The realities of evaluating educational technology in school settings

MEGAN VENN-WYCHERLEY*, Open Lab, Newcastle University, UK
AHMED KHARRUFA, Open Lab, Newcastle University, UK
SUSAN LECHELT, University of Edinburgh, UK
REBECCA NICHOLSON, Digital Learning Lab, Northumbria University, UK
KATE HOWLAND, University of Sussex, UK
ABRAR ALMJALLY, Al Imam Mohammad Ibn Saud Islamic University/University of Sussex, Saudi Arabia/UK
ANTHONY TRORY, University of Sussex, UK
VIDYA SARANGAPANI, Open Lab, Newcastle University, UK

HCI researchers are increasingly interested in the evaluation of educational technologies in context, yet acknowledge that challenges remain regarding the logistical, material and methodological constraints of this approach to research [18, 53].

Through the analysis of the authors' contributed thematic research vignettes, the following paper exposes the practical realities of evaluating educational technologies in school settings. This includes insights into the planning stages of evaluation, the relationship between the researcher and the school environment, and the impact of the school context on the data collection process.

We conclude by providing an orientation for the design of HCI educational technology research undertaken in school contexts, providing guidance such as considering the role of modular research design, clarifying goals and expectations with school partners, and reporting researcher positionality.

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**; **HCI design and evaluation methods**; • **Applied computing** → **Education**.

Additional Key Words and Phrases: Educational Technology, Classroom Evaluation

ACM Reference Format:
Megan Venn-Wycherley, Ahmed Kharrufa, Susan Lechelt, Rebecca Nicholson, Kate Howland, Abrar Almjally, Anthony Trory, and Vidya Sarangapani. 2018. The realities of evaluating educational technology in school settings. *J. ACM* 37, 4, Article 111 (August 2018), 33 pages. https://doi.org/XXXXXXX.XXXXXXX

*The first five authors of this work all contributed equally to the development of this paper

Authors' addresses: Megan Venn-Wycherley, m.venn-wycherley@newcastle.ac.uk, Open Lab, Newcastle University, Newcastle Upon Tyne, UK; Ahmed Kharrufa, ahmed.kharrufa@newcastle.ac.uk, Open Lab, Newcastle University, Newcastle Upon Tyne, UK; Susan Lechelt, University of Edinburgh, Edinburgh, UK, susan.lechelt@ed.ac.uk; Rebecca Nicholson, rebecca.nicholson@northumbria.ac.uk, Digital Learning Lab, Northumbria University, Newcastle Upon Tyne, UK; Kate Howland, University of Sussex, Falmer, Brighton, UK, k.l.howland@sussex.ac.uk; Abrar Almjally, Al Imam Mohammad Ibn Saud Islamic University/University of Sussex, Riyadh/Falmer, Brighton, Saudi Arabia/UK, Aamjally@imamu.edu.sa/A.Almjally@sussex.ac.uk; Anthony Trory, University of Sussex, Falmer, Brighton, UK, a.trory@sussex.ac.uk; Vidya Sarangapani, Open Lab, Newcastle University, Newcastle Upon Tyne, UK, vidya.Sarangapani@ncl.ac.uk.

Manuscript submitted to ACM

1 INTRODUCTION

Over the past decade, interest in educational technology that supports the creation, use, and management of appropriate technological processes and resources to facilitate learning and improve performance [50] has grown exponentially [49, 71, 80]. As a result, this has revolutionised traditional approaches to learning and improved access and support for diverse learners [18]. Overall, the growth of the educational technology sector has seen a marked focus on the importance of high-quality educational technology research, including the need to balance internal rigour and external relevance through research design [71, 80].

Despite the continued research highlighting its potential benefits, the continued use and retention of educational technology in practice has not kept pace [39]. Where HCI and educational technology are concerned, there has been a history of calls for more evaluative research to be undertaken to address this disconnect.

Attempting to define evaluation is complex and is highly dependent on the perspectives of those involved [45]. In educational technology, evaluation can focus on the impact of technology on determining educational effects such as learning outcomes, or user satisfaction, motivation, experience, and usability [57].

Approaches to the evaluation of novel educational technologies have ranged from randomised controlled trials (RCTs) to quasi-experimental, to mixed methods studies [57], with many popular methods seeking to isolate the benefits of learning when using educational technologies, and as such often aim to isolate the causal factors that impact upon learning outcomes [57]. When conducted in a classroom context this presents an extra layer of methodological and logistical constraints [18, 53, 93]. The research in-the-wild paradigm has gained prominence [78], providing researchers with the tools and guidance to undertake evaluative studies in situ to understand how users adapt and appropriate technologies in context.

However, data gathered in situ are often not conducive to 'clean' pre-post tests when compared to lab-based study contexts [29, 53], generating "messy" data outside the scope of typical research models that are often considered inappropriate for reporting and reflection despite their ability to encourage rich, new understandings in response to a situated challenge[24]. While there are benefits to collecting this rich data, it can present challenges for evaluating the impact of the technology intervention in a field that often values experimental studies for clean data, rigorousness, and generalisability [89]. The focus on evaluation in context has been seen as a particular priority for HCI researchers working at the intersection of learning and education, given their desire to understand the interactions between humans and educational technology [49]. This can be partly attributed to the fact that the practicalities of school logistics can often remain unconsidered when designing evaluations of educational technologies in context [65, 71], despite the fact that these constraints often influence the perceptions of teacher confidence, adoption, and utility of the proposed technology. These considerations might include aspects of physical space in the classroom, the relevance of the school curriculum, availability of time, teachers' energy to support engagements, and cultural constraints of the school [20, 33, 67]. The reality is that these classroom-based evaluations of educational technology cannot always practically, or even ethically, be perfectly controlled experiments.

Furthermore, until recently it was considered taboo to share negative results, positioning such research as unpublishable [24, 28]. Although we have seen a shift towards classroom-based evaluation methodologies [19, 76], choosing methods, refining study designs, and finally reporting on these evaluations is complex. As such, data from classroom studies is often not reported, or can be seen as less rigorous despite their ability to provide a myriad of data that can inform wider design considerations, including the potential to better understand the lack of integration in the classroom.

The reporting of research with unexpected or messy outcomes merits attention, enabling future researchers to avoid common pitfalls and challenges by encouraging new study designs and perspectives.

Recognition of the complex constraints of the classroom context has become more evident in the reflective practice of both researchers and practitioners, as a way of sharing their experiences to further advance the field [14]. This is reflected in HCI more generally, as reflexive methods have been used more frequently by researchers in recent years to open conversations regarding the implications of HCI at the intersection of multiple disciplines [7, 58], the development of methods [60] and the development of researcher practice [2]. However, when examining reflexivity in educational technology research, these often take place in Higher Education settings [88], outside of a formal classroom environment [90] or is seen as a skill to be supported by the technology [52]. Therefore, there is value in exploring the learning of the educational technology community from an HCI perspective, considering how we synthesise knowledge from education, educational technology and school-based research to outline guidance on evaluating educational technologies in school settings within an HCI framing.

HCI, although a multidisciplinary field, has always ultimately concerned itself with interactions and intersections [43, 77]. Where HCI intersects with Educational Technology, it concerns itself with the design and development of interactive technologies that support the pursuit of learning goals [68]. This focus is reflected in the recent relational turn in the wider HCI field [23], one which is concerned with exploring phenomenologically situated ways of understanding the 'rich, complex, and messy situations at hand around them' [43]. In doing so, understanding the realities of educational technology use in the classroom becomes a central concern for HCI researchers, one which requires us to revel in the messy realities of classroom life.

In the following article, we adopt reflexivity as an angle to explore the complexities of conducting educational technology research in school settings. Through reflecting on our own work including the commonly experienced pitfalls and the complexities we seek to surface the practical realities of evaluating novel educational technologies in the school environment. In doing so, we look to encourage other HCI educational technology researchers to reflect on the messiness of conducting research in the classroom and provide points of reflection and practical recommendations for those looking to conduct educational technology research within the field of HCI.

To arrive at these recommendations, the authors of this paper each contribute a range of vignettes and reflections on our research carried out in schools, across a variety of age ranges and countries, and with a range of methodologies for classroom-based evaluations of educational technologies. Reflecting upon these experiences and drawing upon practices from the fields of educational technology, education and HCI research, we provide interdisciplinary guidance for HCI researchers to support the methodological design of their research whilst recognising the realities and challenges of school-based evaluations of educational technologies, towards the betterment of the evaluation and adoption of educational technologies in school settings.

Therefore, our aims and contributions are threefold: (i) we discuss in detail the challenges of conducting classroom-based evaluations of novel educational technologies, offering insights into the day-to-day logistical challenges that can arise during HCI-based research studies of educational technologies in the classroom, (ii) we present detailed reflections on the practical realities of designing and conducting classroom-based evaluations from the perspective of HCI researchers, and finally (iii) we offer a set of recommendations based on these reflections, aiming to serve as a practical guide for the HCI research communities initiating classroom-based evaluations of educational technologies.

2 BACKGROUND

There is a long history of educational technology implementation in the classroom, aiming to provide their audience with improved opportunities to learn and develop [41]. Methodologies for their design, development and evaluation have historically drawn from the pool of educational research, and more recently from Human-Computer Interaction and Learning Sciences. Throughout these fields of research, "evaluation in context" has grown in popularity, situated as a method to help resolve the uncertainties of new educational technologies or technology-based pedagogies [31] by examining them in their intended learning environment.

In HCI, there is an increasing interest in understanding what it means to work within these real learning environments and share findings that can support their adoption, sustainability, and scalability in situ. This is evidenced by the recent increase in dedicated special interest groups [68, 93], workshops [18, 53] and systematic reviews [14, 57].

In the following paper, we aim to motivate a reorientation of HCI researchers engagement in the evaluation of educational technologies, reflecting upon some of the common challenges inherent to the reality of classroom-based evaluations of technologies. To frame this research, we first consider current approaches to the design of educational technology research, including its strengths, limitations, and developments. We then discuss the impact of the research approach on the related to conducting classroom-based evaluations of educational technologies.

2.1 Methodologies for Educational Technology research

In this section, we explore the strengths and limitations of educational technology research design, as well as the latest developments in this field. By examining the design of educational technology research, we provide an understanding of the current state of the field and the opportunities and challenges that lie ahead.

Strategies for evaluation are often based on assumptions that fail to identify the true impact of an educational technology on learners beyond measuring learning outcomes [14, 57]. This is despite the fact there is growing interest in measuring 21st-century skills [29] that are difficult to quantify in traditional empirical research [53]. Diversifying our approach to the innovation and evaluation of educational technologies will be vital in promoting long-term learning in areas such as collaboration, critical and reflective thinking, creativity, and media literacy [53, 54].

A central tool in educational technology research is the randomised-control trial (RCT). RCTs are often described as the gold standard for evaluating impact [42], using randomisation to balance groups and allow attribution of any differences in outcome to the study intervention. Proponents of randomised experiments in educational research may maintain that no other type of study design can tell us what would have been the outcomes for a group of learners had they not received the particular intervention [89]. However, critics note that the outcomes of RCTs can become more ambiguous depending on the characteristics of a school, such as access to materials, teacher engagement, or school culture [14, 41]. Furthermore, there are concerns about the ethical use of RCTs as a research methodology, in which only certain groups are provided with access to the intervention to provide comparisons [62]. Contemporary research demonstrates how RCTs are increasingly used alongside qualitative and quantitative data from ‘implementation and process evaluations’ to better demonstrate the relationship between the intervention and the processes and mechanisms underpinning demonstrable impact [48].

This call for complementing quantitative experimental data about nationwide interventions with qualitative data is echoed by Lowther et al. [61], who report on the use of a mixed-methods study to examine the impact of a statewide technology coaching program for teachers where the quasi-experimental method was implemented. It was found that while student assessment and achievement data showed mixed gains in high-stake testing, the classroom observations

helped reveal changes in students' behaviour that explained the quantitative data, such as more frequent engagement in research, project-based learning, and use of technology.

Qualitative approaches to the evaluation of educational technology seek to explore the relationships between participant attitudes, experiences, and perceptions in relation to their use of technology for teaching and learning [90]. However, qualitative approaches to educational evaluation are sometimes not as widely used due to a perceived lack of 'scientific rigour' regarding the subjective nature of interpretation of data, along with concerns about researcher bias and a lack of generalisability to the broader context [3, 32, 70, 86].

Through the development of wider HCI methodology, focusing on lived experience, contextual inquiries and pragmatic cultural-historical roots [12], so too has the toolkit of educational technology evaluation undergone a change. The popularity of the "in-the-wild" paradigm has grown rapidly, acknowledging the range of interdependent factors at play when engaging with technology for teaching and learning purposes, shifting the research away from isolating specific effects and toward providing a more situated and holistic understanding of a given technology's effect on interaction in a particular context [76]. The case study approach can offer researchers the opportunity to explore the implementation of educational technology in a particular setting, providing the opportunity for nuanced insight into the processes, practices, and challenges associated with implementing new technology in schools [38, 86].

