Open Research Online
The Open University's repository of research publications
and other research outputs

An Investigation of Privacy and Utility Tensions in Learning Analytics

Thesis

oro.open.ac.uk

An Investigation of Privacy and Utility Tensions in Learning Analytics

Maureen Maina Korir

Master of Science in Information Technology

Bachelor of Science in Computer Science

Thesis submitted to The Open University

for the degree of Doctor of Philosophy in Educational Technology

Institute of Educational Technology (IET)

The Open University

May 2022

LEVERHULME
TRUST

iet

Abstract

Through learning analytics (LA), higher education institutions have put student data to various uses which aim to be beneficial for students, lecturers, and the institutions. Despite the potential benefits of LA, however, there are research gaps in understanding inherent privacy and utility tensions. Using four research studies, this thesis is an investigation of factors that contribute to these tensions.

An examination, using Delphi study techniques, of how LA experts (n=12) conceptualised privacy in LA and what they thought of as key privacy issues demonstrated a collective agreement on institutional responsibility, including to empower students to manage their privacy. As such, the findings exposed gaps between existing institutional applications of LA and the views of the experts.

A laboratory study (n=111) with follow-up semi-structured interviews (n=4) identified that students are not concerned about the use of their data for LA. However, knowing that student data could be shared with third parties evoked feelings of discomfort. The qualitative data suggested that students' privacy concepts differed from those of the LA experts, highlighting a need to operationalise LA with a shared understanding of what privacy means for stakeholders.

Using an experimental design (n=447), privacy concern was further examined through the lens of students' data use preferences. The findings suggested that participants' data use preferences were not influenced by an awareness of the possible privacy risks and benefits of data use for LA. Consequently, other factors might influence students' data use preferences. The qualitative data shed light on a "dual nature" to participants' data use preferences, suggesting both support for and reservation about the use of student data for LA, the latter due to ethics and privacy concerns. Further examination using follow-up interviews (n=15) suggested a need to align institutional data use practices with students' expectations.

Taken together, the research findings suggest that privacy in LA combines several concepts, expressed in different ways across stakeholder groups. To better understand students' perspectives of privacy in LA requires unpacking the dimensions of privacy that contribute to students' privacy concern, or lack thereof. Most importantly, while some uses of student data for LA do not concern students, other data uses might not meet their expectations. Taking steps to address these tensions will contribute towards ethical LA.

Acknowledgements

Only I could run this PhD marathon, yet I am blessed to have had a community around me to enrich my journey. My sincere thanks go to my supervisors: Dr Sharon Slade, Dr Wayne Holmes, Dr Tim Coughlan, and Professor Bart Rienties. Working with them, I was able to shape the research into what I thought it should be and their guidance helped me to tackle challenging and interesting questions.

Dr Maria Leedham, my third-party monitor, was readily available when I needed to think through how to overcome possible obstacles. My thanks also go to the faculty administrative team, including Christine Griffiths, Wendy Whiteley, and Anne Foward for their support in dealing with university requirements. I would also like to extend my gratitude to the 12 learning analytics experts who made time to participate in the first study. Dr Kyle Jones very kindly provided feedback to an early version of the study questionnaire and Dr YingFei Héliot facilitated research access for the second study. I would also like to thank all the students who took part in the research studies.

Further thanks go to Dr Carolin Decker-Lange, Anke Nink, Dr Annika Lohstroh, Maria Geissler, and Dr Thomas Ullmann who I worked with to translate one of the questionnaires for Study 3 from German to English.

My fellow PhD students at the Open University provided their support, friendship, and responded to my requests for critical analysis to test my research. I had the privilege to share the experience of being academic mums with Saman and Gosia. Dr Mojisola Erdt was a solid sounding board and friend on this journey.

I would also like to thank my friends and family who have offered companionship, laughter, and support, and made this journey enjoyable. Paul, Natalie, and Michael provided welcome distraction when I needed it. My parents and siblings have been there through the highs and lows. Mum, it gives me great pleasure to finally say that I am done.

This thesis is dedicated to the memory of my brother Eric. He didn't get to see me finish but I think he would have been proud of my persistence and determination to get to the end of it.

Declaration of Authorship

I declare that the work contained within this thesis is my own. Several sections of this thesis have been edited and published or are being prepared for publication. Publications to date and articles currently being prepared are listed below. No publications are given in their entirety; however, portions of the publications are adapted within the narrative of this thesis, particularly in Chapter 4 through to Chapter 7.

While the publications have multiple authors, I was responsible for the research conceptualisation and design, data collection, data analysis, and write-up of the work in this thesis. My academic supervisors Dr Sharon Slade (Earth Trust, UK), Dr Wayne Holmes (University College London), Dr Tim Coughlan (The Open University, UK), and Professor Bart Rienties (The Open University, UK) are credited co-authorship for their contribution in providing critique and insightful feedback to my research conceptualisation and study design, and draft articles. Similarly, Dr YingFei Héliot (University of Surrey) is credited co-authorship as she facilitated access that made the research possible and provided helpful critique on a full draft of the article.

References to Relevant Work:

Korir, M. M., Slade, S., Holmes, W., and Rienties, B. (2021). Eliciting students' data use preferences in learning analytics: A crowdsourced approach. In: Rienties, B., Hampel, R., Scanlon, E., and Whitelock, D. eds. Open World Learning: Research, Innovation and the Challenges of High-Quality Education. Routledge.

Korir, M. M., Mittelmeier, J. and Rienties, B. C. (2019). Is Mixed Methods Social Network Analysis Ethical? In: Froehlich, Dominik E.; Rehm, Martin and Rienties, Bart C. eds. Mixed Methods Approaches to Social Network Analysis. Routledge.

Korir, M. M., Slade, S., Holmes, W., Héliot, Y., and Rienties, B. (Submitted). Investigating the dimensions of students' privacy concern in the collection, use, and sharing of data for Learning Analytics. *Computers in Human Behavior Reports.*

Korir, M. M., Slade, S., Holmes, W., and Rienties, B. (Manuscript in preparation) Privacy and learning analytics in the datafied university: Is there consensus among the experts?

Korir, M. M., Slade, S., Holmes, W., Coughlan, T., and Rienties, B. (M*anuscript in preparation*). A systematic and privacy-oriented analysis of ethics and privacy frameworks for learning analytics.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

F-T-F – Face-to-face

HEI – Higher Education Institution

HREC – Human Research Ethics Committee

LA – Learning Analytics

OU – Open University, UK

VLE – Virtual Learning Environment

1   Introduction

Educational technology, in my view, is characterised by applying new and creative approaches to the ways by which educators teach and the ways students can learn. Recent advances in educational technology include the use of chatbots to support learning (Kukulska-Hulme, et al., 2021). Chatbots are computer programs that can carry out conversations using text or audio (Winkler & Söllner, 2018) and have been likened to offering support similar to a student raising their hand in a class (Wollny, et al., 2021). The application of these technologies to learning demonstrates to me the exciting possibilities at the intersection of technology and education.

Learning analytics (LA) is one other example of innovation in educational technology applying data mining and analysis techniques that were initially developed and used on organisational data to educational data (Daniel, 2015). And while learning institutions have collected and used student data for years, LA further extends what is possible, for instance, shifting the use of student data from an individual educator looking to understand students' learning patterns to inform their teaching, to faculty- and university-wide initiatives to recruit and retain students over the course of their study program. Thus, LA holds potential and promise for students, educators, and institutions.

It would seem that the potential and promise of LA applications for higher education is limited only by one's ability to imagine what can be achieved. As a learner, I can imagine using an application developed using LA principles that will prompt me to revisit material at the right time to master a topic of interest. As an educator, I think there is value in an LA application with which students can challenge themselves to be their best in their studies. And I can see how learning institutions I am and have been affiliated with in the past can leverage insights from student data to enhance their course offering for students. However, my enthusiasm and vision for the potential of LA applications is dampened by pressing ethical and privacy questions about whether there is a price to pay for such convenience, and whether the price would be too steep. As such, in my view, applications at the intersection of technology and education rightly must be accompanied by consideration of ethics and privacy. The work reported in this thesis focuses on how this has currently been achieved in the LA context and identifies how these efforts can be extended even further.

1.1 Research Motivation and Background

LA is commonly defined as "the measurement, collection, analysis and reporting of data about learners and their contexts for the purposes of understanding and optimising learning and the environments in which it occurs" (Long & Siemens, 2011, p. 34). There is growing interest in the implementation of LA applications in higher education, with an increase in the number of publications reporting on LA practices (Wong, Li, & Choi, 2018). Varied levels of interest are seen in higher education institutions (HEIs) in several countries, including the UK (Newland & Trueman, 2017; Sclater, Peasgood, & Mullan, 2016), Australia (Colvin, et al., 2016), New Zealand (Mahroeian, Daniel, & Butson, 2017), among other countries (Waheed, Hassan, Aljohani, & Wasif, 2018). The observed interest is driven by the promise and potential of LA applications for higher education, for instance, to improve student retention (Herodotou, Naydenova, Boroowa, Gilmour, & Rienties, 2020), to inform learning design (Rienties & Toetenel, 2016), and to support student learning such as through recommendation of learning resources for students (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015) as well as timely feedback and intervention (Tempelaar, Rienties, & Giesbers, 2015; Pardo, Jovanovic, Dawson, Gašević, & Mirriahi, 2019). Thus, it is often argued that there are numerous opportunities for HEIs to apply LA for the potential benefit of students and other stakeholders.

Against this background of potential for progress and development, however, lie tensions with respect to ethics and privacy in LA. Generally, ethics is concerned with evaluating whether human action is positive or negative (White, 2017). In the LA context, ethics has been defined as "the systematization of correct and incorrect behaviour in virtual spaces according to all stakeholders" (Pardo and Siemens, 2014, p.439). Early influential work in the legal domain posited that privacy was the right to be left alone (Warren & Brandeis, 1890), whereas within the context of information privacy, it can be seen as the "claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent the information about them is communicated to others" (Westin, 1967, p. 7). Related to this definition of privacy is the notion that people might want to control the information others have about them and keep that information within self-determined limits (Roessler, 2018). In the LA context, privacy has been expressed as "freedom from unauthorized intrusion: the ability of an individual or a group to seclude themselves or the information about them, and thus to express themselves selectively" (Ferguson, Hoel, Scheffel, and Drachsler, 2016, pg. 11). On the basis of these definitions, in this

thesis, LA ethics and privacy is viewed as being concerned with determining the right thing to do when using student data.

Privacy and utility tensions are evident when one analyses what is appropriate with respect to privacy when using student data. In this thesis, utility refers to the possibly beneficial uses that student data can be put to. For example, LA could lead to students thought to be at potential risk of failing their courses receiving valuable support to progress in their studies. An example is seen in a randomised control trial by Herodotou and her colleagues (2020) where there was an increasingly higher proportion of students from an intervention group at different course milestones, from the start of the courses (7%) to completion (22%), compared to students in the control group. While all participating students could receive an intervention such as follow up for not submitting an assignment, students in the intervention group additionally received a phone call from a member of the student support team to discuss their thoughts on starting the course, any concerns they may have had, and if they knew how to access support when they needed it. At the same time, there is a perceived conflict between such support and students developing their autonomy, that is, their ability to make their own decisions, including about their learning (Rubel & Jones, 2016; Johnson, 2017). For example, teaching staff who took part in a qualitative study were concerned that LA could potentially limit students' responsibility for their learning (Tsai, Perrotta, & Gašević, 2019). Students may also express a preference to determine their own academic path or want flexibility to fit their studies to their responsibilities (Silvola, Näykki, Kaveri, & Muukkonen, 2021). These preferences might not align with expected student behaviour guided by output from LA applications.

Questions have also been raised about whether students are aware of and explicitly consent to the use of their data for LA. Out of the 112 students who took part in semi-structured interviews in the work by Jones et al., (2020), none could recall consenting to the collection and use of their data. Students might emphasise how important their informed consent is (Roberts, Howell, Seaman, & Gibson, 2016), and may also request to review their consent periodically, to facilitate them modifiying their consent as needed (Adejo & Connolly, 2017). At the same time, research suggests that students are concerned about being negatively impacted by LA feedback (Roberts, Howell, Seaman, & Gibson, 2016), for example, being demotivated due to negative feedback about their performance (Knight, Brozina, & Novoselich, 2016). As these examples demonstrate, deeper insights are needed to understand the causes of these privacy and utility tensions in LA and to propose solutions to address them.

Ethics and privacy frameworks have been developed by researchers and LA experts to enhance the understanding of ethics and privacy issues arising from the use of student data (Slade & Prinsloo, 2013). They also guide HEIs to adopt LA in ethical ways that are respectful of the privacy of the students (Pardo & Siemens, 2014; Drachsler & Greller, 2016). The frameworks seek to achieve these aims by systematically identifying ethics and privacy issues and detailing how HEIs can address them (Sclater, 2016; Drachsler & Greller, 2016; Hoel & Chen, 2018). Given the prominent role of these frameworks for privacy in LA, LA experts, many of whom were involved in their development, were chosen as the first group of stakeholders for the thesis research. While there is evidence of privacy and utility tensions in LA, there have also been calls for more research on privacy in LA focusing on students and including their perspectives (Roberts, Howell, Seaman, & Gibson, 2016). Thus, this research selected students as the second group of stakeholders. The next section details the research aims and research questions.

