From garnering the views of academic colleagues, the twelve-point blueprint for success in the conclusion section was created by King and Willmot. This journal paper was adapted from a conference paper that was written by King, edited and presented by Willmot.

Paper 8

King, M., Dawson, R., Batmaz, F., & Rothberg, S. (2014). The need for evidence innovation in educational technology evaluation. In J. Uhomoibi (Ed.), Global Issues in IT Education (pp. 9–23). British Computer Society.

Paper 9

King, M., Rothberg, S. J., Dawson, R. J., & Batmaz, F. (2016). Bridging the edtech evidence gap: A realist evaluation framework refined for complex technology initiatives. Journal of Systems and Information Technology, 18(1), 18–40.

Papers 8 and 9 are the result of a focused investigation undertaken by King, comprising desk-based research, under the supervision of Dawson, Rothberg and Batmaz. Paper 8 is a conference paper that presents the initial findings of the research. Paper 9 is a journal paper that provides a methodological review of Realist Evaluation (RE) and a tailored RE framework, devised by King. Both papers were prepared by King with input from Dawson, Rothberg and Batmaz. Ideas presented in early drafts of Paper 9 also received input from Dr Gill Westhorp, outlined in the acknowledgements section.

Paper 10

King, M., Dawson, R., Batmaz, F., & Rothberg, S. (2016). What are the behaviours and attitudes of staff that make their students use lecture capture more? In ICERI2016 Proceedings (Vol. 1, pp. 2350–2357). IATED.

Paper 11

King, M. R. N., Dawson, R. J., Rothberg, S. J., & Batmaz, F. (2017). Utilizing a realist evaluative research approach to investigate complex technology implementations: an e-learning lecture capture exemplar. Journal of Systems and Information Technology, 19(1/2), 22-41. http://doi.org/10.1108/JSIT-04-2017-0027

Papers 10 and 11 present the findings from an in-depth evaluation undertaken by King under the supervision of Dawson, Rothberg and Batmaz. Paper 10 is a conference paper that presents the findings of one component of the evaluation (the staff survey). This survey was devised, deployed and analysed by King. The paper was prepared by King with input from Dawson, Batmaz and Rothberg.

Paper 11 is a journal article presenting the research strategy, approach, findings and recommendations. King undertook the day-to-day research activities except for the derivation of some system usage data, the student interviews, student survey and statistical analysis of attendance information. Acknowledgements for these additional contributions are made in the paper. Early drafts of the paper were prepared by King with input from Dawson, Rothberg and Batmaz, who also made significant contributions to its re-structure, after comments were received from journal reviewers. Dawson, Rothberg and Batmaz contributed to the conclusion section and subsequent edits and amendments to the final draft.

Appendix 2 - Acronyms and abbreviations

ACTL	The Advisory Committee on Teaching and Learning (EDUCAUSE)
APIs	Application Program Interfaces
BBC	British Broadcasting Company
BCW	Behaviour Change Wheel
BUFVC	British Universities Film and Video Council
CAL	Computer Aided Learning
CARES	Centre for Advancement in Realist Evaluation and Synthesis
CDROM	Compact Disc Read Only Memory
CEDE	Centre for Engineering and Design Education (Loughborough University)
CETL	Centres for Excellence in Teaching and Learning
CMM	Capability Maturity Model
CMO	Context Mechanism Outcome
CMOC	Context Mechanism Outcome Configuration
COM-B	Capability Opportunity Motivation Behavioural model
CRM	Customer Relationship Management
DRM	Digital Rights Management
edtech	Educational Technology
EEC	Engineering Education Centre (Loughborough University)
EER	Engineering Education Research
EEVL	Internet Guide to Engineering, Maths and Computing
eMM	E-Learning Maturity Model
engCETL	Engineering Centre for Excellence in Teaching and Learning (Loughborough University)
EngTLSC	Engineering Teaching and Learning Support Centre (Loughborough University)
EPSRC	Engineering and Physical Sciences Research Council
FDTL	Fund for the Development of Teaching and Learning
FE	Further Education
FERL	Further Education Resources for Learning
HE	Higher Education
HEA	Higher Education Academy
HEFCE	Higher Education Funding Council for England
HEI	Higher Education Institution
HESA	Higher Education Statistics Agency
HPC	High Performance Computing
ICT	Information and Communication Technology
IMechE	The Institution of Mechanical Engineers
IMS	IMS Global Learning Consortium, Inc
IP	Intellectual Property
IPPR	Institute for Public Policy Research
IPR	Intellectual Property Rights
IT	Information Technology
JISC	Joint Information Systems Committee
JISCMAIL	JISC's Mailing list system
JORUM	JISC's Online Repository for Learning & Teaching Materials
LC	Lecture Capture
LCP	Library Catalogue Plus
LISU	The Library and Information Statistics Unit (Loughborough University)
LO	Learning Object
LORS	Loughborough Online Reading List System
LSS	Lean Six Sigma
MERLOT	Multimedia Educational Resource for Learning and Online Teaching
MIT	Massachusetts Institute of Technology

MOOC	Massive Open Online Course
MSc	Master of Science
NSS	National Student Survey
OA	Open Access
OAI	Open Archives Initiative
RAEng	Royal Academy of Engineering
RAeS	Royal Aeronautical Society
RAMESES	Realist and Meta-narrative Evidence Synthesis: Evolving Standards
RERO	Release Early Release Often
REST	Renewable Energy Systems Technology
RETI	Realist Evaluation of Technology Initiatives
RLMS	Reading List Management System
SCRAN	Scottish Cultural Resources Across the Network
SENDA	Special Educational Needs and Disability Act
SESR	Student Engagement Success and Retention
SESR-MM	Student Engagement Success and Retention Maturity Model
TAM	Technology Acceptance Model
TBIE	Theory Based Impact Evaluation
TDF	Theoretical Domains Framework
TEL	Technology Enhanced Learning
TLTP	Teaching and Learning Technology Programme
TOGAF	The Open Group Architecture Framework
TQEF	Teaching Quality Enhancement Fund
TRILT	Television and Radio Index for Learning and Teaching
UCISA	Universities and Colleges Information Systems Association
UK	United Kingdom
US	United States
UTAUT	Unified Theory of Acceptance and Use of Technology
VICTORE	Volitions, Implementation, Contexts, Time, Outcomes, Rivalry, Emergence
VLE	Virtual Learning Environment