While each of these methodologies has begun to offer unique, contextually-driven insights into the interplay of user and technology in education settings, they are not without critique. Qualitative studies in context are acknowledged to be difficult to control, meaning that it can be challenging to isolate specific effects of the technology [13, 38]. These paradigms can also render participants vulnerable to misinterpretation and representation by the researcher [90].

Therefore, while the evaluation of educational technologies in context can begin to help resolve the uncertainties around the introduction of new technologies or associated pedagogies [31], further work is needed to prepare educational technology researchers in the planning of research in context, supporting their reflection on power and positionality when conducting and reporting upon their research [41], how the daily reality of the school environment will influence the outcome of research and begin to understand what this means for the transferability of findings across educational contexts and progression of educational technology research [38].

2.2 Challenges facing evaluation of Educational Technologies

Public trust in educational innovation is low, due to perceptions regarding its applicability in real classroom settings, and this is no different for educational technology innovation [14, 38].

Evaluation can be defined as the process by which people assign value and worth, and in educational technology can be used to understand the application of educational technologies to understand their role in supporting learners [66]. The introduction of educational technologies in school settings can often begin with great promise but are later rejected due to their incompatibility with the reality of their context (e.g. [27]). Evaluation in context is seen as a method to address the negative perception of the research and provide systems and resources that were created with consideration of users, their needs, and the environment of deployment. Calls for educational technology evaluation in context have been seen across many subfields such as Computer Supported Collaborative Learning [51], Learning Analytics [74, 84] and HCI [55, 64], but many do not explore or refer to the challenges that arise in day-to-day classroom environments [92] and instead focus on measurements of learning [46].

Within the realm of the evaluation of educational technology, we are well underway in the shift to conducting research in context to engender the compatibility of educational technologies in their context of use. Increasingly, this means that researchers are leaving controlled lab environments to conduct research in the natural and messy context of

delivery [14, 19, 57, 76], working alongside people to understand the realities of experience and values in day-to-day classroom environments. An implication of such an approach is the complexities of designing and conducting research in situ is the ability for research to support, change or disrupt existing practices with consideration of the material and social circumstances of the research [76, 87]. However, there are methodological concerns regarding researcher power, reflexivity and the physical and socio-cultural constraints imposed created by the school environment [38, 41, 63, 88, 90].

To gain an understanding of how these challenges influence research, we turn to Lai and Bower's [57] literature review regarding the evaluation of educational technology research. Of the 365 papers reviewed, the majority of the papers focused on evaluating learning, perceptions, and behaviours. Only 11.8% focused on perceptions of the technology, such as perceived usefulness, ease of use and adoption, while 11% examined the pedagogical practices and strategies to support the use of the educational technology. Up to 1.4% explored institutional capacity, policy and support to credibly integrate the evaluated technologies. These figures can give us a critical insight into the role that the challenges of evaluating educational technology present, and how the multifaceted and complex context of evaluation can shape the role of research produced.

Further challenges facing educational technology evaluation include the dichotomy of methodological approach, exacerbated by the challenging nature of the school environment, technological development, and researcher positionality. Some researchers believe that "quick and dirty" evaluations are better placed to respond to the rapid development of educational technology, providing quicker insights in line with a field undergoing rapid development [41]. Others believe that studying educational technology in context can be done using long-term case studies that examine an entire programme, addressing the potential for novelty effects and ensuring that observable phenomena are stable, replicable, and observable [38, 55]. While long-term studies can be better situated to expose the complexities of an educational environment and lead to outcomes that may better support the long-term use of educational technologies and meet the needs of stakeholders in school settings, these benefits must be balanced alongside the realities of researcher availability and the culture of academic research engagements.

Furthermore, methodological designs must consider the position and positionality of researchers. Researchers must take into account the experience of teachers and students on the design of educational technology evaluations, including understanding how technology impacts the teaching and learning process [21, 47], the impact that being evaluated can have on attitudes and perceptions in context, and what this means for the representation of these findings through publication [90]. An improved researcher focus on the practical realities of educational delivery, such as an understanding of educational policies and pedagogical strategies can provide a basis for stakeholder integration in a project, addressing the limited focus on perceptions regarding the adoptability of educational technology research [14]. For example, in Bond et al's review of five decades of educational technology research, they note that current limitations of approaches to educational technology research that do not consider the relevance of relationships, experience and institutional context in the research process [14, 57]. In response, researchers are recommended to explore the development of reflective, mututally beneficial relationships with school-based beneficiaries to provide insight into the influence of relationships, experience and institutional context on the process of developing educational technologies. Current work encourages inviting practitioners into the process of evaluation to address the relevance of the research to the environment, encouraging effective adoption and application [41]. However, this approach can introduce further challenges in balancing the needs of multiple stakeholders (whether researchers, developers, teachers, managers, or policymakers) who all may want to identify and measure the educational potential of such technologies and gauge their degree of success or failure [54]. In this paper, we look to expand upon these concepts in relation to the evaluation

of educational technologies from the perspective of HCI research and contribute insights into the logistical challenges and practical realities of designing and conducting school-based evaluations of educational technologies.

Further challenges are introduced when looking purely at the evaluation stage in the lifecycle of educational technology development research. Changes in educational policy and availability of innovative technologies have meant that the challenges surrounding the evaluation of educational technologies are constantly evolving [29, 57, 68]. Additionally, the interdisciplinarity between HCI, educational technologies and education presents particular challenges for the methodological approaches to the evaluation of educational technologies, making evaluation particularly challenging for researchers, and integration particularly challenging for educators and learners [53].

Methodological decisions behind the evaluation of educational technologies in the classroom setting is a complex and multifaceted task, requiring careful consideration of the technology, the experiences of stakeholders, the role of researchers, and the culture and values inherent to school settings. A lack of consideration for these contextual social and cultural factors must be paid during the methodological design to promote sustainability and scalability [14, 38, 41], or result in a failure to reproduce results beyond the lab or limited school environment.

The pragmatic and practical considerations of conducting educational technology evaluation in the classroom environment, regardless of the instruments used, are less reported in the available literature. Therefore, this paper provides an overview of HCI researchers' experience with evaluating educational technologies in a classroom environment, highlighted through a series of focused vignettes. After surfacing challenges, we then explore potential strategies for future educational technology research, synthesising guidance for future educational technology research regarding the design of classroom-based educational technology evaluations.

3 METHOD

The following section outlines the methodological approach for this paper, including the roots of its conception, approach to data collection and analysis, and overview of the limitation of the reflexive approach adopted.

3.1 Research approach

This paper takes a reflexive approach to the consideration of educational technology design and evaluation within classroom settings, an approach that has gained popularity within HCI and educational technology research for its focus on continuous exploration of how we, as researchers, design and implement research while remaining cognisant of our own motivations and biases. e.g. [7, 58, 88, 90]. Stemming from qualitative research approaches, reflexivity can be considered a confessional account of methodology in which the relationship between researcher, the researched and the socio-cultural makeup of the research space is co-constituted [36]. It is well placed to support the exploration of position, perspective and presence of the researcher when engaged in situated research, and engages in the evaluation of the research process, methods and outcomes [37].

While there must be care taken to adopt rigorous methods for the evaluation of educational technologies in situ, researchers are also integral to this research environment. Reflexivity, as a developmental approach to research methodology, can provide the tools for researchers to adapt to the environments in which they operate [6, 36]. In HCI, there is growing traction for reflexivity as a method for design that is cognisant of ethical harms [40] and engaging in participatory plurality for design [72]. In education and educational technology circles, reflexivity is increasingly evident in its role in the evaluation process [5, 88]. By reflecting upon the relationship between researcher and research, drawing upon the bodies of knowledge across HCI, Educational Technology and educational research, and sharing

these experiences, we can begin to expand our understanding of educational technology evaluation in its environment from the perspective of HCI researchers.

3.2 Research design

This paper began following a workshop at Interaction Design and Children in 2020, focusing on the evaluation of emerging technologies in the classroom [53]. Through structured discussions, the authors shared challenges they had faced when evaluating technologies in school settings, and began to realise that there were common themes of logistical, material, ethical and social challenges that we had encountered in our work. While there was extensive literature on the evaluation of educational technologies, these were largely focused on the Higher Education context, did not explore the experiences of the evaluation stage of the research process, or did not consider HCI methodologies and approaches.

Following this workshop, the authors - all HCI-based researchers involved in the design of educational technology and its evaluation - met to discuss the methodological, epistemological, and practical challenges of design and evaluation in classroom settings. The meetings were a space to discuss and reflect upon our individual and collective experiences when involved in the evaluation of educational technologies, particularly focusing on challenging experiences that had occurred during the research process and whether there were unreported findings that we had deemed to be too "messy" or beyond the scope of our research projects, to be shared in academic research with the HCI community.

Following a number of these meetings, authors all individually reflected on their own experiences in relation to the challenges inherent in the design and evaluation of educational technologies 'in the wild'. Each author was then asked to write vignettes that highlighted the biggest challenges that they had faced while conducting their own research. In a qualitative paradigm, vignettes can allow authors to define a situation encountered and provide the space for the interpretation and exploration of actions in context. Within our research, these brief reflexive narratives could reflect our internal dialogue, and surface themes and points of discussion, illuminating the diverse, disparate and connected challenges of evaluating educational technologies in context.

On returning to the meeting with these vignettes, authors shared them and spent time exploring the stories and research behind the 'problems'. This prompted several authors to write new vignettes or to revisit theirs to ensure that their vignettes reflected the true 'messiness' of conducting research in a classroom setting. For this reason, we chose not to include further vignettes from external research in our analysis, as we could not adequately engage in the reflective reshaping process with their authors.

3.3 Data analysis

Once we had a series of vignettes from each author, between meetings we read each other's vignettes and began a process of reflexive thematic analysis [16]. Within thematic analysis, an inductive approach allows for the generation of identified thematic areas without attempting to fit them into an existing framework [15], where no previous frameworks need to be identified for analysis, and ideas can be generated freely before being refined. This approach was chosen as we sought to understand the potential patterns that were beginning to be visible between our shared experiences. Despite research that spanned a wide range of cities, educational stages, and methods it was clear that there were shared challenges that we had all experienced. Reflexive thematic analysis allowed us to examine these patterns across a range of data including experiences, practices, perceptions, concepts and social processes [16, 56]. Using the collaborative whiteboard software Miro, we recorded our individual vignettes on challenges ranging from logistical issues such as planning studies around school timetables, to distractions that impacted students' attention during our studies.

The final stage was a process of data immersion, reading, reflecting, and discussing our experiences to prompt joint development of thematic areas for further discussion [16]. Through joint discussions of our experiences, vignettes, and positionality, we began to develop core themes that encapsulated our commonly experienced challenges when conducting evaluations in classroom settings. These form the data for the paper, together with further reflection on the possible opportunities and challenges of carrying out educational technology research in the classroom context. In the sections that follow, we first present our vignettes grouped into themes in the findings detailing the challenges we faced while conducting research in classroom contexts. We then go on in the discussion to present further detail from our reflections explored during the meetings and, drawing upon our collective insights, offer an orientation for researchers who wish to undertake similar research.

3.4 Limitations

Following the guidance of a reflexive methodology, we would be remiss to not discuss the methodological limitations of this paper and its impact on the findings contained within.

Firstly, the adoption of a reflexive approach introduces inherent limitations. One prevailing criticism of reflexivity pertains to the inherent subjectivity of this method, which can result in "muddy" and ambiguous reporting [36, 75]. Furthermore, detractors argue that the focus of research becomes centered on the researcher and their internal cognitive processes rather than the subject under investigation [73]. However, Probst notes that reflexivity can benefit the epistemological rigour of research by ensuring positionality, subjectivity and reactivity become more apparent [75]. In light of this, we have incorporated individual statements of positionality from all contributing authors, allowing readers to engage not only with our collective perspectives and conceptualisations of challenges but also with the presented findings and recommendations.

Another notable limitation of this study is rooted in our deliberate decision to confine the vignettes exclusively to our personal experiences. While this helps us to interrogate and explore these occurrences in greater depth, this choice unavoidably excludes the analysis of niche experiences, distinct challenges, and corresponding recommendations that may exist beyond the scope of this paper.

Further to this point, we cannot claim to be a representative sample of the HCI and Educational Technology communities, nor would this be a feasible single paper to author. Nonetheless, we firmly believe that this paper possesses the potential to invigorate the community to reflect on their "messy" research experiences and document these reflections, contributing to the body of knowledge regarding the evaluation of educational technologies in school settings from the perspective of HCI.