1.2   Research Aims and Questions

1.2.1   Main Research Question

In this research, the aim was to investigate the causes of privacy and utility tensions in LA. Thus, the main research question was:

> *What contributes to privacy and utility tensions in learning analytics?*

To answer this main research question, two stakeholder groups (i.e., LA experts and students) were identified, as well as a conceptual framework and a theoretical framework. The conceptual framework considered privacy concepts, privacy concern, and data use preferences, whereas the theoretical framework consisted of three well-known theories in the privacy domain: contextual integrity (Nissenbaum, 2010), privacy calculus theory (Culnan & Armstrong, 1999; Dinev & Hart, 2006), and privacy self-management (Solove, 2013). These three elements (stakeholders, conceptual framework, and theoretical framework) are illustrated in Figure 1.1 which is a summary of the focus of the thesis research. Each element is discussed in detail in this section.
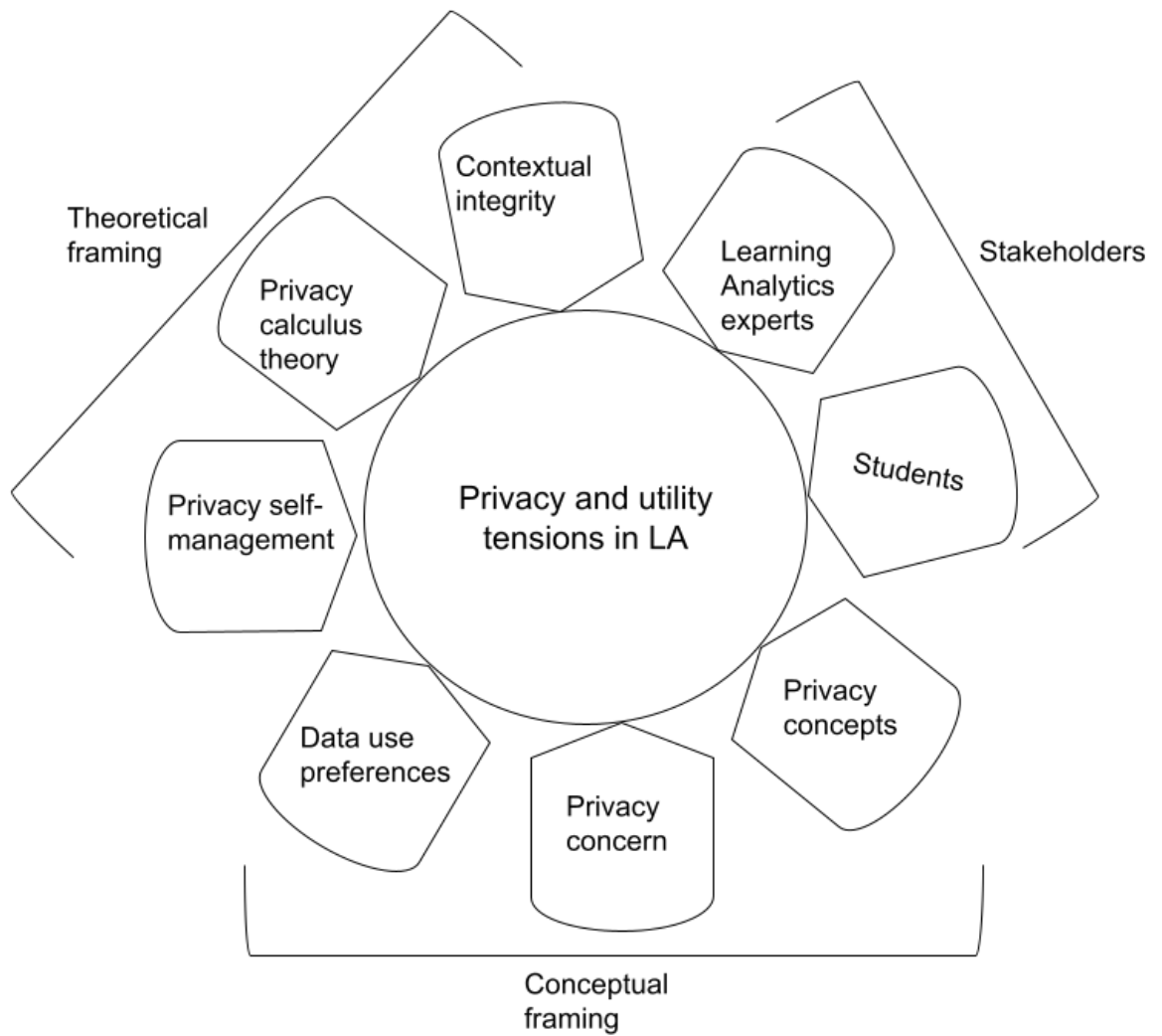
Figure 1.1: Summary of the focus of the thesis research highlighting who the stakeholders are, and the conceptual and theoretical frameworks

1.2.1.1   Stakeholders

As illustrated in Figure 1.1, the research focused on two stakeholder groups, namely, LA experts and students. LA experts have made significant contributions to how privacy is attended to in the design and development of LA and its applications in HEIs, including proposing numerous and influential ethics and privacy frameworks, for example (Greller & Drachsler, 2012; Slade & Prinsloo, 2013; Slade & Boroowa, 2014; Pardo & Siemens, 2014; Cormack, 2016; Drachsler & Greller, 2012). However, the privacy concepts of members of this stakeholder group, that is, how they think about and describe privacy, including the terms that they use, has not received much attention in the literature. Thus, this observation motivated the focus on LA experts.

In a survey with 1,647 students in the USA, Vu, Adkins and Henderson (2019) informed their study participants what data is collected, who has access to it, and how it might be used. Participants were asked three questions to determine whether they were concerned about their data being collected and used by their instructors for academic or research purposes (on a Likert scale where 1 was not really concerned and 5 was really concerned). Here, the mean total value for participants' responses was 2.23. However, it is plausible that this lack of concern might have been due to the authors making the data recipient and data uses clear to participants. Findings from a survey with 286 UK-based students (Slade, Prinsloo, & Khalil, 2019) were that students in that study accepted institutional use of their data for the specific purpose that it would benefit their learning. In contrast, in a laboratory study with 330 students in Germany (Ifenthaler & Schumacher, 2016), students were unwilling to share all the data that can be used for LA. Here, students were more willing to share course-related data for example test results on their learning strategies (78%), compared to personal data or data about their online behaviour such as their social media data (10%) and medical data (8%). Furthermore, a moderated forum discussion with 35 UK-based students (Slade & Prinsloo, 2014) identified some students' concerns about surveillance or tracking. This lack of agreement in the literature as to whether students are concerned about privacy in LA highlighted an area of privacy and utility tension that required further unpacking for in-depth insights. Therefore, students formed the second stakeholder group selected for the thesis research.

1.2.1.2   Conceptual Framework

Three organising concepts formed the conceptual framework for the thesis research. These organising concepts were *privacy concepts* (how the stakeholders thought about and described privacy and the terms they used), *privacy concern* (individuals' perspectives whether the collection and use of their data is fair (Malhotra, Kim, & Agarwal, 2004)), and *data use preferences* (which data they would choose for use in LA). Privacy concepts guided the focus of the research carried out in Study 1 (reported in Chapter 4) and Study 2 (reported in Chapter 5) with the aim of identifying and comparing LA experts' and students' privacy concepts. Students' privacy concern with the use of data for LA was investigated in Study 2 (reported in Chapter 5), while student's preferences for the use of their data was the focus of Study 3 (reported in Chapter 6) and Study 4 (reported in Chapter 7). These three organising concepts were chosen for their contribution towards addressing the main thesis research question. Each of these concepts is detailed in Section 2.3

1.2.1.3 Theoretical Framework

Three theories have guided the design and development of the studies carried out for the thesis research. They are summarised here and elaborated further in Section 2.3. (i) Contextual integrity (Nissenbaum, 2010) is an approach to thinking about privacy which considers users' expectations (referred to as norms) for the flow of information in different contexts. It considers elements such as what data is being transmitted and who will receive it to anticipate or identify how users will respond to stated or proposed information flows. Contextual integrity also recognises that users' expectations for the flow of information can change over time. (ii) Privacy calculus theory (Dinev & Hart, 2006; Culnan & Armstrong, 1999) posits that users determine whether to share personal information after weighing several factors, including the risks and benefits. (iii) With privacy self-management, users are informed by organisations about the personal data that is collected and how it is used. They have an opportunity to decide whether to accept the stated collection and use of their personal data and if they do, can go on to provide their consent (Solove, 2013).

1.2.2 Further Research Questions

Several research questions were developed to address the main research question. The research questions were split across four research studies reported in this thesis as illustrated in Figure 1.2. Figure 1.2 also aligns the conceptual and theoretical framing to the research questions in the four research studies. These research questions were developed to identify and understand the privacy and utility tensions in LA.

Main RQ:
What contributes to privacy and utility tensions in learning analytics?

Study 1	Study 2	Study 3	Study 4
RQ 1: How do learning analytics experts describe and talk about privacy in learning analytics?	**RQ 3**: To what extent are students concerned about the collection, use, and sharing of their data for learning analytics, and compared to e-commerce?	**RQ 6**: To what extent does an awareness of the possible privacy risks and benefits of data use for learning analytics contribute to a difference in students' preferences for the use of their data?	**RQ 9**: What insights can be obtained about the motivation for students changing or not changing their data use preferences?
RQ 2: What do the experts consider to be key privacy issues in learning analytics?	**RQ 4**: To what extent are students' general privacy concerns and behaviour related to their concern about the collection, use, and sharing of student data for learning analytics?	**RQ 7:** To what extent does not having access to learning analytics applications influence students' preferences for the use of their data?	
	RQ 5: What issues contribute to students' concern or lack of concern about data collection, use, and sharing for learning analytics?	**RQ 8:** What do students indicate as the motivation for their preferences for the use of their data?	

Privacy Concepts —————————————— Data Use Preferences —————

Conceptual framing ——— Privacy Concern ———

Theoretical framing ——— Contextual Integrity ——— Privacy Calculus Theory / Privacy Self Management

Figure 1.2: Summary of the thesis research questions and the conceptual and theoretical framing for the four thesis research studies

RQ 1 and RQ 2 focused on identifying how LA experts conceptualised privacy in LA and what they thought key privacy issues were. Given that the LA experts who took part in Study 1 had been involved in developing or implementing ethics and privacy frameworks for LA applications, including in higher education, and Study 1 was conducted in the early stages of this research, the LA experts' insights were valuable to extend the findings of the literature review carried out for this thesis as well as to identify a way forward for the research. More specifically, the LA experts' feedback provided further impetus to focus on students as stakeholders. Work done to address RQ 1 and RQ 2 is discussed in detail in Chapter 4.

RQ 3, RQ 4, and RQ 5 focused on teasing apart the dimensions of privacy in LA from students' perspectives, to identify those which were and those which were not of concern to students. An additional aim was to identify whether students' general privacy concern was related to their privacy concern in the LA context. Finally, the motivation for whether students were concerned about privacy in LA or not were also analysed. Work done to address RQ 3 – RQ 5 is discussed in detail in Chapter 5.

Following on from these research questions, RQ 6, RQ 7, and RQ 8 focused on identifying students' data use preferences. This was done in a context where students were given an opportunity to choose whether to participate in LA and what data could be used, and made aware of the potential privacy risks and benefits arising from the use of student data for LA. The motivation for students changing or not changing their participants' data use preferences was also of interest and therefore investigated. Work done to address RQ 6 - RQ 8 is discussed in detail in Chapter 6.

Finally, following on from RQ 6 - RQ 8, RQ 9 focused on seeking in-depth insights into the motivation for students' data use preferences. Work done to address RQ 9 is discussed in detail in Chapter 7. The next section provides an outline of the content in the remainder of this thesis.

1.3   Thesis Structure

This research seeks to identify and understand some of the issues that contribute to privacy and utility tensions in LA. Thus, this first chapter has set the stage by discussing why this topic is of interest and why it is an important topic deserving further study.

Chapter 2: Background and Literature Review

Chapter 2 is a discussion of relevant background information for the thesis topic and a review of literature on ethics and privacy in LA. The chapter contains a description and justification for the conceptual and theoretical framing of the thesis research. It also contains a summary of empirical research on students and privacy in LA. The chapter supports the identification of gaps in research which led to the research questions addressed in this thesis.

Chapter 3: Methodology

Chapter 3 has a focus on the research methodology. Beginning with the research questions identified to address the research gaps, this chapter contains a summary of the methods that researchers have used to address questions in the domain of students and privacy in LA. A mixed methods research approach is identified as suitable for addressing the research questions and the reasons for this decision are elaborated in this chapter. Furthermore, this chapter contains a justification for the design of the four research studies carried out to address the research questions. Throughout this chapter, details are provided about the steps taken to enhance the quality of the thesis research.

Chapter 4: Study 1 Methods and Results

Chapter 4 is a discussion of the method and results for Study 1. Study 1 focused on understanding how LA experts conceptualised privacy in the field. A literature review carried out for the thesis research identified numerous ethics and privacy issues that HEIs need to address when implementing LA applications. What remained unclear was which of these issues were a priority from a privacy perspective. Thus, a second objective of Study 1 was to identify what the LA experts considered as the key privacy issues. Study 1 used Delphi study techniques to interview twelve LA experts. These experts represented seven countries and had rich and diverse backgrounds in ethics, privacy, and LA as well as practical and/or academic experience.

Chapter 5: Study 2 Methods and Results

Chapter 5 is a discussion of the method and results for Study 2. Study 2 sought to clarify whether students are concerned about privacy in LA. The study focused on identifying and understanding the dimensions of students' privacy concern with the use of their data for LA. 111 students studying for a Masters' degree in Organizational Behaviour took part in a laboratory study as part of their learning. In addition, four students took part in follow-up interviews providing further in-

depth insights into their privacy concern, or lack thereof. They also discussed how they conceptualised privacy in LA. This qualitative data facilitated a comparison of privacy concepts between students and those of the LA experts identified in Study 1.

Chapter 6: Study 3 Methods and Results

Chapter 6 is a discussion of the method and results for Study 3. Study 3 investigated whether and how an awareness of the possible benefits and privacy risks of data use for LA would influence students' data use preferences. Using a survey experiment conducted over a crowdsourcing platform, 447 participants were randomly assigned to one of four groups: a privacy risks group whose participants received a privacy risks intervention, a benefits group whose participants received a benefits intervention, a privacy risks and benefits group whose participants received both the privacy risks and benefits interventions, and a control group whose participants were shown how student data can be used in a LA dashboard. Participants indicated the motivation for their data use preferences in open responses to the survey questions.

Chapter 7: Study 4 Methods and Results

Chapter 7 is a discussion of the method and results for Study 4. Study 4 sought to identify additional factors which contribute to students' data use preferences. Semi-structured follow-up interviews were carried out with 15 of the students who took part in Study 3. Participants were invited to take part in Study 4 based on whether they changed or did not change their data use preferences after interacting with the privacy risks and benefits interventions distributed in Study 3. In this way, further insights were obtained on the factors which contributed to participants changing and not changing their data use preferences in Study 3.

Chapter 8: General Discussion and Conclusions

This final chapter is a summary and synthesis of the thesis research and a discussion of the implications of each study's findings. The chapter indicates what the thesis research has contributed to knowledge and further, identifies its methodological contributions. Recommendations are provided for researchers working in ethics and privacy in LA and for administrators of HEIs who are interested in ethical LA applications. Reflecting on what the thesis research has achieved, this chapter also contains proposals for future research.

1.4 Chapter Summary

This chapter has introduced the thesis research which focused on understanding what contributes to privacy and utility tensions in LA. Several research questions were identified based on three organising concepts (the conceptual framing), namely, privacy concepts, privacy concern, and data use preferences. Four research studies were developed in order to address the research questions. A mixed methods approach was used to allow in-depth insights to be obtained. Three theories – contextual integrity, privacy calculus theory, and privacy self-management - were brought together to develop the theoretical framework, to guide the study design and data analysis process, as well as to make sense of the studies' findings. In Chapter 2, background information about the research is provided alongside a literature review to help identify the research gaps which were addressed.

2 Background and Literature Review[1]

In this chapter, an overview of LA is presented in Section 2.1 along with a discussion of the potential benefits and anticipated challenges of LA. At the same time, some of the potential benefits and anticipated challenges of LA are discussed. Section 2.2 focuses on one of the major challenges of LA, that is, considerations of ethics and privacy, and provides background information on research ethics and ethical theories. Finally, ethics and privacy concerns and ongoing debates in the literature are outlined. Ethics and privacy have long been considered in LA, indeed, since its inception (Slade & Prinsloo, 2013; Pardo & Siemens, 2014; Drachsler & Greller, 2016). In 2018, the General Data Protection Regulation (GDPR) came into force across Europe, with requirements for data controllers (those who collect data) and data processers (those who process data collected by controllers) to meet to protect the rights of data subjects (in this context – students). While other regulations had been in place, for example, the Data Protection Act (DPA) 1998 in the UK and the Family Educational Rights and Privacy Act (FERPA) in the USA, the GDPR brought about changes to how HEIs processed student data. Reeve (2017) highlights relevant changes including: i) requiring an assessment of privacy risks to individuals and determining how these would be addressed before any potentially high-risk processing of personal data, ii) explicit opt-in consent rather than implicit or opt-in by default, and iii) greater transparency about how data is used. Consequently, privacy issues faced heightened visibility and greater public awareness. Therefore, Section 2.3 provides background information about how privacy is conceptualised for this research alongside relevant privacy theories. Section 2.4 then examines insights obtained from students' perspectives of ethics and privacy in LA. Finally, Section 2.5 highlights the research gaps which are identified from the literature review and the resulting research questions that are addressed in this thesis.

2.1 An Overview of Learning Analytics

Historically, HEIs have collected and used student data to improve operations and course delivery (Siemens, 2013), and for research purposes (Griffiths, 2017). They have been associated with measuring and recording information about students and courses and related administrative tasks, and this has increased significantly over time (Selwyn, Henderson, & Chao, 2018; Hakimi,

---

[1] Parts of Section 2.2.2 in this chapter were published in the following book chapter:
Korir, M., Mittelmeier, J., & Rienties, B. C. (2019) Is mixed methods social network analysis ethical? In *Mixed Methods Social Network Analysis* (pp. 206 - 218). Routledge.

Eynon, & Murphy, 2021). Even as HEIs have offered courses online through e-learning technologies, more student data is being converted into digital form and more teaching practices are carried out using various software tools (Williamson, 2017). Thus, even more data is available, some of it demonstrating whether and how students interact with their course and course material through virtual learning environments.

The process by which HEIs use the vast amounts of student data to improve teaching and learning is referred to as learning analytics (LA). An accepted definition in the field is that it involves the

> *"measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs"* (Long & Siemens, 2011, p. 34).

This definition is noted to emphasise the source of the data and the purposes for which the data is collected. For simplicity, this thesis will refer to the collection and use of student data when referring to LA, in this way, acknowledging the varying uses the data can be put to, and which this thesis research focuses on.

Since the first international conference on LA was held in 2011, the field has continued to develop with journals (e.g., Journal of Learning Analytics), conferences (e.g., Learning Analytics and Knowledge Conference), and societies (e.g., Society for Learning Analytics Research) disseminating research findings for use in academia and practice.

With the use of computing and educational technology to facilitate teaching and learning in higher education, even more opportunities have become available to collect data about students as they leave digital traces behind them (Hakimi, Eynon, & Murphy, 2021). The digital traces from students include number of clicks, time spent on the site, number of videos viewed, and number of forum posts (Daniel, 2015; Ho, 2017). Such digital trace data can be collected at various stages, for example, when prospective students express their interest in a university and before they submit their application, once they go through the application process and enrol as students, through to their use of online learning resources, their access to campus facilities, and their use of library resources. Once collected, student data can be aggregated and analysed using various statistical methods to make predictions, analyse relationships between individuals, and analyse links between different data sets (Jones, 2019). The insights obtained are used in various ways with the aim of improving teaching and learning. For example, LA has been used to demonstrate the contribution that communication activities in learning design make to students' completing

and passing a module (Rienties & Toetenel, 2016), providing early warning signals to tutors about students' performance (Aguilar, Lonn, & Teasley, 2014; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014), and supporting students through recommended learning resources (Bodily, Ikahihifo, Mackley, & Graham, 2018).

Use of student data has also been influenced by changes in UK HEIs in recent years. There have been cuts to government funding for free university education and a shift to a tuition fee paying model, with courses becoming commodities, leading students to be perceived, or to perceive themselves as customers (Canning, 2017). Consequently, UK HEIs have sought to further use student data to survive these changes and determine how to use available resources (Daniel, 2017). This practice is referred to as academic analytics and is distinguished from LA as it focuses on institutional needs including making improvements to student recruitment and retention (Sclater, 2016).

The next section discusses some of the potential benefits and anticipated challenges of LA.

2.1.1   Potential Benefits and Anticipated Challenges

HEIs look to implement LA applications to improve students' academic performance, increase the number of students who complete their courses within the registration period, and enhance students' engagement with the courses and course materials (Foster & Francis, 2020). To identify actionable insights from student data, LA involves various activities such as prediction modelling to predict student behaviour and cluster analysis to group similar students together based on their learning (Viberg, Hatakka, Bälter, & Mavroudi, 2018). It has been argued that LA can be used to support learners in their learning, support lecturers and student support teams to identify students in need of support, improve teaching practices and contribute to knowledge development in educational research (Hommel, Egetenmeier, & Maier, 2019).

With HEIs at various stages of implementing LA university-wide (Tsai, Whitelock-Wainwright, & Gašević, 2020), various challenges have been noted (Daniel, 2015), one of which is the acceptance of LA applications by teachers. In a workshop with 95 teaching staff, Rienties and colleagues (2018) found that teaching staff were sceptical about how easy the LA applications would be to use and desired training and further support. A second challenge involves the costs associated with collecting and storing data and mining it, as it can be time consuming and complex. Additionally, previously disparate data systems have to be configured to work together (Daniel & Butson, 2013) and skilled staff are needed (Ferguson, et al., 2014) with multiple areas of expertise

required for LA projects (Siemens, 2013). Another prominent challenge focuses on ethics and privacy in the LA context. This is discussed in the next section.

2.2   Ethics and Privacy in Learning Analytics

With the institutional collection and use of student data, there are prominent challenges related to ethics and privacy (Ferguson, 2012; Siemens, 2013; Slade & Prinsloo, 2013; Pardo & Siemens, 2014). Ethics is concerned with doing the right thing (Sax, 2018), analysing actions for their potential benefit and harm to others (Herschel & Miori, 2017, p. 33). There are numerous definitions of ethics in the LA context. It has been described as a "system of fundamental principles and universal values of right conduct" (Ifenthaler & Schumacher, 2016, p. 924), and as "the systematization of correct and incorrect behaviour in virtual spaces according to all stakeholders" (Pardo & Siemens, 2014, p. 439). Both definitions focus on ethics as seeking to identify the right course of action, while the latter also emphasises a systematic approach and introduces stakeholders' perspectives, suggesting that there may be varied views as to what is correct and incorrect. Thus, consensus among different stakeholders might be needed to address ethical challenges.

Regarding a definition of privacy, Warren and Brandeis (1890) first conceptualised privacy as a right to be left alone. They were motivated by potential infringements on privacy brought about by technological developments with the availability of small and cheap cameras to the public. While privacy has gone on to be conceptualised and defined in different ways over the years, the (potential) influence of technological developments on privacy has remained and, indeed, grown. In the dimension of information privacy, one of the earliest definitions of privacy is that it is the "claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent the information about them is communicated to others" (Westin, 1967, p. 7). This definition, as stated previously, highlights that users may want control over the information others have about them (Roessler, 2018).

Numerous definitions of privacy are also found in LA. One such definition is that it is the "regulation of how personal digital information is being observed by the self or distributed to other observers" (Pardo & Siemens, 2014, p. 438). Additionally, it has been defined as "freedom from unauthorised intrusion: the ability of an individual or a group to seclude themselves or the information about them, and thus to express themselves selectively" (Ferguson, Hoel, Scheffel, &

Drachsler, 2016, p. 11). These privacy definitions emphasise information, an audience, and the ability to control what is shared with the audience.

Accompanying definitions of privacy are claims about why it is valuable. Privacy's value is seen to lie in its ability to promote positive relationships and autonomy (Rubel & Jones, 2016). For example, a professional relationship might be better maintained where personal information is kept private, or selectively shared, e.g., one provides information on shared hobbies to enable the development of relationships rather than political preferences which are potentially divisive. Additionally, privacy supports people making decisions according to what they value or consider important, without outside interference.

Privacy is conceptualised differently by individuals and these differences can be based on their experiences, background, and culture (Norberg, Horne, & Horne, 2007; Cho, Rivera-Sánchez, & Lim, 2009). As a result, a specific focus on privacy relevant for the thesis research was developed and is detailed in Section 2.3.

The next section briefly discusses ethical theories that offer a way to think about ethical dilemmas.