In conclusion, while the reflexive methodology employed in this study offers valuable insights, it is essential to acknowledge its limitations in terms of subjectivity, scope, and representation. These limitations shape our findings and underscore a deeper need for continued reflexivity and exploration within the domain of embedded, Educational Technology research in HCI.

4 ACKNOWLEDGEMENT OF POSITIONALITY

The following section acknowledges the positionality for each contributing author, including an overview of our areas of research, approach to research and location. All but one author (AA - Saudi Arabia) are based in the UK, and thus our collective experiences are predominantly rooted in the UK educational context.

MVW: MVW works in action research, typically embedding herself within educational communities to co-develop computing curricula in the UK. Focusing on long-term engagements, she employs a range of qualitative data collection

and analysis methods to understand technology use to support curricular development in teaching and learning environments.

AK: AK has carried out research around the development, deployment and evaluation of novel educational technologies in the classroom context exploring the development of unquantifiable aspects such as higher-level thinking and collaboration skills. As such, he mainly adopts qualitative or mixed research methods with longitudinal studies carried out in close collaboration with the educators.

SL: SL has expertise in exploratory design research and research through design. Her work involves designing tangible interfaces to support collaborative and embodied learning. She typically utilizes observation techniques and video analysis to provide descriptive accounts of how new technologies in classrooms contribute to how children interact with learning content and with each other.

RN: RN carries out work exploring the use of digital technologies in performing arts classrooms. She takes a critical realist approach in order to understand the underlying causes that affect teachers' experiences of using technology in the classroom. This research typically draws on participatory action research methods which are used in embedded longitudinal projects that seek to design and evaluate technology 'in the wild'.

KH: KH carries out design-based research on novel technologies for learning, including mixed-methods evaluations in the classroom that investigate the impact on learning, motivation and attitude. She has expertise in computing education and has also designed and evaluated systems for creative writing, reading and language acquisition.

AA: AA carries out computing education research, including mixed-methods evaluations in schools that investigate the impact on learning, attitude, and enjoyment. Her work focuses on exploring the use of embodied approaches such as tangible manipulation and gestures in supporting children's learning. She is based in Saudi Arabia.

AT: AT conducts research on the effects of conceptual representation and embodiment within learning technologies. He has applied mixed methods through design-based research processes that culminate in quantitative evaluations within school settings.

VS: VS is an interdisciplinary researcher working at the intersection of human-computer interaction and educational technologies. Her work is inspired by participatory action research methods, cross-cultural interactions and inclusive digital applications. Her work seeks to establish the notion of educational brokerage through collaborations with diverse stakeholders by creatively integrating digital technologies to impact service delivery mechanisms in schools.

5 FINDINGS

The following section outlines some of the common pitfalls experienced by the authors across their research experiences, identifying some of the previously undiscussed elements of published or "unpublishable" research. We begin with Table 1 summarising the study context, including pupil age, country the study took place in and the authors involved. We also include an overview of our methodological choices and the purpose of the research. We then present the themes developed through our analysis, with each theme supported by vignettes arising from the presented study contexts.

Context	***Pupil age***	***Country***	***Authors Involved***
A: Interactive tabletops in the classroom	*12-13*	*UK*	*AK*

This was a 6-week study with two mixed-ability Year 8 classes (aged 12-13 years) with an average of 24 students in each session in a local high school. The aim was to explore design implications, potential and challenges of deploying multiple tabletops (7 in this study) in a realistic setting (i.e., in school with sessions carried out by teachers using teacher-prepared, curriculum-based tasks). The researchers worked with 5 teachers in the subjects of History, Geography, and English over eleven 60-minute sessions in total. Data was collected through in-class observations, analysis of recordings from three camera angles, semi-structured interviews and unstructured discussions with the teachers and groups of students at the end of the study. The strategic aims of the study from the school's perspective (developing students' higher-level thinking skills such as reflection and metacognition, collaboration skill, and collaborative writing skills) were set out in discussions with the school's headteacher rather than the teachers themselves.			
B: TUIs and GUIs for primary programming	*6-7*	*Saudi Arabia*	*AA, KH*
This study compared the use of a tangible user interface (TUI) and a graphical user interface (GUI) for learning to code with primary school pupils. 42 children aged 6-7 took part in the study, which was conducted at a primary school in Saudi Arabia. The study examined how interface type (TUI or GUI) affected children's learning outcomes, attitudes towards computing, enjoyment, and spontaneous use of gesture. Learning sessions were video recorded and videos subsequently coded for participant and researcher gestures. Pupils completed a pre-test and attitudinal survey; a learning session involving pairs of children programming a physical robot; a post-test and attitudinal survey, and an enjoyment survey.			
C: Classroom Orchestration in the Performing arts	*14-15*	*UK*	RN
This study explored teachers' experiences using digital technology in a secondary performing arts classroom. It was an embedded, longitudinal research project which ran over eighteen months and used co-teaching as a method of participatory action research. Field notes were kept by the researcher throughout the process and were analysed alongside interviews with the teacher to understand their experiences of and practices with digital technology. The researcher and the teacher involved in the study co-taught lessons three times a week which in practice meant planning and delivering lessons and assessing student progress together. The primary focus of the project was the teacher, although student progress was also measured throughout.			
D: Physical Computing in the Classroom	*12-13*	*UK*	*MVW*
A secondary school in the North East of England contacted the university about an opportunity to collaborate on the development of a Key Stage 3 computing curriculum for Year 8 pupils (12-13 years old). They highlighted challenges with pupil engagement and the transition from block to text-based programming and wanted to introduce physical computing devices known as the BBC micro:bit. The project lasted a full academic year and the researcher was present for each timetabled computing class per week.			
E: Concreteness Fading for Computing Concepts	*9-10*	*UK*	*AT, KH*
This study evaluated four variations of a low-fidelity prototype learning environment that aims to teach primary school pupils about internet routing using a concreteness fading approach. 59 children aged 9-10 took part in the study, which was conducted at a primary school in England. Four groups were compared: abstract, concrete, concreteness fading, and concreteness introduction, with each differing only in the conceptual representation used in the prototype with which they interacted. All participants progressed through three learning stages using the prototypes. Pupils completed pre- and post-tests and pre- and post- attitudinal surveys.			

F: Teaching Critical Thinking About Sensors and Sensing	*9-11*	*UK*	*SL*
The study involved one-off sessions in a number of schools in England, with children 9-11 years of age, to understand whether a tangible interface together with a set of designed learning activities would support the schoolchildren in engaging with critical thinking about technologies. The session was planned to be three hours long and to be replicated in all six of the schools.			
G: Content Creation Tools and Cross-Cultural Learning	*11-13*	*UK and India*	*VS*
This project involved a research activity with 2 schools in England and one school in India. The vignette discussed in this paper specifically relates to research activity conducted with the school in India. This was a private school and the study was carried out with 30 Year 8 students (ages 11-13) for a period of 5 weeks. The research was carried out by a UK-based PhD student who is originally from India. In the planning phase, the researcher coordinated with the headteacher and the class teacher to identify a suitable topic for the study (which focused on aspects of content-creation, cross-cultural learning, critical-peer feedback and deep learning) and other logistics required to conduct the research activity. Students were given the topic 'Culture of England' and had to use Project Based Learning approaches to develop a project and generate a physical artefact. Simultaneously students also had to use a content-creation tool, to develop a digital artefact to externalise their learning. The researcher was present in the classroom through the length of the activity and worked with students through all sessions.			
H: Analysing Indicators of Collaboration with a Physical Computing Toolkit	*9-11*	*UK*	*SL*
A series of studies were run in 5 primary school classrooms in England with children 9-11 years of age. In all of the sessions, the students were asked to use computers to write code for a bespoke microcontroller device with sensors and actuators, and the aim of these studies was to evaluate whether and how the device would support collaborative learning. The interface was designed to encourage collaborative work amongst children and students. The research focused both on observing physical/gestural interactions between children, as well as analysing their dialogue with their peers/others. The studies took place in each school's computer rooms, which had a variety of configurations. For example, in some, desktop computers were arranged by the wall, with all the children facing the wall; in others, desktop computers were arranged in rows, with the children facing the teacher but sitting next to each other. In another classroom still, the children used laptops and sat at tables of four, facing each other.			

Next, we summarise our thematic groupings of commonly experienced pitfalls in classroom-based evaluations of educational technologies in HCI research. We present a core vignette under each sub-theme, together with the study context, with a goal to sensitise other researchers to the challenges that often occur when working in school settings and what implications they may have on classroom-based evaluations. We then return to these challenges and their impact on the evaluation of educational technology in the classroom in our discussion.

5.1 Planning for Classroom Research

A key challenge experienced by all of the authors has been working to ensure that expectations for research studies are aligned between schools and researchers, particularly in the planning process. While not a novel suggestion, the novelty is presented in the examination of these experiences and acts as a springboard for practical methods and design choices future researchers can adapt into their own work [41] In our experiences, challenges with aligning expectations can be

subdivided into three sub-themes: 1) understanding school and teacher values, 2) understanding timetable pressures and implications on study outcomes and 3) anticipating student selection.

5.1.1 *Understanding school and teacher values.*

Study Context A: Interactive tabletops in the classroom. During a 6-week study with mixed-ability Year 8 classes and multiple tabletops in a school setting, carried out by AK and colleagues, it became obvious that the research team, teachers and school management had their own goals for participating in the research. Where researchers were focused on technology evaluation goals, the school's strategic goals centred upon developing students' higher-level thinking, collaboration and collaborative writing skills. However, these goals were set by the headteacher rather than the teachers themselves. During the study and the post-study interviews with the five teachers who took part, it was clear that encouraging reflective and metacognitive thinking and collaborative work, while being the reasons the headteacher wanted the school to take part in the study, were not part of the school culture. Most teachers had little expectation of the students, focused mainly on teacher-led information delivery and considered this to be how things are and should be at that school. This significantly affected the teachers' engagement with the study. The study was conducted with two-mixed ability classes and an observation reported in the published work states that "*many students lacked the motivation to learn or to engage with the tasks at the tabletops. A number of students did not listen to the teachers and showed behaviour issues.*"

The teachers' low expectations of the students may have also contributed negatively. In our interviews with the teachers, T1 and T2 commented on the lack of resilience and perseverance among the students. T1 also commented that their students are "not used to properly thinking for themselves" and that they were not used to "proper collaborative work". The lesson to here is to ensure that the teachers and the management both have the same goals and have a disposition to teaching and learning that aligns with those of the researchers. A follow-up study using the same setup and software but with a different school that considered developing higher-level thinking skills a priority, resulted in significantly better engagement from both the teacher (one teacher was involved in this case) and the students.

5.1.2 *Understanding timetable pressures and implications on study outcomes.*

Study Context B: TUIs and GUIs for primary programming. Following a successful study investigating the use of a tangible programming environment with pairs of children at a school in Saudi Arabia, AA, with support from KH and her other PhD advisor, discussed returning for a second study to evaluate the impact of the physicality of the robot being programmed. However, the school was concerned about pressures on their timetable. Another school was happy to host the study, but to better fit with their school day, it was necessary to reduce the session times with each pair of children. The researchers decided to cut out two of the learning activities that were included in the previous study to reduce the session length. Unfortunately, after the study was completed the results showed that one of these was seemingly the most crucial activity in terms of learning outcomes and in prompting spontaneous use of gestures, an aspect the researchers were particularly interested in studying. Although this made it hard to answer the original research questions, it opened up the opportunity for an interesting comparison with the first study and encouraged us to investigate whether spontaneous gestures are more likely to occur alongside more in-depth and complex learning activities.

5.1.3 *Anticipating student selection.*

Study Context D: Physical Computing in the Classroom. During a full-year study on introducing problem-based learning with a physical computing toolkit to the classroom, MVW first had to negotiate timetabling of the study with

the school. On paper, the two classes were chosen due to timetabling decisions made by the school. However, in early conversations with members of staff, it emerged that student ability had played a part in the decision, and that there was a perception that the higher-ability students would be 'easier' to work with due to good behaviour and engagement with their learning. Ensuring that research was carried out with well-behaved pupils was perceived by the teachers and the school to provide a better experience in the process of research, and therefore a better image in the eyes of the researcher, without the challenge of disruptive classroom behaviour and lack of engagement perceived by lower ability classes. However, MVW was concerned that in developing a curriculum based on high ability pupils that the potential attainment gap between ability-setted pupils could increase.

5.1.4 Reflection. Working to align expectations from the outset, and understanding the impact of changes in study design is key to successful deployment and evaluations of educational technologies in school settings. Reflecting upon the complexities presented in the vignettes above, and aligning ourselves with existing bodies of research [38, 41] we suggest that researchers seek to understand the realities of school schedules and the competing demands on students and teachers at the outset of any research project. We suggest a series of practical implementations to align expectations in the design of evaluation research.

Firstly, from our collective experience, UK schools may be more flexible in the later summer term. While this may mean competing with other informal school activities such as sports day or school trips, as well as other external agencies who may be working with students in the school, this can avoid core examinations that may prevent teachers from engaging in "risky" innovations [54] Typically, there is more pressure on timetables in years where standardised testing takes place, and teachers will often encourage evaluation with other year groups for whom this may not be a constraint.

Beyond the practicalities of understanding school schedules, there is a clear need to communicate goals and expectations of the research with schools and teachers from an early stage in the research process, potentially including them in the design of evaluation research [22]. Within HCI, approaches to the evaluation of educational technology are moving towards objectives that go beyond just learning outcomes [53]. However, schools and teachers are less interested in the specific designs of technology, but rather with an improvement to teaching and learning outcomes in the first instance. For example, schools may be more focused on the resource implications of research regarding planned activities and impact on teacher time, while we should be centering the importance of managing expectations between teachers and school management and how this influences their alignment with researchers.

An additional consideration is the challenge posed by a mismatch of expectations of students between researchers and teachers, impacting their approach to the evaluation and consequently affecting data collection (as discussed in [35, 69]). Understanding the reasons why particular participant groups may have been chosen for a study can help researchers in their plans for classroom research in the evaluation of educational technologies.

Nonetheless, understanding the realities and constraints of school environments that influence teachers' beliefs and practices is essential for designers and researchers of educational technology when planning for research in the classroom, with a particular focus on noting and reflecting upon differing expectations and priorities of those involved. Through the chosen vignettes, we demonstrate how researchers made compromises to minimise disruption or align expectations, particularly when changes occurred or became apparent part-way through the project, having unpredictable effects on study outcomes. However, it is also important to note that these changes also revealed new insights and directions for the researchers involved. Understanding the pressures and constraints that teachers are working under can help us design technology that can be used in real classrooms. Seeking out 'ideal' settings for

evaluations could give a misleading impression of how novel educational technologies will be used in real school settings.

5.2 Considering the Impact of Researcher Presence

While evaluation research is often aimed at being carried out in a typical classroom environment, it is important to be sensitive to how researchers and research studies are perceived in schools - as the very presence of a researcher can change how the typical behaviour of the students and the classroom context unfolds during the study. In our experiences, researcher impact on the evaluation of educational technology in the classroom must be considered in the following ways: 1) understanding the complexity of the researchers' role, 2) the influence of being an outsider, 3) perceived power of the researcher, 4) researcher gender effect.

5.2.1 Understanding the complexity of the researchers' role.

Study Context C: Classroom Orchestration in the Performing arts. At various points during the longitudinal, co-teaching study run by RN, the partnered teacher was absent for a variety of reasons and had a cover teacher provided by the school. This was frequently done with short notice, often after lessons had been planned in detail. To ensure the continuity of the project, the teacher asked the researcher to teach the lesson rather than the supply teacher. The researcher taught the lessons on each occasion, despite the fact that these sessions did not offer any opportunities for data collection as the teacher who was the focus of the study was not present. The researcher placed importance on student learning and was aware of the importance of not allowing the research to overshadow the continuity of their learning experience. Ultimately these sessions were necessary to ensure the success of the project but created complex considerations of the role of the researcher in the classroom in longitudinal and embedded research settings. Understanding where the researcher priorities lay in this scenario was complex and centred around questions of understanding the research setting and ensuring students involved in the study were not affected by their participation. For the teacher involved, their priority was student learning, meaning that they had expected the researchers' role to align with theirs in prioritising student learning, sometimes resulting in the researcher being perceived more as a teaching assistant than a researcher.

5.2.2 The influence of being an outsider.

Study Context D: Physical Computing in the Classroom. Following on from a previous case study within the same school, MVW went on to work with two further sets of Year 8 students. As she was present for each of their computing lessons, often spending full days supporting the school's computing department, she felt that integrating into the school environment was an important step in the research process and was eventually given a staff badge to allow her to access the school.

In the second set of Year 8 students, behaviour and engagement deteriorated throughout the planned lessons. Upon discussion with the teacher, this was considered to be influenced by the researcher's position within the school, with students considering her *"just another boring teacher"* from the computing department, whom they had yet to meet. In planning for a subsequent iteration study with a new group of Year 8s within the same school, the teacher suggested that a clear distinction of MVW as a visitor might improve pupil engagement, as students enjoyed having guests in the classroom as it made them feel special. In the following iteration, the research design built upon the affordance of the 'researcher as an outsider' where the researcher returned to wearing a visitor's badge, noting the improvement in engagement and behaviour from participating students.

781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832

5.2.3 *Perceived power of the researcher.*

Study Context G: Content Creation Tools and Cross-Cultural Learning. VS, the lead researcher in this study, led a 5-week activity with Year 8 students in a school in India on a project-based learning (PBL) activity to embed cross-cultural understanding between students in England and India. The lead researcher noticed how her background of pursuing her research work in a western university within the field of HCI, opened up avenues to identify necessary research partners in India and roll-out workshops and evaluation activities that would have otherwise been difficult for local researchers. The school had a strict policy on digital technology usage where students could only access devices in the technology room, and students could not use any of the school's devices that were connected to the internet. The researcher was able to discuss the needs of her project and gain permission to bring into the school individual iPads for students taking part in the research that also had data connectivity. The headteacher provided permission for students taking part in the research activity to be out of lessons and to take part in evaluation activities at the end of the research. The school also contacted the parents on short notice to obtain parental permission for their children to stay back after school and take part in the end of activity presentation which was specially organised in an external venue with a high-quality audio-visual set-up.

5.2.4 *Researcher gender effect.*

Study Context E: Concreteness Fading for Computing Concepts. This study was carried out in a UK primary school by AT, with support from KH and his other PhD advisor. Due to the early stage of the prototype learning environment being evaluated in this study, it was necessary for the researcher to support interaction on a one-to-one basis with children. When analysing pre- and post-test data, it was found that male pupils outperformed female pupils on learning gains across all four study groups. Although differences did not reach statistical significance, this pattern raised concerns for the researchers. There was no evidence of any differential effects of gender on scores from a Cognitive Abilities Test the school had recently carried out, yet there was a notable trend for greater learning gains for male pupils in all four groups. One explanation the researchers considered was the biasing of results, as there is evidence that the gender of a researcher can influence the outcome of a study. All sessions were run exclusively by a male researcher (AT) in a one-to-one setting in a corridor outside the classroom. Each participant interacted closely with the researcher during the instructional phase and then completed the testing phase in relative proximity to the researcher (due to the constraints on space, but this may have been misinterpreted by the children as observation). No between-gender differences in ability or attitude were noticed by the researcher during data collection, potentially due to their subtlety and the implicit nature of the effect.

5.2.5 *Reflection.*

The phenomena of researcher presence and school performativity impacting upon research studies have been experienced by a number of the authors and highlight the importance of clarifying the role of the researcher within the classroom environment. One ongoing consideration is understanding and acknowledging the potential impact of a researcher's presence in the classroom environment and how that aligns with or can affect the aims of the study. This is a particularly pertinent consideration when carrying out long-term or embedded work [22, 25, 27].

Whilst the most naturalistic educational technology evaluations are those in which researchers are not present in the classroom; there are many reasons why researcher presence is important in evaluating novel technologies, including the need for observation and other data collection, and the facilitation and management of prototype systems. In such cases, researchers who are seeking naturalistic technology evaluations are often aiming to integrate themselves into the classroom environment in order to gain insights into the socio-cultural dynamics of the area of study [76].

While this can offer important insights, there are some research studies that are designed to benefit from the researcher being an 'interesting outsider', as someone with whom students can engage and where the insights rely on students and staff being able to share their experiences with a researcher [91]. Understanding and reflecting on the potential impact of the researcher's role in the classroom and in the wider study is key to ensuring that the insights gained are those that align with the aims of the study. Where the research seeks to implement a co-design or action research approach, this can increase the complexity of the researcher's role, particularly when negotiating shared aims of the research [71]. There is a necessary importance placed on achieving community aims when using participatory approaches [44, 85]. However, balancing this approach to educational community-engagement alongside the specific research aims can present a challenge. The negotiation of a researcher's role in such studies is often a challenge experienced on a personal level for many researchers seeking to balance their priorities regarding the research alongside the well-being of participants and meeting the agreed community aims. Approaches taken by researchers in these contexts will often then have an impact upon the school response and engagement [38]. Although this is often an ongoing challenge, in acknowledging how their position impacts the classroom environment, researchers can gain a deeper understanding of the school and classroom environment in which the educational technologies are evaluated, particularly in cases where educational technology is planned to integrate into an existing classroom setting.

The acknowledgment of positionality and elements such as gender effect are both complex and individualistic to the context of research [1]. These can be tied more deeply to interpretation rather than planning but are important elements for researchers to reflect upon in their documentation and presentation of research. Both culture and gender were felt by authors to have an effect on studies in a variety of ways. In a similar way to the discussions of the researcher's role above, these effects and their potential impacts changed depending on the aims of the study in which they were experienced. In one case, the impact allowed the researcher access to what they may not have otherwise had, whereas in another case, the impact was felt to affect the results of and therefore insights from the study. While these are not effects that can always be mitigated, consideration of the potential impacts in the study design phase can lead to specific reflections on the data and the resultant impact when reporting the findings.

5.3 The reality of the school environment

Classroom research requires researchers to be constantly ready to adapt to the unexpected realities of the school environment [41]. Schools are typically lively and dynamic places, and changes and disruptions to research study plans are commonplace. Here, we highlight the types of disruptions and classroom changes that have impacted our work, categorised as 1) disruptions to session structure, 2) disruptions during sessions, 3) disrupted access to technology, and 4) the impact of setup and time.

5.3.1 Disruption to planned session structure.

Study Context F: Teaching Critical Thinking About Sensors and Sensing. As part of an evaluation study to assess the value of a physical computing toolkit for supporting critical thinking about computing concepts, SL and colleagues carried out one-off data collection sessions at six primary schools for an evaluation study. Upon arriving at one of the schools, the researchers were informed by the class teacher that they would have one hour less than expected with the students. The reasoning given for this was that there was a last-minute assembly organised by the school that all students were asked to attend. This meant that the researchers had to think quickly on their feet to cut out some of the activities they had planned for the session. It also meant that the amount of material they were able to cover in this session was not the same as in sessions deployed at the other five schools participating in the study. Not having data

for the final hour of the planned activities was problematic for our analysis, which aimed to compare findings between schools. Ultimately, in the publication that emerged from this study, the focus of the analysis was narrowed to focus on one specific section of the sessions, and the missing data from the final hour was not needed; however, the missing data from this session played a role in the decision of which learning segments to ultimately analyse.

5.3.2 *Disruptions during sessions.*

Study Context D: Physical Computing in the Classroom. During the physical computing in the classroom project, there were frequent interruptions caused by fire alarms as part of an ongoing school-wide behavioural problem external to the class with whom MVW was working. In the first six weeks of the iteration, fire alarms could be triggered several times over the course of the 50-minute lesson, and teaching would have to stop so that the class could safely evacuate to the tennis courts on the far side of the school. What little time remained of the lesson was spent revisiting incomplete topics, with teachers noting the lack of flow to lessons, which contributed to the decreasing engagement and poor behaviour from the Year 8 pupils. This culminated in the teachers and researchers being unable to continue with the curriculum development research for that term, and pupils returned to their original scheme of work.

5.3.3 *Disrupted access to technology.*

Study Context C: Classroom Orchestration in the Performing Arts. Access to technology for some school subjects can be a challenge, and teachers often rely on a booking system to access computer rooms. These computer rooms are commonly bookable by the whole school, and changes can happen regularly. Towards the end of this particular project, the students were finishing e-portfolios which included gathering necessary evidence into a final portfolio document and writing reflections on their experiences throughout the project. The whole research project was predicated on the use of digital technology to do this. However, due to exams, the lesson was regularly re-timetabled back into the music classroom, with no access to desktop or laptop computers. This was done regularly without prior warning, meaning that plans were often changed, and RN had to bring in university Chromebooks to ensure students could continue with the project. Ultimately, this had an effect on the observations of the use of technology as Chromebooks were not how students and teachers would typically access the technology being evaluated.

5.3.4 *Impact of Setup and Time.*

Study Context A: Interactive Tabletops in the Classroom. AK and colleagues carried out research on tabletops in school but were not given a dedicated space for the 6-week duration of the study. For each session, the researchers needed about 30-60 minutes to set up the seven interactive tabletops and observation cameras in the given space, and then needed another 30-60 minutes to take the tables out of that space and store them away. This meant that when a space was picked, the school needed to make sure it was not booked for the hour before and after the session which limited the options of space available leading to some sessions being carried out in less than ideal spaces. It also meant that the researchers needed to allow for up to 180 minutes in school for each 60-minute session (and this was done for 11 sessions in total).