2.2.1   Ethical Theories

Apart from the ethical principles, ethical theories provide a structured way to evaluate human action to help determine the right course of action (White, 2017). An action can be evaluated based on the results or consequences, and whether the action is (in)appropriate. In the results-oriented approach to ethics, also referred to as consequentialist ethics (West, Huijser, & Heath, 2016), an action is justified by the outcome and the aim is to maximise the good over any harm. Ethical decisions are based on the outcome or the consequences of the actions and an action is ethical if it results in maximising the benefit for an individual or for the wider society (Wiles, 2013; Brooks, te Riele, & Maguire, 2014). As an example, HEIs can implement LA applications to benefit a large majority of students, while potentially disadvantaging a few. However, there are challenges with the results-oriented approach, including questions surrounding how 'good' and 'benefit' are defined and measured (Brooks, te Riele, & Maguire, 2014), and that an action can be judged as right only on the basis that its benefits outweigh its harms (White, 2017).

Due to these challenges, the act-oriented approach to ethics (White, 2017), also referred to as deontological ethics (West, Huijser, & Heath, 2016) evaluates actions based on the idea that there are appropriate ways to treat people based on an equal respect for persons (Brooks, te Riele, &

Maguire, 2014). There is an expectation that there is a duty to treat people with dignity and respect because they are human. However, the act-oriented approach to ethics fails to address conflicts in one's obligations and has been criticised for its emphasis on people as rational beings and minimising the influence of emotion and not accounting for what leads people to respond differently to ethical issues. Thus, its principles have been extended to suggest necessary duties such as beneficence (doing good), non-maleficence (not causing harm), and justice, among others (Brooks, te Riele, & Maguire, 2014).

A third way to determine the right course of action uses virtue ethics, in which the person's character is considered rather than principles, rules, or consequences of an action. The question that is asked with virtue ethics is not what one should *do*, rather what kind of person should one (aim to) *be* (Brooks, te Riele, & Maguire, 2014).

A fourth approach – the ethics of justice - focuses on equal distribution of resources (West, Huijser, & Heath, 2016) (that is, distributive justice), while an extension to this concept assesses social structures and relationships based on the extent to which they facilitate self-development and self-determination (Johnson, 2017).

Finally, a fifth approach to determine the right course of action - the ethics of care - emphasises involving others, seeking harmony in relationships, and considering the needs of others (Botes, 2000).

While these ethical theories offer a way to think about ethical dilemmas, different criteria can be used therefore different results and decisions can be reached on what the right thing to do is (Wiles, 2013). These theories have found support in the LA context, including virtue ethics (Kitto & Knight, 2019) and both justice and care ethics (Prinsloo & Slade, 2017).

To understand why ethics is important for LA, relevant background information is provided on research ethics and ethics principles.

2.2.2 Research Ethics and Ethics Principles

It remains open to debate whether LA is categorised as research involving human participants in the tradition of medical research, or as operational research conducted as part of HEIs' business to enhance effectiveness and to develop strategy (Griffiths, 2017; Kitto & Knight, 2019). Research ethics providing guidelines for research with human participants has its foundations in medical research ethics (National Commission for the Protection of Human Subjects of Biomedical and

Behavioral Research, 1978; World Medical Association, 2013) with codes and guidelines developed to communicate ethical research practices. One of the earliest codes was the Nuremberg code, which was developed in 1947 during the trial of Nazi doctors (Mandal, Acharya, & Parija, 2011; Shuster, 1997). The code had 10 points which are summarised below, based on Shuster (1997):

1. The researcher should obtain the participant's voluntary consent.

2. The experiment's results should be for the good of society.

3. The experiment should be based on previous knowledge.

4. The experiment should avoid harm or injury to the participants.

5. No experiment should be conducted where it is known that death or injury will occur.

6. The degree of risk should not exceed the importance of the problem to humanity.

7. Provisions should be made to minimise harm to the participants.

8. The experiment should be conducted by qualified persons.

9. The participant can withdraw from the experiment at any time.

10. The researcher can terminate the experiment at any time if there is risk of harm or injury to the participants.

The Nuremberg code emphasised the need for participants' voluntary and informed consent, the need to minimise harm or injury to participants, that there are societal benefits to the research and that these should be greater than the risks to society or participants, and finally, that participants have a right to withdraw from the research without facing any consequences.

Following on from the Nuremberg code, the Helsinki declaration was developed by the World Medical Association in 1964 and has been updated several times since then (World Medical Association, 2013). The Helsinki declaration includes the recommendation that research protocols should be reviewed by an independent committee before the research is carried out. The Belmont report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978) was developed in response to the Tuskegee syphilis study in which African American males were recruited, without their consent, to take part in a study on how syphilis progressed. Those who had the disease were left untreated even after treatment was

identified. Therefore, the Belmont report contained three basic principles (Mandal, Acharya, & Parija, 2011):

1. Respect for persons.

2. Beneficence, meaning to minimise harm and maximise benefits to participants.

3. Justice, where the risks and benefits of the research must be distributed fairly.

Numerous ethics codes and guidelines have been proposed since these initial codes. Present-day ethics codes and guidelines have built on or extended these early principles, highlighting their relevance and importance in the conduct of ethical research. Examples of present-day ethics codes and guidelines include the Ethical Guidelines for Educational Research from the British Educational Research Association (BERA), the British Psychology Society (BPS) Code of Ethics and Conduct, and the Ethical Guidelines for Internet Research from the Association of Internet Researchers (AoIR). These ethics codes and guidelines provide principles for researchers to use. As these codes and guidelines have related foundations, this thesis research followed the guidelines from the British Psychology Society (BPS) which were relevant for the studies reported in Chapter 4 to Chapter 7.

The next section discusses ethics and privacy concerns and ongoing debates in LA.

2.2.3   Ethics and Privacy Concerns and Debates in Learning Analytics

The ethics and privacy of LA is an area with ongoing debate and discussion. Through the Learning Analytics Community Exchange (LACE), Griffiths and other colleagues (2016) emphasised the need for careful implementation of LA applications and highlighted the supporting role of ethics and privacy guidelines as well as the need to infuse relevant values in the design of these applications. Among several examples, the authors cited the inBloom case (Singer, 2014; Bulger, McCormick, & Pitcan, 2017). inBloom received $100 million initial funding with a vision to develop a centralised software system for schools to allow data sharing as well as provide learning applications and curricula. Privacy advocates, teachers and parents were excluded from the development of the system, and as such, the ethics and privacy concerns - including questions on misuse of student data, re-use of student data for commercial interests, and questions on secure storage of data - were raised at later stages and proved insurmountable. As a result, inBloom was shut down in 2014. Despite involving data collection and use at schools, the case of inBloom holds valuable lessons for LA in higher education as well.

Ethics and privacy has also been framed as enabling the development of LA, with the argument being made that once these issues are addressed sufficiently, then they can result in improved opportunities for further progress in LA applications (Gasevic, Dawson, & Jovanovic, 2016).

It has been argued that achieving ethical LA requires greater involvement of students. At present decisions around LA implementation may not include students and the design and development of LA applications are driven by institutional actors rather than students. Moving forward, students can be included in the decision-making process around LA applications, as well as providing input to determine LA features to support their learning (West, Luzeckyj, Toohey, Vanderlelie, & Searle, 2020).

Scholes (2016) examines the practice of intervening at the individual level based on information about group risk (p. 941). The author identifies several intervention practices from related work, including restricting the level of a course a student can study, restricting the number of courses a student can study, requiring pre-requisites such as the student taking a bridging course, offering support such as phone calls to encourage participation, referring a student to academic support services, and discouraging a student from continuing with study. The author's analysis identifies that LA contributes to discrimination to the extent that it facilitates students being treated not as individuals but rather as members of some group, based on features of other students which excludes the student's past choice and planned future decisions. As a solution, the author advocates for LA to integrate individual agency to account for the student's effort and achievement more effectively than other characteristics such as age, ethnicity and gender can.

Willis III (2014) and Braunack-Mayer, Street, Tooher, Feng and Scharling-Gamba (2020) advance the argument that ethics has not kept pace with innovations in LA. Willis gives the example that ethics in LA seeks to do what causes the greatest good for most people (p. 2) (refer to Section 2.2.1 for a discussion of various ethical theories). To keep pace with innovation since it might be difficult to determine the greatest good, Willis III proposes that questions are asked to identify tensions in LA and effort made to consider how the tensions can be addressed. He suggests asking questions related to identifying the very best outcome, which he refers to as utopia. Where the best outcome cannot be determined, he refers to taking action where it is allowed (for example, if the action in question is not prohibited legally) while carefully observing the outcomes for any negative impact and assessing for alignment with values. To handle unknown outcomes, for

example, he suggests determining if institutions are obliged to act and how to address unexpected results.

Criticisms levelled against LA based on ethical and privacy issues have also come from outside the field itself. Selwyn (2019) detailed challenges inherent in LA focusing on what the applications are intended to do compared to what they actually result in (its consequences) and the dominant ideas in LA (its values) and what they result in. A number of these challenges are specifically of interest for this literature review, given that they focus on ethics and privacy. For example, from the ethics perspective, Selwyn points out that LA fails to capture data that adequately represents the complexity of students' lives. Additionally, he argues that it reduces students' capacity to make decisions for themselves by removing opportunities for them to exercise their judgement, leading students to work only to what they will be measured against and to ignore other options that might be beneficial to their learning. With respect to privacy issues, Selwyn cautions that LA can become indistinguishable from surveillance given that students are "supervised" (p. 13). Similar to Mai (2016), Selwyn also points to the ability that LA affords to learn new information about students that they have not chosen to reveal.

In response to Selwyn, Ferguson (2019) acknowledges that some of the issues raised remain open including opportunities for students to control how they are seen, maintain separation between their private and public personas, and avoid constant monitoring.

The next section conceptualises privacy for the purposes of the thesis research and introduces relevant privacy theories that were used.

2.3   Conceptual and Theoretical Framing of Privacy

Given that this thesis research focused on identifying and understanding what contributes to privacy and utility tensions in LA (as discussed in Chapter 1), the following three privacy-related issues were deemed relevant and therefore selected as the conceptual framing for the thesis research: privacy concepts, privacy concern, and students' data use preferences. Each of these is described in turn in this section. In doing so, a greater understanding of privacy and a definition of privacy for the thesis research is identified. While this section briefly describes these three privacy-related issues, Section 2.4 examines their treatment in the LA literature.

2.3.1   Privacy Concepts

A privacy concept is taken as an 'abstract mental picture of what privacy is and what makes it unique and distinct' (Solove, 2008, p.13). As such, for the thesis research, privacy concepts were seen as the way stakeholders (in the case of the thesis research, this refers to both LA experts and students) think about or describe privacy. While there is a comparative lack of literature on stakeholders' privacy concepts in LA, some examples are found in adjacent fields of research. For example, research suggests that users hold multiple mental models, and draw from these mental models at different times, and sometimes simultaneously (Prettyman, Furman, Theofanos, & Stanton, 2015). Prettyman and colleagues were interested in mental pictures (as defined above) of online privacy to design effective training materials. Their work suggests that users can hold multiple privacy concepts often at the same time. For example, while embracing technological progress, participants were suspicious of new technology and its impact on their privacy. Furthermore, participants held the view that privacy is no longer available, yet went on to try and protect their privacy, e.g., by using passwords. The authors found that these mental models were often partially formed and contradicted one another.

Focusing on privacy concepts proved useful to identify tensions between how the LA experts and students thought about privacy. The rest of this section details common ways privacy is conceptualised. As discussed in Section 2.2, privacy has numerous definitions across a variety of fields. Solove (2009, p. 13) describes a "traditional" approach to defining privacy where the aim is to distinguish privacy from other things, and proposes an alternative approach to defining privacy – a "pluralistic" way (p. 40), which brings together concepts that are related in order to define privacy. This section contains both approaches and takes a first step to define privacy for the thesis research.

2.3.1.1   *Contextual Integrity*

Contextual integrity (Nissenbaum, 2010) is used as part of the theoretical foundation for Study 2 (discussed in Chapter 5) and Study 3 (discussed in Chapter 6). It is detailed in this section as relevant background information to understand the focus of the thesis research.

Contextual integrity (Nissenbaum, 2010) is an approach to understanding privacy and identifying possible privacy violations in the use of individuals' information. Nissenbaum argues that social life is governed by norms of information flow, that is, what type of information is passed on from one entity to another, and under what conditions. These norms are identified from various

sources including culture, law, history, and convention among others. Nissenbaum's work identifies two norms: of appropriateness and of flow. Norms of appropriateness govern what information about a person can or cannot be revealed in a given context, for example, one might feel free to talk about politics with their immediate family and close friends but not their colleagues. Norms of flow govern the movement of information from one party to another, for example, one can tell their doctor things about their health status and not expect this information to be shared with others (apart from other health professionals). Contextual integrity is used to identify when privacy is breached and to understand why this is the case: it would be violated if either the norms of appropriateness or flow are breached. To identify privacy violations, contextual integrity identifies: i) the context, for example, where data is collected and where it is used, ii) the actors involved, namely, the senders and recipients of information and the information subjects, iii) the attributes or information types, and iv) the transmission principles guiding the flow of information between different actors.

Researchers have applied contextual integrity to LA in various ways. Heath (2014) used it to analyse data use scenarios for LA and identified where potential privacy violations could arise. It has also been used in empirical literature on privacy in LA, primarily to explain the research findings. For example, participants in the work of Tsai, Whitelock-Wainwright, and Gašević (2020) were found to conceptualise privacy as contextual integrity. To determine the appropriateness of data sharing, the participants in the study considered the data that was to be shared, who was involved (e.g., the tutor and the student) and the type of relationship between the tutor and student. Research findings (Jones, et al., 2020) also suggest that students are unaware of institutional use of their data and are uncertain about providing consent for the same. The authors call for effort to re-align institutional data use practices with students' expectations to re-establish contextual integrity. Ifenthaler and Schumacher's work (2016) also supports contextual integrity in the LA context as students in their study did not want data which had been freely shared in one context (social media) then used in a different context (LA). In this thesis, contextual integrity was applied to identify the norms of appropriateness and flow held by LA experts and students and to better understand the study results.

2.3.1.2 *Privacy as a Right and Freedom to Act*

The origins of the perspective of privacy as a right can be traced to Warren and Brandeis's (1890) definition of privacy as "the right to be left alone" (Smith, Dinev, & Xu, 2011, p. 994). Human rights documents, such as the Universal Declaration of Human Rights (UN General Assembly,

1948) also provide for the right to privacy. Additionally, privacy as a right is viewed as "positive freedom" (van der Sloot, 2018, p. 82), as one is free to pursue chosen activities (certainly constrained by regulation, social norms, and other relevant limits).

Pardo and Siemens (2014) argue that it is hard to capture what the right to privacy means for LA applications. In their analysis of student privacy in LA, Rubel and Jones (2016, p. 153) concur with the view of privacy as a right in the LA context, arguing that efforts should be made to avoid privacy losses irrespective of the supposed benefits which may result for students from the use of LA applications.

2.3.1.3 Privacy as Control over Personal Information

This conceptualisation of privacy considers that privacy is the "claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1967). Whitley (2009, p. 155) contends that in an age of analytics, consideration should also be given to possibilities not only to limit the data that others can access, but also to control further uses of the data, what Cormack (2016) refers to as downstream uses of data.

2.3.1.4 Taxonomy of Privacy Harms

The taxonomy of privacy harms by Solove (2009) identifies privacy problems to better understand privacy. This approach differs from those discussed above as there is room for multiple concepts to be considered. The taxonomy focuses on activities that can and do create privacy problems. It identifies problems that affect a private matter but does not define what a private matter is, with the author arguing that these vary across different cultures. This section summarises the taxonomy of privacy harms in Table 2.1 based on Solove (2006; 2009).

Table 2.1: Summary of the taxonomy of privacy

Category	Privacy breach	Definition	Example of possible privacy harm in the learning analytics context
Information collection		Possible privacy harm due to gathering of information	
	Surveillance	Observe students and their activities whether they are aware of it or not	Make students uncomfortable or cause them to change their behaviour, including while unaware of any surveillance
	Interrogation	Question students and pressure them to divulge information	Students concerned about how them not wanting to share information is perceived by others
Information processing		Use, store, or process data that has already been collected	

	Aggregation	Combine separate pieces of information about a person to form a partial picture about them	Use the disconnected and incomplete information in the profile to make decisions about students
	Identification	Link a digital profile to a person in the real world	Tie students to their past behaviour limiting their self-development
	Insecurity	Risks from the handling of students' information	Information collected from students facilitating identity theft
	Secondary use	Use data collected for one purpose for other purposes, without students' consent	Students fearful and uncertain about how their information can be used in the future
	Exclusion	Students are not informed about how information about them is used and have no influence on this	Sense of vulnerability and uncertainty in students especially where this involves important decisions
Information dissemination		Where information about students is shared with others possibly resulting in privacy harms	

Breach of confidentiality	Students' information, which they did not expect to be shared or made public, is shared with others	Potential future employers learn that students consulted with mental health services during their studies
Disclosure	True information about students is revealed to others and spreads beyond expected boundaries	Damage to students' reputation
Exposure	Revealing to others certain physical and emotional attributes about a person involving their health or body	Embarrass or humiliate students
Increased accessibility	Increased access to information such as digitising publicly available records	Greater possibility of disclosure
Blackmail	Where a student is coerced to cooperate with a blackmailer who threatens to expose the student's personal secrets	Create a power relationship where the student is dominated or controlled by another person

28

	Appropriation	One takes and uses the name or likeness of another person for their benefit	Limit to students' freedom and self-development as they lose control over how they are presented to society
	Distortion	Manipulating how a student is perceived and judged by others	Lead to students being embarrassed, humiliated, stigmatised or suffering reputational harm
Invasion		Incursions into people's private affairs	
	Intrusion	Interfere with one's life	Interrupt or interfere with students' activities due to the presence or activities of another
	Decisional interference	Interfere with certain decisions in people's lives	Limit students' autonomy and liberty

Adapted from Solove (2006; 2009)

As can be seen in Table 2.1, the taxonomy (Solove, 2006; 2009) contains four basic categories of potentially harmful activities, namely 1) information collection, 2) information processing, 3) information dissemination, and 4) invasions. Each category contains related sub-categories of harmful activities. The taxonomy identifies harmful activities that may occur when information is collected from the data subject (in this case the student), processed (stored, combined, manipulated, searched, and used) by the data holders, and disseminated to others. With these three categories, information moves further away from the individual's control. With activities in the fourth category – invasion – the intrusions progress towards the individual and may not involve information. While the taxonomy identifies possible harms, it is important to point out that there are legitimate uses for use of student data, thus, for example, not all information collection will result in intrusion or undesired outcomes. However, in an educational context, collection of information from social media may lead to undesired outcomes.

For the thesis research, the taxonomy of privacy harms (Solove, 2006; 2009) was used to provide a clearer understanding of privacy. In particular, the categories of information processing and information dissemination were relevant for the thesis research.

2.3.2   Privacy Concern

Given that privacy is a complex concept (Solove, 2009) which has been defined and studied in different ways across a variety of fields, researchers have turned their attention to proxies to better understand privacy (Xu, Dinev, Smith, & Hart, 2011). One measure that is used as a proxy for privacy is *privacy concern* (Smith, Dinev, & Xu, 2011), which can be defined as ``the desire to keep personal information away from others'' (Buchanan, Paine, Joinson, & Reips, 2007). In the context of privacy and privacy concern, personal information, for example a name or an IP address, is that which ``relates to an identified or identifiable individual'', and that allows them to be ``distinguished from another individual directly from the information in question, or indirectly from that information in combination with other information.''[2] Privacy concern focuses on individuals' opinions whether the collection and use of their personal data is fair (Malhotra, Kim, & Agarwal, 2004, p. 337).

---

[2]   https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/

Studies that have used privacy concern as a proxy to understand privacy have found individual differences in how users share personal information and used these differences to categorise users according to their privacy concern (Sheehan, 2002; Taylor, 2003). Additionally, users' privacy concerns have been shown to influence their online interaction, for example, to determine the extent to which they are willing to share information (Olson, Grudin, & Horvitz, 2005). These studies have also shown that users' stated privacy concerns and attitudes can differ from their privacy-related behaviour, a phenomenon referred to as the privacy paradox (Acquisti & Gros&klags, 2005; Norberg, Horne, & Horne, 2007). For example, users indicate that they are concerned about privacy but go on to use social media or other online services which compromise the privacy of their data. Kokolakis (2017) identifies several explanations for the privacy paradox, including that users disclose information where possible gains outweigh possible losses (Dinev & Hart, 2006; Xu, Dinev, Smith, & Hart, 2011; Jiang, Heng, & Choi, 2013), where users may not be free to make decisions about information sharing (Zafeiropoulou, Millard, Webber, & O'Hara, 2013), where their decisions are affected by biases and heuristics (Acquisti & Grossklags, 2005), or where they are not aware of or do not understand the potential dangers. As a result of the work summarised in Kokolakis (2017) and other researchers, the relevance of the privacy paradox to explain the disparity between privacy attitude and behaviour is contested.

2.3.3   Students' Data Use Preferences

In the course of carrying out the research for this thesis, the understanding of privacy continued to develop, not only to encompass the perspective of privacy concepts and privacy concern discussed in the sections above, but to also investigate students' data use preferences. Students' data use preferences referred to a hypothetical situation where students were made aware of the use of student data for LA and given an opportunity to determine whether to participate and which data items they would be willing to have used for LA. The types of data that can be used for LA include data about the student including their demographic data such as age, gender, and previous education, and data about their activities in a VLE including details about the number of times they access a learning resource and for how long and what sections of the resource they engage with (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015; Joksimović, Kovanović, & Dawson, 2019). Therefore, students' data use preferences were chosen as an alternative organising concept to better understand students' privacy concerns or lack thereof.

The idea that was explored in the thesis was that students' data use preferences and how they expressed whether and how their data could be used for LA, could offer some insight into

whether they were concerned about privacy in LA. This is based on a meta-analysis of privacy research (Baruh, Secinti, & Cemalcilar, 2017) in which studies suggested that users expressed an intention to share less information where they had high privacy concern (no. of studies: 37; average r: -0.21; total N: 13,934) and shared less information where they had high privacy concern (no. of studies: 37; average r: -0.12; total N: 18,249).

An additional reason to focus on students' data use preferences was that in HEIs, students might have no choice but to provide their data in order to enrol as students at an institution and participate in various learning-related activities. Therefore, rather than focusing on their comfort with the *collection* of their data, the focus on students' data use preferences aimed to pay more attention to students' comfort with the different *uses* their data can be put to for LA.

2.3.4   A Definition of Privacy for the Thesis Research

For this thesis, the understanding of privacy has developed over the course of carrying out the four studies reported herein. While a unitary concept of privacy was held when the research begun, the understanding that has developed over time is that privacy in LA is best understood as a cluster concept, bringing together multiple perspectives on privacy. In addition, as observed while reviewing the literature on ethics and privacy frameworks for LA, the understanding of privacy also integrates concepts of transparency, informed consent and student involvement and agency.

Transparency: With transparency, the expectation is that stakeholders have "access to the description of how the analytics process is carried out and should be informed of the type of information that is being collected, including how it is collected, stored and processed" (Pardo & Siemens, 2014, p. 445). Research has demonstrated that transparency, that is, informing uses about who is collecting data and what data is being collected, can reduce privacy concern (Oulasvirta, Suomalainen, Hamari, Lampinen, & Karvonen, 2014).

Informed Consent: With its origins in medicine, informed consent considered that users are informed about some medical solution or intervention and can state that they understand the risks before a procedure is carried out (Prinsloo & Slade, 2018). However, informed consent is different in the digital context because even as users consent to the use of their data, they are not (and in some cases cannot be) given the full picture regarding who is collecting the data and how it will be used (Kitchin, 2013). It is expected that students are informed about how information about them will be used and are given an opportunity to agree, that is consent, to the stated

practices (Jones, 2019). However, students' ability to provide truly "informed" consent has repeatedly been called into question (Prinsloo & Slade, 2015) as research suggests that it is time-consuming for users to read the privacy notices on all websites that they visit (McDonald & Cranor, 2008).

Student Involvement or Agency: Involving students in decisions about whether and how their data is used is linked to empowering them (Tsai, Perrotta, & Gašević, 2019), and is thought to promote students' autonomy (Jones, 2019). It involves making students aware that their data is used, and how (which as defined above is transparency), as well as involving them as active participants in this process (Kruse & Pongsajapan, 2012; Slade & Prinsloo, 2013; West, Luzeckyj, Toohey, Vanderlelie, & Searle, 2020).

2.3.5   Privacy Calculus Theory

According to the privacy calculus theory, which was first proposed by Laufer and Wolfe (1977), individuals' intention to disclose or share personal information results from a calculus of behaviour (that is, the privacy calculus) where individuals weigh multiple factors (Li, 2012). A privacy risks and benefit analysis is a common example of a privacy calculus, where individuals weigh the privacy risks and benefits of sharing personal information to determine whether to share personal data. Individuals might put up with a loss of privacy if they can expect some benefit in return (Kokolakis, 2017; Culnan & Armstrong, 1999).

In the LA literature, Ifenthaler and Schumacher (2016) propose a privacy calculus model, demonstrating students' decision-making process to determine whether or not to share their data with LA applications. Additionally, Slade, Khalil and Prinsloo (2019) found that students in their study considered issues such as the value of services received in their decisions to determine their preferences. It is noted, however, that people are rarely rational (Acquisti & Grossklags, 2005), and may instead respond to immediate benefits without considering any long-term ramifications (whether positive or negative).

2.3.6   Privacy Self-Management

Privacy self-management refers to the practice of notice, choice and consent, where users are informed by organisations about the data collected and how it is used (notice), have an opportunity to decide whether to accept the stated collection and use of their data (choice), and if they do, can go on to provide their consent (Solove, 2013).