5.3.5 *Reflections.*

Schools are fast-paced and changeable environments with a large number of competing interests and policies, especially given the range of stakeholders who are involved in day-to-day school life [33]. While this makes in-the-wild evaluation all the more important as a model of evaluating educational technologies, it can also present complex challenges, especially to do with navigating disruptions to carefully planned session structures and material setups.

Researchers working in school settings will necessarily find themselves having to change their plans to accommodate the ever-changing realities of school life [1, 64].

One particular challenge can be the impact of disruptions on the available time for the study, whether because of specific time constraints or behaviour challenges that mean activities take longer or need to be truncated, as we saw in the vignettes above. Having alternative activities planned can support researchers' flexibility in these settings and is something that is a recommended best practice based upon our collective experience.

The replicability of studies can sometimes also be challenging, especially given the variability of school environments; this variability can be seen in everything from the types of technology used to the range of social actors and physical space available in any given school [38, 41, 82, 83]. For researchers working in-the-wild, the importance of reflexivity and detailed accounts of context become key to enabling the research to be interpreted and used by others [78]. Understanding and highlighting where contexts changed between studies can provide interesting and sometimes unforeseen insights into educational technology and its uses.

Understanding and exploring the possible variables of a particular school setting at the outset of a research project can give a researcher some initial insight into the potential mitigations they may have to make along the way. When designing research projects and learning activities, flexibility is key to enable researchers to respond and react to changes in the expected process.

5.4 Spaces, Objects and Devices

In many cases, evaluation studies require setting up the physical space, testing instruments and devices where the setup and use of study equipment can strongly influence the quality of data gathered during a study [8, 33, 69]. In classroom evaluations, this presents a range of potential challenges, including 1) the novelty of recording equipment, 2) seating type as a disruption, 3) seating arrangements and observation, 4) desk configuration and collaborative activities and 5) spaces assigned for research.

5.4.1 Novelty of recording equipment.

Study Context H: Analysing Indicators of Collaboration with a Physical Computing Toolkit. In a study where sessions were run in a series of schools with 9-11-year-old children, SL planned to both video and audio record the children during the sessions. The children were asked to work in pairs in their typical classroom environment, and their dialogue within their pairs was integral to the analysis. For this reason, one audio recorder was placed on each pair's table. In the first session in this series, the audio recorders were briefly introduced to the children before the session began. Specifically, the children were told what the audio recorders' purpose was and were then asked for their consent to their voices being recorded. The learning activity itself was designed to be high energy and playful - leading the children to interact with each other and the pedagogical materials around them in a way they might not in a typical classroom. A trend started during the session where the groups of children would play with the audio recorders - turning them repeatedly on/off, singing into the recorders, etc.

All of the data from this session had to be discarded because of this. In the next session (with a new group of students), a longer discussion was held with the children at the start, engaging them more with thinking about data collection and why it's important to leave the audio recorders alone during recording. Stickers saying "please do not touch me" were also placed on each recorder. Placing a higher emphasis on the importance of the data collection instruments and reasoning for them was crucial to ensuring the data collection went smoothly.

5.4.2 Seating type as disruption.

Study Context A: Interactive Tabletops in the Classroom. In a study introducing tabletops to the classroom, due to running sessions in different spaces in the school each time, the first session AK ran was conducted in a room with swivel chairs. The smart tables were spread over the room in a grid-like manner, with two to four swivelling chairs around each table. The use of swivel chairs in combination with the layout meant that students spent a lot of their time swivelling around on the chairs rather than paying attention to the smart table task. While this bit was not reported in the published work, the use of these chairs presented challenges to teachers in controlling the class (resulting in frequent interruptions trying to keep students focused) and to the researchers in terms of the quality of the data collected. For later sessions as part of that study, the use of swivelling chairs was avoided wherever possible.

5.4.3 Seating arrangements and observations.

Study Context B: TUIs and GUIs for Primary Programming. In a study where AA and KH were exploring young children's use of spontaneous gestures when taking part in programming activities, the school assigned a science lab as the study space. After 8 participants had taken part, AA noticed a pattern that the children were gesturing much less than in previous studies. She realised that the high stools used as seating the lab seemed to be inhibiting children's spontaneous gestures, possibly because they were using their hands to ensure they stayed balanced. The data from this group of participants had to be excluded for this reason. The experience made the researchers wonder about the broader implications of this for lab-based teaching for younger children, given the evidence that gesturing can be important in the learning process.

5.4.4 Desk configuration and collaborative activity.

Study Context H: Analysing Indicators of Collaboration with a Physical Computing Toolkit. SL and colleagues ran a study in 5 separate schools with children 9-11 years old to evaluate whether and how a novel tangible microcontroller device would enable collaborative learning to take place in classroom settings. The researchers began observing that the set-up of the classrooms seemed to have a strong effect on the way the children in each school/classroom interacted and collaborated with others. The set up of the classrooms across the schools was diverse, including benches where children would sit side by side, facing the teacher, tables where groups of four sat facing each other, and computer rooms where all of the computers and seats faced the wall of the classroom. The configuration of seating within the classroom afforded/constrained who each child could collaborate with and to what extent.

This had an effect on what actions and gestures would be visible to others around the child. For example, configurations where there was more visibility toward children further away was found to support children in talking/working with those not in their immediate proximity. Configurations where the children were able to face each other rather than sit side by side, was found to also support a larger range of gestures and interactions. In sum, the configuration of each room was seen to be a contributing factor to the findings of how and to what extent, the children collaborated with the interface.

5.4.5 Spaces for research.

Study Context E: Concreteness Fading for Computing Concepts. AT initially piloted the study in a spacious and quiet workspace separate from the classroom, as part of an ongoing project with KH. With this setting in mind, the planned study design included the use of A/V equipment to capture detailed interactions with the physical prototype. For the second pilot, this workspace was no longer available and, instead, an area in a busy staff room was provided. The floor-standing camera tripod became a trip hazard and background noise meant the participants' comments were difficult to interpret. For the full study, the only available setting was a corridor connecting multiple classrooms.

Although the space physically allowed for the use of A/V equipment, the high footfall would have led to recording children for whom no parental consent was obtained and, therefore, it was not possible to collect these data. In addition, the corridor workspace was directly adjacent to the participants' classroom, leading to many viewing the materials and activities before their assigned session. As the study was evaluating distinct representations in a between-groups design, outcomes may have been influenced by false preconceptions, particularly within attitudinal measurement. Further to this, the novelty of the presence of a researcher and the fun, colourful, prototype encouraged some children (and staff members) to interrupt the study session to ask questions. This distraction did not occur evenly between groups as some representations were more 'engaging' than others, potentially negating their effect.

5.4.6 Reflections.

Whilst the 'novelty effect' of educational technology is something researchers are well aware of as an issue [9]. However, the unique position of educational technology research in HCI presents new considerations in research design. The focus on evaluation beyond learning outcomes transforms research in order to explore factors beyond learning outcomes and into the wider implications of its design and affordances in context [7, 58, 80]. The qualitative nature of these types of evaluation can require video and audio in addition to the educational technology involved resulting in additional challenges to the evaluation process. Therefore, the impact of novelty on other aspects of the study context (such as recording equipment, and rooms and furniture that are unfamiliar to students) should also not be overlooked.

Our vignettes illustrate how equipment and objects that are secondary to qualitative evaluation can be distracting to students. Video cameras and audio recorders can provoke feelings of shyness and/or encourage students to perform for the recording. Our experiences indicate that briefly pointing out the recording equipment as part of the informed consent process can sometimes result in drawing attention to the devices in a negative way. We recommend a longer discussion of the importance of data collection to a research project in addition to pointing out the devices, which can bring students on board and give a better sense of the importance of their contribution. Beyond the study equipment, unfamiliar spaces and objects within the school can also cause distraction. Whilst the ideal context for in-the-wild studies is students in their regular classroom setting, the practicalities often mean that an alternate classroom or space is used. Researchers should establish exactly where the study will take place within the school in advance and inspect the space before the study begins, wherever possible.

Furthermore, particular attention should also be paid to the space of a classroom, including furniture and configuration as it may affect data collection. This is especially important where the goals of the evaluation include analysis about physical interactions, gesturing and embodied indicators of collaboration. As our vignettes demonstrate, the configuration of desks can influence the individuals that each student can interact with when completing a learning activity - be they other students or the teacher. Additionally, the shape and size of seating arrangements can also influence the body language and gestures students are able to use with others, as well as what can be observed by the researcher and recording equipment. Where the setting and seating arrangements accurately reflect the real classroom environments in which the novel technology is likely to be used, any issues that occur have clear relevance as part of the evaluation of the technology itself. Where a non-standard setting is being used, it is harder to draw conclusions and is more indicative of a problem with the design of the evaluation.

6 DISCUSSION

Work in educational technology research commonly looks at the challenges regarding implementation [27, 71], evaluation of technologies outside of formal classroom environments [88], or review of existing technologies [1]. While we

recognise the importance of these forms of educational technology studies, these do not help to address the complexities of HCI researchers conducting in-the-wild research of novel technologies that they have developed in classroom contexts. Therefore, our work contributes a series of practical recommendations that together offer an orientation for HCI researchers looking to conduct evaluations of educational technologies in school settings. The provision of these insights is offered with the aim of starting a discussion amongst researchers and encouraging reflection as a way of starting to address some of the pitfalls experienced when evaluating educational technologies in situ.

Through this reorientation to educational technology evaluation in school-based contexts, we suggest that as a community we can begin to address some of the logistical, material, and methodological constraints present in modern educational technology evaluations in school-based contexts [18, 53, 68].

As a result of our reflections on the vignettes in the section above, we offer a series of recommendations (R1-11) for educational technology research design, developed by reflecting on the most commonly experienced challenges by the authors. These recommendations offer opportunities for other researchers to build on authors' insights regarding the impact of the school environment, the relationship between schools and researchers, a researcher's impact, and the design of and reporting on educational technology research.

6.1 Research in Complex School Environments

Schools are naturally complex environments, representing an interplay of multiple constraints and opportunities [33]. Therefore, it is important for HCI researchers to prepare for these complexities and their disruptions through the design of their research, considering the realities of a classroom environment and the disruptive influence introduced by their research process. This is particularly relevant for HCI researchers given their focus is often on the development of novel technologies, and understanding student and teacher interactions with these technologies in context is key.

R1. Adopt a modular research design with a critical path. Deviations from a research plan are not uncommon in school settings, whether in terms of session cancellation or interruptions, change of physical space, or technology/infrastructure. These types of failures are a regular reality that teachers have to deal with on a daily basis, with considerable educational research dedicated to the process of designing and adapting lesson plans in response to disruption and change [20, 67]. However, this approach to research planning is rarely discussed in HCI literature [53].

Therefore, we recommend that during the planning of research evaluation activities, HCI researchers should identify a critical path for data collection, that represents the absolute minimum for the success of the research. This includes designing research with a modular approach, within which activities can be added or removed to this 'critical path' [91] dependent on the potential provided by the setting. This approach can allow for planned activities to be adapted in response to student engagement, classroom disruption, and unexpected, unplanned change. Moreover, researchers should ensure contingency plans for critical path activities including ensuring the availability of secondary research dates should unavoidable disruptions be experienced during the planned research. This is particularly important to consider when conducting research towards the end of the school year, where an interrupted session can mean no further data collection is possible with that class before the holidays begin or is in conflict with other informal school activities.

R2. Prepare for the physical classroom environment. The physical environment of a classroom is also a key factor impacting the evaluation of educational technologies in situ, and can present a range of pitfalls to consider when designing the evaluation of educational technology. Where existing educational technology research discusses space, this is most frequently during the evaluation of existing technology products, rather than novel technologies developed over the course of an HCI-based research project (e.g. [61]. When considering the impact of the physical classroom

environment from the perspective of an HCI research evaluating a novel technology, there must be some consideration of how the physical classroom environment will impact the development and process of the research. For example, the impact of physical equipment setup time on research sessions needs to be taken into consideration in timetable planning with schools and in the design of research, or how this translates into classroom use [64]. Furthermore, it is particularly important to consider the physical environment if the analysis involves physical interactions or gestures, or embodied indicators of collaboration. Importance should be placed on visiting the room(s) planned for the research before conducting the research to try and anticipate and address any potential issues regarding the physical environment prior to the collection of research data. This recommendation also relates to contingency planning (R1) which helps mitigate problems related to timetabling and the physical environment as well.