Solove (2013) points out several challenges facing privacy self-management. The first is that users are not, in practice, informed about how their data is used because few look in detail at privacy notices. Second, he contends that users may not have enough experience or understanding to make an informed decision on the use of their data, and may use heuristics, or be influenced by how the choice is framed. Third, Solove points to the lack of feasibility relating to scalability of privacy self-management given the large numbers of entities that collect and use user data. A fourth challenge is that data shared individually at one point in time can be combined with other data revealed at a later stage, thereby compromising privacy. Finally, he indicates that it is difficult for users to assess harm, and they may prefer immediate benefits.

2.4 Students' Perspectives of Ethics and Privacy in Learning Analytics

Relevant literature on empirical studies with students on ethics, privacy and LA was identified by searching for these terms in the following leading databases[3] : Association for Computing Machinery (ACM), the *dblp* computer science bibliography, IEEEXplore, ScienceDirect, SpringerLink, and the Web of Science Core Collection, following Webster and Watson's (2002) structured approach. Additionally, the search was carried out in the Journal of Learning Analytics and in the proceedings of the Learning Analytics & Knowledge conference. These are prominent publication avenues for research in LA. The search was first conducted in April 2018, and updates were made over the course of conducting the thesis research. A final search was carried out for relevant updates while writing this thesis. Sixteen unique articles (that is, excluding duplicate results) were identified using this search strategy. In reviewing empirical literature on students, privacy, and LA, three major research strands were identified as summarised in Table 2.2.

---

[3] The search string used was "ethics privacy students "learning analytics"

Table 2.2: Summary of related research on ethics and privacy in learning analytics focusing on students' perspectives

Paper	Method	Sample size	Country	Outline and summary of ethics and privacy-related findings (Students' views)
Concern about Learning Analytics				
Nevaranta, Lempinen, and Kaila, 2020	Survey	201 students	Finland	• Find learning analytics potentially useful. • Concerned about ethics of data storage and handling of data by third parties. • Concerned about how the data will be used. • Want to restrict the use of their data.
Falcao, Ferreira, Rodrigues, Diniz, and Gasevic, 2019	Interviews and focus group	22 students	Brazil	• Not concerned about higher educational institution using their data. • Do not recall consenting to the use of their data. • Want to remain responsible for their learning.
Slade, Prinsloo, and Khalil, 2019	Survey	286 students	UK	• Near equal split between acceptance and non-acceptance of institutional use of student data. • Participants favoured having control over the collection and use of their data. • Participants expressed high levels of trust in the University.
Sun, Mhaidli, Watel, Brooks, and Schaub, 2019	Semi-structured interviews	4 developers; 8 advisors; 20 students	USA	• Students favoured control over use of their data.
Knox, 2017	Pilot testing in an educational context	12 students	UK	• Concerns that some students' learning activities are excluded therefore the analytics is inaccurate. • Concerns about algorithmic decision making about students' learning.
	Focus group	41 students	Australia	• Students wanted responsibility for their education.

Roberts, Howell, Seaman, and Gibson, 2016				• Potential to demotivate students. • Invasion of privacy due to sense of being monitored all the time.
Willingness to Share Data				
Li, Sun, Schaub, and Brooks, 2021	Survey	119 students	USA	• Identify factors influencing participants' willingness to take part in learning analytics.
Ifenthaler and Schumacher, 2019	Survey	330 students	Germany	• Control over data was positively associated with willingness to share data. • Students who have been at an institution for longer are more willing to share data. • The more the expected benefit, the higher the willingness to share data.
Ifenthaler and Schumacher, 2016	Lab study	330 students	Germany	• Students unwilling to share all the data needed for learning analytics.
Specific Privacy Concerns				
Jones, Asher, Goben, Perry, Salo, Briney, and Robertshaw, 2020	Semi-structured interviews	105 students	USA	• Students lack awareness of institutional use of student data for learning analytics. • Students desire control over who can access their data. • Students do not recall consenting to the use of their data. • Students trusted their learning institution. • Students expected to provide their data in exchange for a service from the learning institution.
Tsai, Whitelock_Wainwright, and Gasevic, 2020	Survey and focus group	Over 600 students	UK	• Differences in students' expectations for ethics and privacy and what is provided.
Vu, Adkins, and Henderson, 2019	Survey	1,647 students	USA	• Students not concerned about institutional use of their data when the recipient and data uses are made clear.

Schumacher and Ifenthaler, 2018	Interviews	20 students		• Two participants unwilling to use LA due to concerns about surveillance.
	Lab study	216 students	Germany	• Participants would use a learning analytics feature if they did not perceive it as invasive.
Arnold and Sclater, 2017	Survey	406 students 425 and 916 students	UK USA	• Students support use of data to prevent students dropping out and help them improve their grades.
Roberts, Howell, and Seaman, 2017	Focus group	41 students	Australia	• Expressed a desire for anonymity.
Slade and Prinsloo, 2014	Moderated forum discussion	35 students	UK	• Concerns about surveillance and tracking.

The first strand seen in Table 2.2 has a focus on concern about LA, and considers, for example, whether students are concerned about privacy in the use of student data for LA. The second strand of empirical research has concentrated on students' willingness to share their data for LA. The third strand has concentrated on specific privacy concerns. It is noted that these issues are examined in some publications at the same time, meaning for example, a publication might focus on both students' acceptance of LA and identifying specific concerns related to privacy. This related work is summarised in this section to identify the gaps in knowledge that this thesis research concentrated on. The lessons learned from the literature will be synthesised and the research gaps this thesis focused on will then be identified in Section 2.5.

2.4.1   Concern about Learning Analytics

In a survey carried out both online and face-to-face at two Finnish HEIs (Nevaranta, Lempinen, & Kaila, 2020), students were asked about their attitudes towards LA, what they expected from it, what they thought it could offer, and what its risks were. While students were identified as being positive about the use of their data for LA, they also expressed concerns about how else the data might be used and who it would be shared with. Thus, this finding demonstrates the nuances present in students' perceptions where they express an acceptance of institutional data use while at the same time expressing concerns about dimensions of privacy.

When informed that their data is used for LA applications, students advocate for informed consent, and want to be involved in decisions on the use of their educational data (Sun, Mhaidli, Watel, Brooks, & Schaub, 2019). The authors carried out 32 semi-structured interviews with different stakeholders to explore their perceptions of the use of student data in an LA tool called Student Explorer, deployed for use by academic advisors. The following stakeholders were interviewed: 4 developers of the Student Explorer (SE) dashboard, 8 academic advisors and 20 undergraduate students. The aim was to understand these stakeholders' perceptions, attitudes, and expectations towards the access, use, and analysis of student data in LA applications. The questions asked of participants during the interview differed based on whether the authors were interacting with the system developers, academic advisors who use the LA tool, or students whose data the tool uses.

A key finding from the interviews with students was that they were unaware of the LA tool. While this was to be expected as the tool was designed for advisors, given that it uses students' data, the authors argued that students should be made aware of the various ways their data is used. There was lack of agreement whether students themselves should gain access to the LA tool although they appreciated the potential benefits they saw it could offer them. Specifically, students could envision using such a tool to track their personal progress, better prepare for

meetings with their advisors, motivate themselves based on comparison with their peers, and determine if a course was suitable for them. Also of particular interest were students' views on their control over the data that was used for the LA tool. Twelve students were in favour of student control and suggested different ways to provide it, including providing e-signatures over email (3 students), students being notified when their data was used in new ways (4 students), students being informed about the potential benefits of data use (2 students) and the tool using anonymised data (2 students). Four students did not support student control over data use as they were concerned that it would impact on data quality and the benefits that could be provided. Both the advisors and developers seemed to agree that the question of whether students should be able to opt out of the use of their data was best handled by the university leadership. This work highlights the differing perspectives of LA held by different stakeholders.

Students appear to expect to exchange their data for learning-related benefits (Slade, Prinsloo and Khalil, 2019). The authors conducted a survey with 286 first year students (215 provided complete responses) at the Open University (UK) to explore their perceptions of online privacy in general, their privacy self-management behaviours, and their perceptions of privacy in their online studies. The work was guided by the privacy calculus theory (Culnan & Armstrong, 1999; Dinev & Hart, 2006) and privacy self-management (Baruh & Popescu, 2017; Solove, 2013). The questions asked were in the following categories: i) awareness of uses of personal data; ii) use of social media; iii) protection of privacy online; iv) online data and practice at the Open University.

Three key issues can be emphasised from the findings: first, participants were noted to accept the use of their data for potential learning-related benefits. However, there was very little difference between these two, as it was noted that the study found participants were almost equally split between accepting (32%) and not accepting (34%) the use of their data for perceived benefits. Second, participants were in favour of having control over the collection and use of their data. They did not favour sharing data with third parties. A majority of participants (72%) did not want their online activities followed without their permission, 69% favoured having control over who could access information about them, and 67% wanted control over information collected about them. Despite expressing a desire for control over data, these participants did not exhibit high privacy behaviours. This might be explained by the study being conducted in the learning context where participants might have low levels of privacy concern. Indeed, participants were more interested in privacy-protection with respect to financial information. Third, participants had high levels of trust in the university. Respondents in this study also cared more about who saw the information rather than what was collected.

Knox (2017) reports on the design and implementation of the Learning Analytics Report Card (LARC), a data literacy tool to give students choice in the use of their data in the educational context, and to raise their own and their teachers' awareness of how student data can be captured, analysed, and used. The LARC allowed students to choose the data that is included in a report on their academic progress. Students could choose the period of time the report would cover as well as its content, choosing from the following options: engagement, attendance, social interaction, performance and personal. Attendance data covered dates and times when students logged into the learning management system, engagement covered the frequency of their interaction with course materials, social interaction looked at data from the discussion forums, performance compared their data to that from other students, and personal looked at their profile information and tasks they needed to do at the start of the course. Rather than aim for a predictive analytics approach in LARC, the authors indicate that the LARC aimed to enable students to reflect on their performance based on the report and take necessary changes to maintain or improve their performance. In this way the LARC seems to prioritise student autonomy over their learning, which is one of the privacy and utility tensions discussed in Chapter 1.

Twelve students at a UK university were given access to LARC over 12 weeks and provided their feedback on the content of the report. Students questioned the accuracy of the analytics. They contended that the LARC system did not consider all the learning activities the student carried out. The analytics were criticised as only focusing on ranking and grouping students rather than understanding the learning activity or related elements around it. One student commented that their trust in the reports would only develop over time. Questions were also raised whether there were elements of the students' learning that were best left private. Another participant was uneasy about losing the opportunity to make decisions on their learning or having decisions about their learning being made on their behalf by algorithms.

Similar results were identified in a study with 22 Brazilian students, who were interviewed individually and in focus groups (Falcao, Ferreira, Rodrigues, Diniz, & Gašević, 2019). They were willing to have their HEI use their data to show them their performance in various courses, as well as identify areas of weakness and provide guidance on available support. However, some students wanted to be responsible for their learning. While they indicated that they did not recall consenting to the use of their data, the students in this study stated that they were not concerned about the HEI using their data.

Roberts, Howell, Seaman and Gibson (2016) focused on students' knowledge, attitudes, and concerns about LA. They conducted four focus groups with 41 students. While participants were

in general positive about LA, they expressed numerous concerns including that the data was incomplete and thus inaccurate, each student was different thus prediction was problematic, there was potential that the feedback would demotivate students, students wanted responsibility for their learning, and that students' informed consent was necessary.

Summary: The literature summarised in this section has revealed a nuanced picture about students' concern about LA. The findings from the literature suggest that students are not concerned about LA, influenced by their trust in the HEI and expectations that the HEI would use students' data to benefit students' learning. At the same time, concerns have been raised focusing on the lack of clarity for students about how the data will be used and who it will be shared with, students' loss of autonomy over their learning, and the potential for LA to paint an incomplete picture about students' learning.

2.4.2   Willingness to Share Data

Li, Sun, Schaub and Brooks (2021) examined students' perspectives on the use of their educational data for LA, and the population and participation characteristics of those who consent and do not consent to the use of their educational data for LA. The authors distributed an email prompt and survey and approached 4000 students. From this original number, 272 responded to the prompt and 119 filled in the survey. The study was conducted in the USA and used the following measures: i) factors affecting students' decision; ii) perceived benefits; iii) concerns with data use; iv) comfort with use of educational data in 5 scenarios; v) level of competitiveness and cooperativeness; vi) institutional trust; vii) levels of individualism and collectivism; and viii) a privacy concern questionnaire - Internet Users Information Privacy Concern scale. The findings included identifying that trust in the learning institution, level of general concern about data collection, and comfort with the instructors' use of data for classroom purposes all influenced participants' decision to take part in LA.