R3. Account for the novelty effects beyond the studied technology. Even with contingency plans, school-based evaluations of educational technologies are still prone to unexpected issues, as demonstrated in the vignettes regarding swivel chairs, voice recorders, and desk placements. Researchers are often mindful of the consequences of the novelty of the investigated technology itself and take steps to overcome their associated challenges [30, 54], yet our vignettes highlight how we, as researchers, can also often overlook the potential negative impact of the novelty of research elements like the recording equipment, and new spaces. Such phenomena are more difficult to plan for, as they may only become apparent at the time of conducting the research. Our recommendation is for HCI researchers undertaking the evaluation of educational technology in school settings to provide sufficient attention to any potentially distracting objects, giving students the opportunity to explore them when possible, and having a clear discussion with them about possible negative consequences on the research and its data. During the vignette about recording equipment, discussions were held about the importance of data collection and the use of informational stickers, and in other experiences, the students were given the opportunity to play with the equipment prior to beginning the data collection (e.g. [59]). In response, students were less likely to become distracted by the novelty of the research equipment with which they were presented.

6.2 School-Researcher Relationships

Teachers, educational technology designers, and researchers often have differing criteria for evaluating the success of educational technology [84, 85]. This, coupled with the complexities of the demands on teachers' time, can make research relationships between teachers and HCI researchers challenging [22] and affect evaluations in context [54]. The drive to create mutual researcher-research partner relationships is neither novel in HCI [58] nor in educational technology research [41], where there is an increasing focus on the place of mutual, respectful relationships between research teams and those within the research context. However, there is limited research focused directly on how HCI researchers can work alongside school partners in the classroom context. Therefore, we recommend that HCI researchers work to develop a communicative relationship between themselves, partner schools, and involved teachers to directly discuss the impact of timetable and scheduling on research plans, and cover the process for classroom management in the face of disruption when evaluating the educational technologies in school settings.

R4. Aim to clarify the goals and expectations for the research prior to commencing the study A mismatch between researcher and teacher plans or expectations when undertaking the evaluation of educational technologies in the classroom can result in a significant negative impact on the research process and its outcomes [38, 41, 54]. For example, current HCI research is tending towards the evaluation of 21st-century skills [29, 55], whilst schools are legally required to demonstrate improvements in pre-determined categories of learning which impact upon school expectations and hopes for education and learning-focused research [54, 84, 92].

When school-based research is designed with collaborative, equitable relationships in mind, wider literature demonstrates how teachers report a better understanding as to how the research was relevant to their own educational environment, as well as providing opportunities for self-reflective teaching practice and professional development opportunities for teachers[33, 64]. Through these relationships, researchers can gain an understanding of the experiences, practices, needs and values of the teachers. Establishing shared goals and expectations can support the design of research projects that result in technologies better suited to the classroom. If the design of educational technology for classroom use does not meet teacher needs, they are likely to revert to what Brown and McIntyre referred to as 'normal desired state' of student activities [17]. Drawing from existing HCI and educational research literature, we propose that HCI researchers include an introductory session to their research for those involved (where feasible) to help address these challenges when conducting school-based evaluation research of educational technologies.

R5. Work to understand the particular constraints of a school, e.g. scheduling and timetable Working in situ to understand the constraints of a project is inherent to increasingly popular approaches to research methodology (e.g. [44, 78]). However, in Lai and Bower's [57] systematic review of educational technology research, only 1.4% of reviewed papers considered the institutional environment when evaluating learning technologies in an educational context, and only 37.9% of these papers focused on formal, compulsory education context. However, there is no clear overview of how many of these reviewed papers that focused on compulsory levels of education were conducted within a classroom environment.

Through our reflections, we highlight how understanding the constraints of the institutional environment can actually support researchers to avoid challenges during the process of their educational technology evaluation research. We realise that not every researcher will have the availability or resources to conduct long-term, embedded research in their educational context, but we do recommend that researchers explore some of the following common constraints most relevant to undertaking evaluation of educational technology research from an HCI perspective.

Firstly, working with the school to understand the timetable of the school day, week and year can support the researcher in identifying a suitable time period in which to conduct the study. It can also help researchers to understand the potential time available for the study, as well as the overall constraints on teachers' time and resources at any given time in the school year. For example, during exam times, there may be conflicts on space in schools, teacher availability or supervision capacity. Secondly, it may be important to understand how school policies, such as behavioural or ability-based setting policies (setting refers to how students are allocated to teaching groups, either grouped by ability or mixed) may impact the demographics of students involved in the evaluation process. Understanding these constraints will aid the researcher in discussions about goals and expectations (R4) once they are aware of the potential constraints [14, 76].

R6. Understand the place of the research within the wider school context.

As a final consideration of undertaking research in complex school environments, is understanding how the research is perceived within the wider school setting. While undertaking a review of understanding the expectation of the research (R4) is its own recommendation, it is important for HCI researchers to understand what potential impact or benefit their research may have upon the school and wider educational community in both the short and long-term [39]. For example, one of Lai and Bower's [57] recommendations was for researchers to reflect on the many aspects of the educational environment, and develop research that exists within the holistic school environment.

Drawing from the educational technology research perspective, this can help engage teachers with the evaluation process, as educational technology research notes that teachers can place more value on a study when they are invested in the research [22]. Where teachers have a deeper investment in the outcomes of the research, our vignettes point to

further opportunities to adapt the study in response to disruption with their help and support. Therefore, we recommend that HCI researchers enter the evaluation process with an understanding of both the school and teacher's priorities and values regarding the research, in order to provide a foundation for researchers to develop contingency plans for their evaluation research, as well as respond to in-the-moment challenges when evaluating in school settings.

6.3 Researcher Effects

Researcher and teacher expectations of children can have significant effects on their learning outcomes [79], challenging the evaluation of educational technology in school settings. For HCI researchers who are often involved in the design of the technologies as well as the evaluation and implementation it is especially important to acknowledge the realities of researcher impact on not only the design of the educational technology itself but also through their interactions with the school how this may impact school behaviour including how this may influence student selection.

R6. Assure teachers and students that it is the technology that is being evaluated and that the research is seeking honest feedback and experiences. Researchers must acknowledge the impact of their presence as observers on wider school, teacher, and student behaviour. This is of particular importance where school stakeholders may modify their behaviour in order to portray a certain image to the researcher [1]. To address this challenge, the goal of the research should determine the role of the researcher. In longitudinal, action research work the researcher may be internal to the educational environment and begin to address the challenge of their presence upon the teacher and students [22, 91] Short-term research, on the other hand, may need to take explicit steps to reassure teachers and students that it is the technology that is being evaluated, not them and that they want honest feedback and experiences so they can improve the technology if needed. In short, researchers should consider their position within the research and the impact this has on their data collection and analysis.

R7. Ensure bias in student selection is avoided. A further element for consideration when conducting evaluative studies of educational technology is the notion of performativity when researchers conduct research in educational contexts. In the UK, neoliberal educational policies (such as league tables) constrain creative freedom in the delivery of school-based teaching [4] and lead to tension when conducting research in the classroom – schools may be selective about which classes participate to provide a 'good research experience' for the research team, at times in contrast to a researcher's aims. This form of experimenter impact can modify perceived school behaviours, for example resulting in the bias of evaluation results where cherry-picked, top-performing students may perform differently from their mixed-ability peers.

Existing literature reviews in the evaluation of HCI already point to the challenge presented by a limited focus on the constraints of the school setting, such as socio-cultural background and policy [57], which can shape participant demographics without potential consideration from an HCI researcher unfamiliar with the educational context. Therefore, We encourage HCI researchers to address this with teachers directly and resist any suggestions that classes or children will be picked based on their expected good behaviour or performance. Discussions of research aims, expectation setting and reminders that the research is evaluating the technology and not the school can help researchers to address these challenges that may not often be considered in educational technology evaluation research, and contribute towards the longevity and impact of research findings.

R8. Report the positionality of the experimenter along with the results of the study. A key part of the realities of educational technology research in school settings is acknowledging the potential for experimenter impact. While this recommendation for a reflection on author position and experience is not novel within the intersectional research

community, existing literature is spread across areas of research such as design research, HCI, Education and Educational Technology [58, 60, 72] and educational levels (with Higher Education featuring most prominently, e.g. [88, 90]).

Throughout these vignettes focused on HCI research in school settings, the research process has been influenced by the complex interplay of epistemological beliefs, cultural background, and other relevant characteristics of the researcher. The positionality of the researcher cannot always be mitigated, nor should it be, as it provides rich insight into the design and interpretation of research [26]. Therefore, from this work we recommend that authors take the time to include a statement of positionality that outlines their epistemological stance and background when working in school environments, providing a lens for readers to interpret research outcomes and allowing for further exploration of reflexivity and vulnerability when conduction educational technology evaluation in school settings.

6.4 Designing and Reporting of Educational Technology Evaluation Research

The design and reporting of educational technology evaluation research can often be divorced from the messy reality of its environment of delivery [28, 53, 89]. In a field such as HCI, where the ecological complexity of contexts such as schools is key to our understanding of the potential role of educational technology, the reporting of the design and evaluation is that much more important. Without detailed reporting, it becomes more difficult to build upon existing work. In this paper, the vignettes presented demonstrate the complex realities of the research, which were not always fully presented in the associated published papers, limiting the progression of this field of research. To address this, we recommend researchers clarify their approach to evaluation, as well as reflect and report the messy realities encountered in their research to support replicability and further validation of their educational technology research.

R9. Clarify the definition of evaluation to be used within the research and make this explicit through the research questions. Returning briefly to the concept of evaluation, we must first acknowledge that evaluation has many meanings amongst educational technology research communities. Evaluation can represent the exploration of the perceptions of learners, as well as evaluation in the context of use, or addressing uncertainties regarding its introduction, acceptance or associated pedagogical development in the classroom [14, 31, 57]. These can encompass evaluation constructs such as learning, affective elements, behaviours, design, technology, pedagogy, presence and institutional environment [14, 57], with the evaluation of these different elements having implications on the design of educational technology research. Lai and Bower's review of the evaluation of educational technology notes the danger in evaluation instrumental, the inconsistency in the definition of evaluation means that it can be challenging to evaluate findings across studies and make relative judgements [57]. Therefore, we recommend that authors are explicit in their definition of evaluation used in their research, and explore its implications on the research questions to help support the balance of rigour and relevance that is currently so challenging in educational technology research in context [80].

R10. Share and report on the messy realities of classroom research. Research involved in the evaluation of educational technologies has begun to demonstrate a turn to embrace the messy reality of educational contexts [34, 76] and towards finding 'what works' in educational literature [10, 11, 81] rather than ignoring or undermining it to demonstrate a systematic and methodological measuring of learning outcomes. However, given the complexities of day-to-day classroom life, there are many more questions to be answered regarding both the potential role and potential usability of educational technology interventions that are not easily answered by only reporting 'non-messy' data. Dismissing or obfuscating the challenges related to messiness may hide important aspects of how humans interact with technologies in their environment, which is a central component of HCI research.

This echoes both our own experiences and literature, in which researchers are unlikely to discuss these disruptions and challenges in the presentation of their work [28], despite this being an intrinsic part of the school classroom

environment. To address this gap and advance the understanding of the evaluation of educational technologies in the school context, we recommend HCI researchers join us in documenting the challenges they experience during their research and evaluations, contributing to the exploration of how these disruptions can present new insights, opportunities and practices within HCI educational technology research.

R11. Embrace reflexivity in classroom research. All this discussion about the challenges of evaluating educational technologies in the classroom raises the question - how can we encourage researchers to share the realities of classroom-based evaluations with the educational technology research community and beyond? We recommend researchers engage in reflexive practices, work with teachers to understand the pragmatic issues encountered in classrooms and most importantly **normalise reporting the challenges of evaluating educational technologies in the classroom.** Fire alarms, swivel chairs, and supply teachers are not an unexpected part of school life, yet these can be both surprising and disruptive to researchers and their research when the process of sharing these disruptions is not commonplace. Normalising the sharing of such challenges can help researchers collaboratively construct a realistic picture of both evaluating and integrating educational technologies in the classroom [28], where design and acceptance are undertaken and shared from the perspective of a messy yet realistic environment.

6.5 Summary of Recommendations

In the following section, we present a table summarising the recommendations that we have derived from our collective, practical experiences. We hope these recommendations serve as a starting point for HCI researchers embarking into the world of school-based educational technology research, as they begin to foresee the complexities they are likely to encounter and consider these recommendations as points of reflection and discussion.

Recommendation	Example(s)	Benefit
Planning for Research in Complex School Environments		
R1. *Adopt a modular research design with a critical path.*	Identify a core 'critical path' for data collection and design activities to be modular so they can be added or removed to this critical path; ensuring the availability of secondary research dates to account for potential disruptions to critical path activities.	Supports successful and complete data collection.
R2. *Prepare for the physical classroom environment.*	Visit the room planned for the research, before conducting the research.	Understanding the constraints and affordances of the physical space; better planning of practical issues, e.g., where to plug in recording equipment.

R3. *Account for the novelty effects beyond the studied technology.*	Acknowledge the novelty of everything involved in the study to the students in class such as data collection equipment, furniture, setup, and researcher presence. Give students an opportunity to acclimate to any novel instruments and ask questions about them.	Reduces disruptions to the study and data collection and improves the quality of the collected data.
Developing a Strong School-Researcher Relationship		
R4. *Endeavour to establish strong relationships with the schools and teachers, even before commencing the study.*	Arrange in-person or video meetings with teachers and relevant members from the school leadership to clarify each others' needs and ensure alignment of expectations.	Ensures teachers and leadership are invested in the study and having the right expectations which consequently increasing school's adaptability and support for contingency planning.
R5. *Work to understand the development of scheduling and timetable, as well schools strategy for working with students of different ability levels.*	Ask teachers about their timetable for the year, and what they prioritise in different months (e.g., exams). Ask about how students of different ability levels are allocated to classes and groups.	Helps to support the researcher in choosing a time period in which to minimise potential distraction to their study, provides opportunity for the researcher to understand the impact of their presence within the school environment, and ensures the study is carried out with the targeted ability levels.
Accounting for Experimenter Impact		
R6. *Assure teachers and students that it is the technology that is being evaluated, and that the research is seeking honest feedback and experiences.*	Let students know that they should not feel they have to perform differently than in their typical classroom.	More genuine data by alleviating modification of stakeholders' behaviour in order to portray a certain image to the researcher.
R7. *Ensure bias in student selection is avoided.*	Communicate to the teachers that it is important that all students are included, not just the best students.	Avoiding bias of evaluation results, where cherry-picked, top-performing students may perform differently to their mixed-ability peers.
R8. *Report the positionality of the experimenter along with the results of the study.*	Acknowledge the epistemological belief, gender, cultural background, and/or other relevant characteristics of the researcher.	Provides insights into the design and interpretation of research.
Designing and reporting of educational technology evaluation research		

R9. *Specify what is meant by evaluation.*	Ensure a clear definition of what evaluation means in the context of this research and what are the constructs to be measured.	Aligns the reader's expectations with those of the authors in terms of what is evaluated and the research methods used.
R10. *Share and report on the messy realities of classroom research.*	Report on disruptions in the session structure, disruptions caused by instruments and materials, and unexpected incidents that influenced the data collection and analysis.	Supports situated knowledge, and the transition away from assuming classroom research should be as "objective" and "neutral" as lab studies.
R11. *Embrace reflexivity in classroom research.*	Write down, and report on, your reflections directly after the study about what went well, and what did not, as well as any observed potential confounding factors.	Ensure the research is responsive to the unexpected aspects of the environment

7 CONCLUSION

The themes and vignettes outlined in the section above examine the experiences of classroom-based evaluations of educational technology and the implications on research. While studying the effects of education technology in controlled settings remains a valid objective of research, it offers much fewer opportunities for research progression than evaluations in-the-wild [28, 38, 41]. Despite the challenges regarding the realities and the messiness of methods, data collection and evaluation of educational technologies in school settings, this paper presents reflections of the authors' own experiences of carrying out evaluation research in classroom contexts, sharing the messy realities of school-based educational technology research. We then use this as a springboard to provide practical guidance for future HCI researchers looking to evaluate educational technologies in the field, drawing upon learnings from educational technology, learning sciences and evaluation studies and contributing to the reorientation of future educational technology evaluation work and practices.

We offer reflections on four specific aspects of carrying out HCI research in classroom contexts, highlighting potential methodological challenges that may arise in the planning of classroom evaluations, the potential impact of researcher presence, the realities of working in the school environment, and the complications arising from introducing new and novel spaces, objects and devices when evaluating educational technologies. However, despite the many challenges, we also provide recommendations toward the reorientation of HCI-based educational technology research in school contexts, outlining practical guidance and strategies to reframe and respond to the challenges of school-based evaluations while elevating the importance of the messiness and richness of the data collected in these classroom contexts.

In sharing and communicating these commonly experienced challenges and vignettes, future researchers will be able to compare and contrast their experiences. While this review of the realities of educational technology evaluation was limited in its range of experiences, we hope that in sharing the often unpublished realities of evaluations in school settings, we will encourage further HCI researchers to share their unexpected outcomes during research and reflect on how to address such challenges. In reorienting how HCI researchers conceptualise, design and report the challenges

of evaluating educational technologies in school contexts, we encourage others to do the same, contributing towards the development of HCI-specific practices in school-based educational technology research across wider contexts and developing further understanding of how to address the realities of evaluating educational technologies in school settings.

ACKNOWLEDGMENTS

We would like to express our sincere thanks to all those who took part in the referenced research, particularly to the staff, pupils and research teams without whom these insights would not have been possible. This work was funded by the EPSRC, including the Centre for Digital Civics award number EP/L016176/1 and EPSRC and BBC iCASE scholarship award number 1623937, and Al Imam Mohammad Ibn Saud Islamic University (IMSIU). We also extend our thanks to Professor Yvonne Rogers for her time to provide feedback on this work in draft.

REFERENCES

[1] June Ahn, Austin Beck, John Rice, and Michelle Foster. 2016. Exploring issues of implementation, equity, and student achievement with educational software in the DC public schools. *AERA Open* 2, 4 (2016), 2332858416667726.

[2] Adriana Alvarado Garcia, Marisol Wong-Villacres, Milagros Miceli, Benjamín Hernández, and Christopher A Le Dantec. 2023. Mobilizing Social Media Data: Reflections of a Researcher Mediating between Data and Organization. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 866, 19 pages. https://doi.org/10.1145/3544548.3580916

[3] M. V Angrosino and K.A Mays de Perez. 2000. Rethinking Observation: From Method to Context. In *Handbook of Qualitative Research* (2nd edition ed.). Sage, Thousand Oaks, 673–702.

[4] Margie Appel. 2020. Performativity and the demise of the teaching profession: the need for rebalancing in Australia. *Asia-Pacific Journal of Teacher Education* 48, 3 (May 2020), 301–315. https://doi.org/10.1080/1359866X.2019.1644611

[5] Marlèn Arkesteijn, Barbara Van Mierlo, and Cees Leeuwis. 2015. The need for reflexive evaluation approaches in development cooperation. *Evaluation* 21, 1 (2015), 99–115.

[6] Mariam Attia and Julian Edge. 2017. Be(com)ing a reflexive researcher: a developmental approach to research methodology. *Open Review of Educational Research* 4, 1 (2017), 33–45. https://doi.org/10.1080/23265507.2017.1300068 arXiv:https://doi.org/10.1080/23265507.2017.1300068

[7] Madeline Balaam, Rob Comber, Rachel E. Clarke, Charles Windlin, Anna Ståhl, Kristina Höök, and Geraldine Fitzpatrick. 2019. Emotion Work in Experience-Centered Design. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300832

[8] Stephen J. Ball. 2003. The teacher's soul and the terrors of performativity. *Journal of education policy* 18, 2 (2003), 215–228. Publisher: Taylor & Francis.

[9] Tehmina N Basit. 2010. *Conducting research in educational contexts*. Bloomsbury Publishing.

[10] Gert Biesta. 2007. Why â€œwhat worksâ€ wonâ€™t work: Evidence-based practice and the democratic deficit in educational research. *Educational theory* 57, 1 (2007), 1–22. Publisher: Wiley Online Library.

[11] Gert Biesta, Ourania Filippakou, Emma Wainwright, and David Aldridge. 2019. Why educational research should not just solve problems, but should cause them as well. *British educational research journal* 45, 1 (2019), 1–4. Publisher: Wiley.

[12] Susanne Bødker. 2006. When Second Wave HCI Meets Third Wave Challenges. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (Oslo, Norway) *(NordiCHI '06)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/1182475.1182476

[13] Susanne Bødker. 2015. Third-wave HCI, 10 years later—participation and sharing. *interactions* 22, 5 (2015), 24–31.

[14] Melissa Bond, Olaf Zawacki-Richter, and Mark Nichols. 2019. Revisiting five decades of educational technology research: A content and authorship analysis of the British Journal of Educational Technology. *British journal of educational technology* 50, 1 (2019), 12–63.

[15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[16] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 18, 3 (2021), 328–352.

[17] Sally A. Brown and Donald McIntyre. 1993. *Making sense of teaching*. Open University Press, Buckingham [England] ; Philadelphia.

[18] Emeline BrulÃ©, Oussama Metatla, Katta Spiel, Ahmed Kharrufa, and Charlotte Robinson. 2019. Evaluating Technologies with and for Disabled Children. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3311757

[19] Alan Chamberlain, Andy Crabtree, Tom Rodden, Matt Jones, and Yvonne Rogers. 2012. Research in the wild: understanding'in the wild'approaches to design and development. In *Proceedings of the Designing Interactive Systems Conference*. 795–796.

[20] Joseph L Chesebro and Alexander Lyon. 2020. Instructor responses to disruptive student classroom behavior: A study of brief critical incidents. *Communication Education* 69, 2 (2020), 135–154.

[21] Sie Wai Chew, I-Ling Cheng, and Nian-Shing Chen. 2018. Exploring challenges faced by different stakeholders while implementing educational technology in classrooms through expert interviews. *Journal of Computers in Education* 5 (2018), 175–197.

[22] Rebecca Cober, Esther Tan, Jim Slotta, Hyo-Jeong So, and Karen D. KÃ¶nings. 2015. Teachers as participatory designers: two case studies with technology-enhanced learning environments. *Instructional Science* 43, 2 (March 2015), 203–228. https://doi.org/10.1007/s11251-014-9339-0

[23] Mark Coeckelbergh. 2012. *Growing moral relations: Critique of moral status ascription.* Springer.

[24] Tina Cook. 2009. The purpose of mess in action research: building rigour though a messy turn. *Educational Action Research* 17, 2 (2009), 277–291. https://doi.org/10.1080/09650790902914241 arXiv:https://doi.org/10.1080/09650790902914241

[25] Thomas D. Cook, Barbara Means, Geneva D. Haertel, and Vera Michalchik. 2003. The case for randomized experiments. *Evaluating educational technology: Effective research designs for improving learning* (2003), 15–37. Publisher: Teachers College Press New York.

[26] Diane G. Cope. 2014. Methods and Meanings: Credibility and Trustworthiness of Qualitative Research. *Oncology Nursing Forum* 41, 1 (Jan. 2014), 89–91. https://doi.org/10.1188/14.ONF.89-91

[27] Larry Cuban. 1986. *Teachers and machines: The classroom use of technology since 1920.* Teachers college press.

[28] Sarah Cuschieri. 2022. *Beyond Your Scientific Publication.* Springer International Publishing, Cham, 107–114. https://doi.org/10.1007/978-3-030-99295-8_9

[29] Manoela MO da Silva, Rafael A Roberto, Veronica Teichrieb, and Patricia S Cavalcante. 2016. Towards the development of guidelines for educational evaluation of augmented reality tools. In *2016 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)*. IEEE, 17–21.

[30] Randall S. Davies. 2011. Understanding Technology Literacy: A Framework for Evaluating Educational Technology Integration. *TechTrends* 55, 5 (Aug. 2011), 45. https://doi.org/10.1007/s11528-011-0527-3

[31] Faiza Derbel. 2017. Technology-capable teachers transitioning to technology-challenged schools. *Electronic Journal of e-learning* 15, 3 (2017), 269–280.

[32] Shiv R. Desai. 2019. Youth Participatory Action Research: The Nuts and Bolts as well as the Roses and Thorns. In *Research Methods for Social Justice and Equity in Education*, Kamden K. Strunk and Leslie Ann Locke (Eds.). Springer International Publishing, Cham, 125–135. https://doi.org/10.1007/978-3-030-05900-2_11

[33] Pierre Dillenbourg and Patrick Jermann. 2010. Technology for Classroom Orchestration. In *New Science of Learning: Cognition, Computers and Collaboration in Education*, Myint Swe Khine and Issa M. Saleh (Eds.). Springer, New York, NY, 525–552. https://doi.org/10.1007/978-1-4419-5716-0_26

[34] Triin Edovald and Camilla Nevill. 2021. Working out what works: The case of the Education Endowment Foundation in England. *ECNU Review of Education* 4, 1 (2021), 46–64. Publisher: SAGE Publications Sage UK: London, England.

[35] Peggy A Ertmer, Anne T Ottenbreit-Leftwich, Olgun Sadik, Emine Sendurur, and Polat Sendurur. 2012. Teacher beliefs and technology integration practices: A critical relationship. *Computers & education* 59, 2 (2012), 423–435.

[36] Linda Finlay. 2002. Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative research* 2, 2 (2002), 209–230.

[37] Linda Finlay. 2002. "Outing" the researcher: The provenance, process, and practice of reflexivity. *Qualitative health research* 12, 4 (2002), 531–545.