Ifenthaler and Schumacher (2019) conducted a survey with 330 students to investigate whether students are willing to release personal information for use in LA applications. The authors based their work on the privacy calculus model (Culnan & Armstrong, 1999; Dinev & Hart, 2006). In this study, the authors hypothesised that: (1a) students expect transparency in LA systems, (1b) students expect sensitive and responsible processing of data (1c) students' willingness to provide personal data is related to their anticipated control over data. Additionally, the authors hypothesised that disclosing personal data was related to (2a) year of study, (2b) course load, (2c) study interest, (2d) use of the Internet for learning, and (2e) use of social media. Finally, the authors hypothesised that releasing personal data is related to anticipated benefits from a LA application (3). The following findings were supported: 1c) participants were more willing to share

data when they were aware about their control over their data (1c), they had been at the university for longer (2a), they used the Internet (2d) and social media (2e) for learning, and they anticipated potential benefits for sharing their personal data (3). Students in this study were unwilling to share personal data with their lecturers, a finding which is in contrast to that of Vu, Adkins, and Henderson (2019).

Ifenthaler and Schumacher (2016) carried out a laboratory study and survey with 330 undergraduate students to get students' perceptions of two privacy principles, namely, sharing data and control of data, as well as their acceptance of LA. The study questions focused on understanding 1) whether students want to use a LA system to support their learning; 2) the extent to which students were willing to share data on LA systems, and 3) the extent to which students' sharing data is related to their preference for LA systems. The findings demonstrated that students prefer LA systems with more functionality, such as support for personalised content and activity recommendations. Students in the study were less willing to share their data with LA systems. They were not willing to share all data needed for LA with more features, i.e., their personal data and data trails from the virtual learning environment. Control over data was positively related with acceptance and expected use of the LA systems. Control over data was also positively associated with willingness to share data. Further, acceptance and expected use of the LA systems was positively associated with willingness to share data. Although well-designed, this study also had a number of limitations which the authors acknowledge in their work, including that the study collected participants' self-reported data which could differ from their actual behaviour.

Summary: The literature summarised in this section emphasises that there are likely numerous factors which will play a role in students' decision to share data for LA where such opportunities are provided.

2.4.3   Specific Privacy Concerns

In general, research findings suggest that students are positive about institutional use of their data to enhance their own and other students' learning (Jones, Asher, Goben, Perry, Salo, Briney and Robertshaw, 2020). The authors carried out semi-structured interviews with 105 undergraduate students from 8 universities in the USA. Their aim was to identify the privacy issues students perceive in LA, their privacy expectations about their data, their expectations about how their data will be used and shared, and finally, their response to the use of LA. This work was guided by the theory of intellectual privacy (Richards, 2012) and contextual integrity (Nissenbaum, 2010). According to the theory of intellectual privacy (Richards, 2012, p. 1946), new ideas develop best when one is not subjected to public scrutiny, where one can think freely

without surveillance. The authors identified six main themes in their findings: i) that the participants were unaware of LA and how HEIs use student data (this result is also reported in related work: (Jones, et al., 2019; Sun, Mhaidli, Watel, Brooks, & Schaub, 2019); ii) while the participants were found to support LA, they were not as clear on what it could achieve; iii) participants wanted to have a say on who could access their data; iv) the students in the study could not recall consenting to the collection and use of their data, similar to findings of Tsai, Whitelock-Wainwright and Gašević (2020); v) the students expressed a high degree of trust in their learning institutions, a finding which has also been identified in other research (Li, Sun, Schaub, & Brooks, 2021; Slade, Prinsloo, & Khalil, 2019); vi) finally, while participants expressed that they were against the sale of their data to commercial entities, they viewed their relationship with the HEI as transactional, such that students use their data to pay for a service. The limitations of this study arose from the use of a convenience sample, and thus, participants may have been those who were interested in privacy issues.

While students are noted to welcome institutional use of their data, this was for what they considered to be legitimate purposes, as identified in the work of Tsai, Whitelock-Wainwright and Gašević (2020), namely, to comply with legal requirements, to improve educational services and to improve the university's overall performance. Using a survey with 674 students and additionally through 26 focus group discussions, the authors focused on identifying i) what students considered as legitimate uses of their personal and educational data for higher education, and ii) whether there were gaps between student perceptions or expectations of privacy and the actions that they had taken to protect their data. This work was guided by two models for understanding privacy: privacy paradox (Norberg, Horne, & Horne, 2007; Kokolakis, 2017) and contextual integrity (Nissenbaum, 2010). The work was carried out in the UK with a mix of international and domestic students. While the authors identified differences between students' desires for ethics and privacy in institutional use of student data and their expectations of what will be done, it is not reported whether these differences were statistically significant. Furthermore, the authors did not address the possibility that these differences were due to students lacking opportunities to express their data use preferences to the HEI. Participants in the focus groups expressed an expectation of exchange of data for benefits similar to related research (Jones, et al., 2020; Slade, Prinsloo, & Khalil, 2019). These students were uncomfortable with sharing data with third parties, and they were split between wanting freedom to choose how their data is used versus being able to make some decisions about the type of data that can be used (rather than all the data). Finally, with respect to consent, students could neither recall whether they gave consent, nor what they consented to, also as reported in related work (Jones, et al., 2020).

Fewer students are noted to have privacy concerns over data use when greater detail is provided during the research about what data is collected, who has access to it and how it is used. In a survey in HEIs in the USA, Vu, Adkins, and Henderson (2019) focused on understanding whether students are aware of and concerned about their instructors use of their data. There were 1,647 participants who were asked the following questions: Are you aware that your learning: A1) activity (such as login frequency) in your online courses could be seen and recorded by your instructors? A2) activity (such as pages viewed or clicked) in your online courses could be seen and recorded by your instructors? A3) profiles in your online courses could be seen and recorded by your instructors? B1) Are you concerned that most of your learning behaviours in your online courses (such as login frequency, page viewed and learning profile) can be monitored and recorded by your instructors? B2) Would you be concerned if your instructors collected your learning data in your online courses without revealing your personal information (name, gender, … etc.) for academic or research purposes? B3) Would you be concerned if your instructors collect your learning data in your online courses without revealing your personal information (name, gender, … etc.) for instructional/teaching improvement purposes? Based on the results, the authors make the claim that participants were aware of the use of their data and did not care. However, the results could also plausibly demonstrate that the study participants were neutral in their outlook. A neutral stance might not be the same as not caring. In the framing of the study questions, the authors addressed two issues which contribute to privacy concerns (Nissenbaum, 2010): the question of 'who has access to the data', that is, instructors and 'how will the data be used' - that is for academic or research purposes. This transparency might explain the results the authors obtained.

Participants in the two studies by Schumacher and Ifenthaler (2018) expressed their preferences for LA features. In doing so, two participants were noted to be concerned about too much surveillance. These two studies were conducted as part of mixed methods research with a qualitative exploratory study with 20 participants followed by a quantitative study with 216 participants. Both studies were conducted in Germany. While this work does identify students' preferences for learning analytics features, it did not investigate what they viewed as privacy violations in the context of LA.

Other legitimate purposes for use of student data were identified in the work of Arnold and Sclater (2017) and includes preventing students from dropping out, helping students to improve their grades, and allowing them to compare their performance with other students. In interviews with 406 students at different UK universities, and an online survey conducted in two rounds with 425 and 916 students in the US, students were asked if they would be happy i) for their learning activity data to be used if it kept them from dropping out or helped them get personalised

44

interventions, ii) for their data to be used if it helped improve their grades, and iii) to have their data visualised and their performance compared to their classmates through an app. The study showed that LA being used to prevent students from dropping out and improve their grades was more acceptable to both UK and US students than the data being used to support comparison with other students. In the latter case, acceptance was low but much lower with UK students. Furthermore, US students were more receptive of LA being used to prevent them dropping out (76% and 72%) than UK students (53% and 54%). Both groups of students were receptive to LA being used to help them improve their grades, but US-based students were more accepting (94% and 91%) than UK-based students (71% and 77%). Finally, US-based students were more receptive (61% and 60%) than UK-based students (25% and 26%) to LA being used to support comparison with other students using an app. While the findings of this study can be used to inform the design of LA, there are several limitations with the study and question design. First, students were asked questions about a hypothetical situation which does not indicate how they would actually respond to LA designed with the same functionality. Second, both the first and third questions combined two issues which would have best been considered separately. The first question referred to stopping students from dropping out and providing them with personalised interventions, while the third question referred to using an app and comparing themselves with their classmates. As a result, it may not be immediately clear which of these issues students supported or were opposed to. While the implementation of these issues in LA could be similar, e.g., personalised interventions only prevent students from dropping out, these nuanced differences might not be clear to students. Third, the paper does not clearly report how the students based in the UK were recruited for the interviews, whether these students were familiar with or had used LA, whether the interviews were structured or semi-structured, and how long they took.

Privacy concerns were also raised by participants in the study by Roberts, Howell and Seaman (2017). The authors explored students' perceptions of LA dashboards and the features they preferred. They conducted four focus groups with a total of 41 students. Five themes emerged. Of particular interest is the theme 'dashboard privacy'. Here, students expressed a preference to be compared anonymously to their peers, that is, without using their names. In addition, they anticipated opening the dashboards in public and did not want others to know e.g., by glancing over their shoulder, how they were performing.

Other dimensions of privacy that concern students have to do with surveillance and tracking of the individual student. Slade and Prinsloo (2014) obtained feedback from approximately 35 students on the LA policy that the OU was developing at the time. The authors focused on: 1. keeping information up to date; 2. transparency (why the university collects data and how it is

45

used); 3. students' experience with student support at the university, and 4. concerns about data collection and storage. Questions were distributed to students on a discussion forum. One of the authors acted as a moderator and released questions regularly for discussion. Participants were asked:

1. Do you regularly keep your StudentHome profile and other information such as study goals up to date? Is there anything the OU could do to make it easier to do this?

2. How often should the OU give you opportunities to check and update your data, and give consent to its use? What would be the most effective way of doing this?

3. Do you think the OU makes clear enough how and why it collects and analyses data? How do you think the OU should communicate its approach to students in the future?

4. Can you think of occasions when the OU has actively used data it has built up about you to offer you support tailored to your needs? Have these initiatives effectively used the information the OU holds about you?

5. Have you ever been offered support that you felt has not been based on relevant, up- to-date and accurate information the University holds about you?

6. Have you any other concerns about data collection, storage, updating and that you think the OU should address?

7. Do you think there is any information that the OU doesn't collect or use at present which it should do in the future?

8. Is there any information which the OU currently collects that you think is not relevant to supporting you as a learner?

9. Is there anything else you would like us to consider? The study found that participants were not motivated to keep their information updated as they did not know how it was used.

They expressed concerns about surveillance and tracking. There was conflict between participants' stated desire for personalised support, for example, emails that target them as individuals, and concerns about surveillance and data collection. Asking about data collection and privacy raised students' concerns. Students also indicated an interest in providing informed consent and having an opt-out option available. However, an opt-out clause was not provided for in the university policy on the use of student data for LA (Slade & Boroowa, 2014). Finally, this study results identified students' expectation that their learning institutions would be transparent about the data used for LA, and inform them what data was collected, how, where, for what purpose and who would have access to it.

Summary: From the literature summarised in this section, it becomes clear that students' specific privacy concerns are varied. These include concerns about surveillance and tracking,

sale of student data, and data sharing. Students express conditional acceptance of LA and some of this is influenced by an expectation of an exchange of their data in return for a service from the HEI. Finally, students' concerns may be lessened where they are informed about data collection and how the data will be used.

2.5   Summary of Research Gaps

The body of research summarised in this section has enhanced understanding on several issues related to concern about LA, students' willingness to share data, and students' specific privacy concerns. As indicated in Section 2.3.1, there is a lack of literature on stakeholders' concepts of privacy in LA. There are specific privacy concerns that have been identified in this context, including about other uses the data could be put to and who it could be shared with (Nevaranta, Lempinen, & Kaila, 2020; Slade, Prinsloo, & Khalil, 2019), surveillance of the students by the institution (Slade & Prinsloo, 2014), and the possibility that student data could be sold (Jones, et al., 2020). A gap identified in this case is that it is unclear what students' conceptualisations of privacy in LA are. Knowing how they perceive their privacy and the privacy of their data could provide insights into what contributes to their privacy concern or lack of privacy concern. As a focus on privacy concepts was deemed useful to identify privacy and utility tensions in LA, this formed the first research gap.

The literature summarised in this section suggests that there is a level of support or acceptance among students of the use of their data for LA (Nevaranta, Lempinen, & Kaila, 2020; Slade, Prinsloo, & Khalil, 2019; Falcao, Ferreira, Rodrigues, Diniz, & Gašević, 2019; Roberts, Howell, Seaman, & Gibson, 2016). This could be attributed to the reported high levels of trust in learning institutions (Jones, et al., 2020; Slade, Prinsloo, & Khalil, 2019). At the same time, in the work of Slade, Prinsloo and Khalil (2019), there was nearly an equal number of students who did not accept the use of their data for LA as those who did. Furthermore, findings from other studies highlight students' concern about privacy being invaded (Roberts, Howell, Seaman, & Gibson, 2016), students' unwillingness to share all the data needed for LA (Ifenthaler & Schumacher, 2019), and their concerns about surveillance (Schumacher & Ifenthaler, 2018; Slade & Prinsloo, 2014) and tracking (Slade & Prinsloo, 2014). This body of literature suggests that there are differing and conflicting perspectives about privacy from students' perspectives. Therefore, the second research gap identified was the need to better understand students' privacy concern with the use of their data for LA.