[38] Wellesley Foshay and D. Quinn. 2010. *Strategies for Evaluating Technology in Education and Training.*

[39] Gregory M Francom. 2020. Barriers to technology integration: A time-series survey study. *Journal of Research on Technology in Education* 52, 1 (2020), 1–16.

[40] Patricia Garcia and Marika Cifor. 2019. Expanding Our Reflexive Toolbox: Collaborative Possibilities for Examining Socio-Technical Systems Using Duoethnography. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 190 (nov 2019), 23 pages. https://doi.org/10.1145/3359292

[41] G.D. Haertel and B. Means. 2003. *Evaluating Educational Technology: Effective Research Designs for Improving Learning.* Teachers College Press. https://books.google.co.uk/books?id=U85YwnEGbxcC

[42] Eduardo Hariton and Joseph J. Locascio. 2018. Randomised controlled trialsâ€"the gold standard for effectiveness research. *BJOG: an international journal of obstetrics and gynaecology* 125, 13 (2018), 1716. Publisher: NIH Public Access.

[43] Steve Harrison, Deborah Tatar, and Phoebe Sengers. 2007. The three paradigms of HCI. In *Alt. Chi. Session at the SIGCHI Conference on human factors in computing systems San Jose, California, USA.* 1–18.

[44] Gillian R. Hayes. 2012. Taking action in your research. *Interactions* 19, 4 (July 2012), 50–53. https://doi.org/10.1145/2212877.2212890

[45] Walter F. Heinecke, Natalie B. Milman, Lisa A. Washington, and Laura Blasi. 2001. New directions in the evaluation of the effectiveness of educational technology. *Computers in the Schools* 18, 2-3 (2001), 97–110. Publisher: Taylor & Francis.

[46] Curtis R. Henrie, Lisa R. Halverson, and Charles R. Graham. 2015. Measuring student engagement in technology-mediated learning: A review. *Computers & Education* 90 (Dec. 2015), 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

[47] Khe Foon Hew, Min Lan, Ying Tang, Chengyuan Jia, and Chung Kwan Lo. 2019. Where is the "theory" within the field of educational technology research? *British Journal of Educational Technology* 50, 3 (2019), 956–971.

[48] Neil Humphrey, Ann Lendrum, Emma Ashworth, Kirsty Frearson, Robert Buck, and Kirstin Kerr. 2016. Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. *Education Endowment Foundation* 1 (2016).

[49] Jennifer Iriti, William Bickel, Christian Schunn, and Mary Kay Stein. 2016. Maximizing research and development resources: identifying and testing â€œload-bearing conditionsâ€ for educational technology innovations. *Educational Technology Research and Development* 64, 2 (April 2016), 245–262. https://doi.org/10.1007/s11423-015-9409-2

[50] A. Januszewski and M. Molenda. 2013. *Educational Technology: A Definition with Commentary*. Taylor & Francis. https://books.google.co.uk/books?id=JO3Yc0UuK74C

[51] Heisawn Jeong, Cindy E. Hmelo-Silver, and Kihyun Jo. 2019. Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005â€“2014. *Educational Research Review* 28 (Nov. 2019), 100284. https://doi.org/10.1016/j.edurev.2019.100284

[52] Peter Kahn, Lucy Everington, Kathleen Kelm, Iain Reid, and Francine Watkins. 2017. Understanding student engagement in online learning environments: The role of reflexivity. *Educational Technology Research and Development* 65 (2017), 203–218.

[53] Ahmed Kharrufa, Serdar Abaci, Emeline BrulÃ©, Petr Slovak, Megan Venn-Wycherley, and Daniel Lambton-Howard. 2020. Evaluation of emerging educational technologies in the classroom. In *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts (IDC '20)*. Association for Computing Machinery, New York, NY, USA, 77–84. https://doi.org/10.1145/3397617.3398065

[54] Ahmed Kharrufa, Madeline Balaam, Phil Heslop, David Leat, Paul Dolan, and Patrick Olivier. 2013. Tables in the wild: lessons learned from a large-scale multi-tabletop deployment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1021–1030. https://doi.org/10.1145/2470654.2466130

[55] Ahmed Kharrufa, Sally Rix, Timur Osadchiy, Anne Preston, and Patrick Olivier. 2017. Group Spinner: recognizing and visualizing learning in the classroom for reflection, communication, and planning. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5556–5567.

[56] Michelle E Kiger and Lara Varpio. 2020. Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical teacher* 42, 8 (2020), 846–854.

[57] Jennifer W. M. Lai and Matt Bower. 2019. How is the use of technology in education evaluated? A systematic review. *Computers & Education* 133 (May 2019), 27–42. https://doi.org/10.1016/j.compedu.2019.01.010

[58] Debora de Castro Leal, Angelika Strohmayer, and Max Krüger. 2021. On Activism and Academia: Reflecting Together and Sharing Experiences Among Critical Friends. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 303, 18 pages. https://doi.org/10.1145/3411764.3445263

[59] Susan Lechelt, Frederik Brudy, Nicolai Marquardt, and Yvonne Rogers. 2021. EvalMe: Exploring the Value of New Technologies for In Situ Evaluation of Learning Experiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445749

[60] Hee Rin Lee, Selma Šabanović, and Sonya S. Kwak. 2017. Collaborative Map Making: A Reflexive Method for Understanding Matters of Concern in Design Research. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5678–5689. https://doi.org/10.1145/3025453.3025535

[61] Deborah L. Lowther, Fethi A. Inan, J. Daniel Strahl, and Steven M. Ross. 2008. Does technology integration â€œworkâ€ when key barriers are removed? *Educational Media International* 45, 3 (Sept. 2008), 195–213. https://doi.org/10.1080/09523980802284317

[62] Keith Morrison. 2001. Randomised controlled trials for evidence-based education: some problems in judging'what works'. *Evaluation & Research in Education* 15, 2 (2001), 69–83.

[63] Sabine Mueller and Olivier Toutain. 2015. *The outward looking school and its ecosystem*. Technical Report. OECD. https://www.oecd.org/cfe/leed/Outward-Looking-School-and-Ecosystem.pdf

[64] Rebecca Nicholson, Tom Bartindale, Ahmed Kharrufa, David Kirk, and Caroline Walker-Gleaves. 2022. Participatory Design Goes to School: Co-Teaching as a Form of Co-Design for Educational Technology. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans. https://doi.org/10.1145/3491102.3517667

[65] Miguel Nussbaum and Anita Diaz. 2013. Classroom logistics: Integrating digital and non-digital resources. *Computers Education* 69 (2013), 493–495. https://doi.org/10.1016/j.compedu.2013.04.012

[66] Martin Oliver. 2000. An introduction to the evaluation of learning technology. *Educational Technology and Society* 3, 4 (2000), 20 – 30. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0005026996&partnerID=40&md5=b68659cdead23bf12c46414ece58dc26 Cited by: 88.

[67] Lyndal May O'Gorman, S Elliott, and EJ Ärlemalm-Hagsér. 2020. Stories of disruption. *Researching early childhood education for sustainability: Challenging assumptions and orthodoxies* (2020).

[68] Viktoria Pammer-Schindler, Erik Harpstead, Benjamin Xie, Betsy DiSalvo, Ahmed Kharrufa, Petr Slovak, Amy Ogan, Joseph Jay Williams, and Michael J. Lee. 2020. Learning and education in HCI: a reflection on the SIG at CHI 2019. *Interactions* 27, 5 (Sept. 2020), 6–7. https://doi.org/10.1145/3411290

[69] Stephen J Pape and Sherri K Prosser. 2018. Barriers to technology implementation in community college mathematics classrooms. *Journal of Computing in Higher Education* 30 (2018), 620–636.

[70] Laura Parson. 2019. Considering Positionality: The Ethics of Conducting Research with Marginalized Groups. In *Research Methods for Social Justice and Equity in Education*, Kamden K. Strunk and Leslie Ann Locke (Eds.). Springer International Publishing, Cham, 15–32. https://doi.org/10.1007/978-3-030-05900-2_2

[71] William R Penuel, Barry J Fishman, Britte Haugan Cheng, and Nora Sabelli. 2011. Organizing research and development at the intersection of learning, implementation, and design. *Educational researcher* 40, 7 (2011), 331–337.

[72] Suvi Pihkala and Helena Karasti. 2016. Reflexive engagement: enacting reflexivity in design and for'participation in plural'. In *Proceedings of the 14th Participatory Design Conference: Full papers-Volume 1*. 21–30.

[73] Wanda Pillow. 2003. Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *International Journal of Qualitative Studies in Education* 16, 2 (2003), 175–196. https://doi.org/10.1080/0951839032000060635 arXiv:https://doi.org/10.1080/0951839032000060635

[74] Luis P. Prieto, MarÃa JesÃºs RodrÃguez-Triana, Roberto MartÃnez-Maldonado, Yannis Dimitriadis, and Dragan GaÅ¡eviÄ‡. 2019. Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. *Australasian Journal of Educational Technology* 35, 4 (Aug. 2019). https://doi.org/10.14742/ajet.4314

[75] Barbara Probst. 2015. The eye regards itself: Benefits and challenges of reflexivity in qualitative social work research. *Social work research* 39, 1 (2015), 37–48.

[76] Yvonne Rogers. 2011. Interaction design gone wild: striving for wild theory. *interactions* 18, 4 (2011), 58–62. Publisher: ACM New York, NY, USA.

[77] Yvonne Rogers. 2022. *HCI theory: classical, modern, and contemporary*. Springer Nature.

[78] Yvonne Rogers, Kay Connelly, Lenore Tedesco, William Hazlewood, Andrew Kurtz, Robert E. Hall, Josh Hursey, and Tammy Toscos. 2007. Why Itâ€™s Worth the Hassle: The Value of In-Situ Studies When Designing Ubicomp. In *UbiComp 2007: Ubiquitous Computing*, John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang (Eds.). Springer, Berlin, Heidelberg, 336–353. https://doi.org/10.1007/978-3-540-74853-3_20

[79] Robert Rosenthal and Lenore Jacobson. 1966. Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological reports* 19, 1 (1966), 115–118. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[80] Lowther D. L. Ross S. M., Morrison G. R. 2010. Educational Technology Research Past and Present: Balancing Rigor and Relevance to Impact School Learning. *Contemporary Educational Technology* 1, 1 (July 2010), 17–35.

[81] Jaap Scheerens. 2015. School effectiveness research. In *International encyclopedia of social and behavioral sciences, 2nd edition*. Elsevier, 80–85.

[82] Neil Selwyn. 2010. Looking beyond learning: Notes towards the critical study of educational technology. *Journal of computer assisted learning* 26, 1 (2010), 65–73.

[83] Neil Selwyn. 2013. *Distrusting educational technology: Critical questions for changing times*. Routledge.

[84] Antonette Shibani, Simon Knight, and Simon Buckingham Shum. 2020. Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education* 46 (July 2020), 100730. https://doi.org/10.1016/j.iheduc.2020.100730

[85] G Shrader, K Williams, J Lachance-Whitcomb, L-E Finn, and L Gomez. 2001. *Participatory design of science curricula: The case for research for practice*. Technical Report. Seattle, WA.

[86] David Silverman. 2013. *Doing Qualitative Research: A Practical Handbook*. SAGE Publications.

[87] Lucy A. Suchman. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.

[88] Young Suzanne and Nichols Helen. 2017. A reflexive evaluation of technology-enhanced learning. *Research in Learning Technology* 25 (Nov. 2017). https://doi.org/10.25304/rlt.v25.1998

[89] Carole J Torgerson and David J Torgerson. 2012. The need for randomised controlled trials in educational research. In *Education Matters*. Routledge, 221–232.

[90] Nompilo Tshuma. 2021. The vulnerable insider: navigating power, positionality and being in educational technology research. *Learning, Media and Technology* 46, 2 (2021), 218–229. https://doi.org/10.1080/17439884.2021.1867572 arXiv:https://doi.org/10.1080/17439884.2021.1867572

[91] Megan Venn-Wycherley, Christine Bennett, and Ahmed Kharrufa. 2020. Design Studios for K-12 Computing Education. In *51st ACM Technical Symposium on Computer Science*. Newcastle University.

[92] Laton Vermette, Joanna McGrenere, Colin Birge, Adam Kelly, and Parmit K. Chilana. 2019. Freedom to Personalize My Digital Classroom: Understanding Teachers' Practices and Motivations. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300548

[93] Benjamin Xie, Erik Harpstead, Betsy DiSalvo, Petr Slovak, Ahmed Kharrufa, Michael J. Lee, Viktoria Pammer-Schindler, Amy Ogan, and Joseph Jay Williams. 2019. Learning, Education, and HCI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–4. https://doi.org/10.1145/3290607.3311761