The literature also suggests that students have come to expect the use of their data in the educational context (Jones, et al., 2020), and they expect to provide their data in exchange for some service from the HEI (Slade, Prinsloo, & Khalil, 2019). That said, little is known about

students' perceptions of providing their data in return for some service from the HEI, where, in order to access LA and its advanced features, students have to accept institutional use of their data (Siemens, 2013, p. 1394). In this regard, students are asked to consent (or are taken to consent) to the use of their data for LA so that the data can be used for potentially beneficial purposes such as providing students with learning recommendations and intervening early to recommend remedial action and improve students' performance (Ho, 2017; Siemens, 2013). Use of student data in these ways has potential for privacy harms such as profiling and identification of the individual whose data is used (Solove, 2009), as discussed in Section 2.3.1.4. While insightful research on students' perspectives of the ethics and privacy of LA is ongoing, little is known about students' perceptions of this exchange, their preferences for the use of their data for LA and whether and how these preferences change when the trade-offs are made salient.

Some challenges identified in the literature review include that students do not recall providing consent for the use of their data (Sun, Mhaidli, Watel, Brooks, & Schaub, 2019; Falcao, Ferreira, Rodrigues, Diniz, & Gašević, 2019), and they express concern that the data used shows an incomplete or inaccurate picture of the students' learning (Knox, 2017; Roberts, Howell, Seaman, & Gibson, 2016). Finally, students favour having control over the collection and use of their data (Sun, Mhaidli, Watel, Brooks, & Schaub, 2019; Slade, Prinsloo, & Khalil, 2019) and express an interest in providing consent for the use of their data (Slade, Prinsloo, & Khalil, 2019; Roberts, Howell, Seaman, & Gibson, 2016) and being responsible for their learning (Falcao, Ferreira, Rodrigues, Diniz, & Gašević, 2019; Roberts, Howell, Seaman, & Gibson, 2016).

Students' willingness to share data for LA is likely influenced by multiple factors including their trust in the learning institution, their levels of concern about data collection, and their comfort with their instructors using their data for teaching and learning purposes (Li, Sun, Schaub, & Brooks, 2021), as well as, how much control they perceive they have over the use of their data, how long they have been at a learning institution, their level of use of the Internet and social media for learning and the anticipated potential benefits in return for providing their data (Ifenthaler & Schumacher, 2019). However, some students were seen to be reluctant to share all the data that could be shared for LA (Ifenthaler & Schumacher, 2019; Knox, 2017). The research carried out does not answer questions about why users are willing or unwilling to share their data for LA. Consequently, seeking to understand students' data use preferences formed the third research gap.

Learning analytics has numerous stakeholders, and therefore the next section identifies the stakeholders the thesis research focused on.

2.5.1 Stakeholders

Drachsler and Greller (2012) identify LA stakeholders as those who contribute to LA for example, students, and those who potentially benefit from it (including students, teachers, learning institutions, and governments). LA stakeholders can also be categorised based on their contribution to or involvement in the development and use of LA applications. Three high-level stages can be distinguished, specifically, the design and development of LA applications where consideration is given to what problems need to be solved, what data is available, and what applications can be developed. The second high-level stage is the deployment stage where the LA applications are implemented in HEIs. Finally, the third high-level stage involves the use of LA applications. Figure 2.1 illustrates these three stages.

Figure 2.1: High level view of stages in learning analytics

Keeping in mind these three high-level stages, it is clear that there are numerous stakeholders involved in LA, including analysts, application developers, institutional representatives, policy makers, practitioners, researchers, students, student support staff, and teachers. While each of these stakeholders are involved in the development and use of LA, the benefits of LA applications may not affect each stakeholder in the same way. As the thesis research focused on privacy and utility tensions in LA, two stakeholder groups were identified for the thesis research.

LA experts are involved in the development of ethics and privacy frameworks and their research influences policy development for LA in higher education. Ethics and privacy frameworks are an important element in the design and implementation of LA applications and in the debates on ethical and privacy concerns in LA (Ferguson, 2012; Slade & Prinsloo, 2013; Pardo & Siemens, 2014; Hoel & Chen, 2018). While LA experts are not included explicitly as stakeholders in prominent LA literature that has identified stakeholders, for example, Greller and Drachsler (2012), they can be categorised as researchers, and they play a key role in the development of ethics and privacy frameworks for LA. This suggests that their individual and collective perspective is crucial for the thesis research. Thus, the first stakeholder group identified for the thesis research was LA experts as they are involved in the development of ethics and privacy frameworks for LA.

Students' perspectives were deemed relevant given an observed lack of agreement in research literature as to whether students are concerned about privacy in LA. Thus, the second stakeholder group identified for the thesis research was students.

The next chapter identifies the research questions and discusses the methodology used to address them.

3   Methodology

This chapter discusses the research methodology which guided the research and identifies the research methods applied to address the thesis research questions. Section 3.1 is an overview of the thesis research questions identified from the literature review in Chapter 2. This is followed in Section 3.2 by a discussion of the various research methods used in related research on students' perspectives of ethics and privacy in LA. This section examines the advantages and the limitations of the different research methods which have been used. Section 3.3 then situates the thesis research within the context of mixed methods research. This section contains a review of the support for and critique against mixed methods research, and a discussion of its suitability for this work. Section 3.4 provides a discussion of the research design and the selected research methods. In Section 3.5, the data analysis techniques applied to the qualitative and quantitative data collected for this research are detailed. Given that this research was carried out with human participants, ethical consent was sought from the university's Human Research Ethics Committee (HREC) before participants were recruited for the studies. Details about relevant ethical considerations for the research are provided in Section 3.6. To account for the role of the researcher in this work, Section 3.7 is a reflexive account, highlighting the assumptions, influences, and preferences brought by the thesis researcher to this work. Issues of validity, reliability, credibility, and dependability are also discussed throughout this chapter. Finally, Section 3.8 provides a summary of the chapter.

3.1   Introduction

The **main thesis research question** was:

> *What contributes to privacy and utility tensions in learning analytics?*

The literature review on students' perspectives of ethics and privacy in LA (discussed in Section 2.4) identified several research gaps. Three organising concepts were identified to help address the thesis research question, namely, privacy concepts, privacy concern, and data use preferences. These organising concepts proved useful to situate the focus of the thesis research and identify where a contribution could be made. Several research aims and corresponding questions were then identified to address the main research questions. These are summarised in Figure 3.1.

AIMS	RESEARCH QUESTIONS
Identify how privacy is conceptualised in learning analytics and identify the key privacy issues	**RQ1:** How do learning analytics experts describe and talk about privacy in learning analytics? **RQ2:** What do the experts consider to be key privacy issues in learning analytics?
Clarify the dimensions of students' privacy concerns in learning analytics	**RQ3:** To what extent are students concerned about the collection, use, and sharing of their data for learning analytics, and compared to e-commerce? **RQ4:** To what extent are students' general privacy concerns and behaviour related to their concern about the collection, use, and sharing of student data for learning analytics? **RQ5:** What issues contribute to students' concern or lack of concern about data collection, use, and sharing in learning analytics?
Investigate the extent to which an awareness of potential privacy risks and benefits of learning analytics influences students' data use preferences?	**RQ6:** To what extent does an awareness of the possible privacy risks and benefits of data use for learning analytics contribute to a difference in students' preferences for the use of their data? **RQ7:** To what extent does not having access to learning analytics applications influence students' preferences for the use of their data? **RQ8:** What do students indicate as the motivation for their preferences for the use of their data?
	RQ9: What insights can be obtained about the motivation for students changing or not changing their data use preferences?

Main RQ:

What contributes to privacy and utility tensions in learning analytics?

Figure 3.1: The thesis aims and research questions

There was a need to better understand how privacy is conceptualised in LA and it was deemed relevant to do so from the perspective of LA experts since they are involved in developing ethics and privacy frameworks which are then used when implementing LA in HEIs. Therefore, two research questions were identified with respect to the first organising concept:

RQ 1: How do learning analytics experts describe and talk about privacy in learning analytics?

RQ 2: What do the experts consider to be key privacy issues in learning analytics?

The review of related research suggests that students are in general accepting of the use of student data for LA by HEIs but at the same time have differing and conflicting perspectives about privacy in LA. Additionally, the research demonstrates the need for further research on factors that influence students' acceptance of institutional use of their data for LA, and within this context, how students balance the need for institutional use of their data to deliver a desired service. Finally, the research points to a need to increase students' awareness of LA and a desire on the part of students to be involved in decisions on the use of their data. Given the mixed results on students' privacy concern in the LA literature, the second aim of the thesis was to clarify the extent to which students are concerned about privacy in LA and to better understand the dimensions of any privacy concerns they may have. Consequently, the following research questions were identified:

RQ 3: To what extent are students concerned about the collection, use, and sharing of their data for learning analytics, and compared to e-commerce[4]?

RQ 4: To what extent are students' general privacy concerns and behaviour related to their concern about the collection, use, and sharing of student data for learning analytics?

RQ 5: What issues contribute to students' concern or lack of concern about data collection, use, and sharing in learning analytics?

The review of related research has further shown that an area requiring more research is that of students' perceptions of the exchange of their data in return for a service in the context of LA (Siemens, 2013, p. 1394) where to access LA and its advanced features, students have to accept institutional use of their data. In this regard, students are asked to consent (or taken to consent) to the use of their data for LA so that the data can be used for potentially beneficial purposes such as providing students with learning recommendations and intervening early to recommend

---

[4] Based on contextual integrity (Nissenbaum, 2010; Nissenbaum, 2004) concern about collection, use, and sharing of data can vary across different contexts. Therefore, two contexts were chosen, learning analytics and e-commerce.

remedial action and improve students' performance (Ho, 2017; Siemens, 2013). Use of student data in these ways has potential for privacy harms such as profiling and identification of the individual whose data is used (Solove, 2009). While insightful research on students' perspectives of the ethics and privacy of LA is ongoing, little is known about students' perceptions of this potential privacy risk and benefit trade-off, their preferences for the use of their data for LA and whether and how these preferences change when the trade-offs are made salient. Therefore, the third thesis aim was to investigate the extent to which students' awareness of the potential privacy risks and benefits of LA influenced their data use preferences. Consequently, the following research questions were identified:

RQ 6: To what extent does an awareness of the possible privacy risks and benefits of data use for learning analytics contribute to a difference in students' preferences for the use of their data?

RQ 7: To what extent does not having access to learning analytics applications influence students' preferences for the use of their data?

RQ 8: What do students indicate as the motivation for their preferences for use of their data?

As further insights were sought regarding participants' data use preferences, the final thesis research question was as follows:

RQ 9: What insights can be obtained about the motivation for students changing or not changing their data use preferences?

With the thesis research questions identified, the next section presents a critical analysis of the research methods used in related research on students' perspectives of ethics and privacy in LA.

3.2   Research Methods Used to Identify Students' Perspectives of Ethics and Privacy in Learning Analytics

There are a variety of research methods which can be used to address the thesis research questions. Before detailing the research methods selected for this thesis research, this section provides an overview of the different research methods which were under consideration to address the thesis research questions. These research methods have been used in related research to determine students' perspectives of ethics and privacy in LA. These research methods were noted in the literature review summarised in Section 2.4 (refer to Table 2.2 for a summary of the research findings and corresponding research methods). Their advantages and limitations are assessed to determine their suitability for the thesis research.

3.2.1 Interviews

LA researchers have used interviews to gather verbal data from participants. This includes work by Sun, Mhaidli, Watel, Brooks, and Schaub (2019) and Jones et al., (2020). There are different types of interviews, namely structured, semi-structured, and unstructured interviews (Deakin & Wakefield, 2014). In structured interviews, the questions are pre-determined and strictly followed by the researcher when engaging with participants. Each participant is asked the same question and there is no variation in the order of presentation. Unstructured interviews adopt a more conversational approach and the topics discussed are determined by participants' responses (Lichtman, 2013). This means that the researcher will ask further questions based on a participant's response. With semi-structured interviews, questions or lines of inquiry are prepared in advance, however, other lines of questioning can be pursued during the interview, based on participants' responses. The interviewer can ask a question that is not in the interview schedule, again, based on a participant's response. This approach allows the researcher to probe for further insights. Doing so may also shed further light on the topic under investigation, highlighting issues which the researcher may not have been aware of when preparing the questions for discussion.

Interviews can be carried out face-to-face (f-t-f) where the interviewer and the participant are in the same physical location, or they can be conducted remotely, where the interviewer and the participant are in different physical locations. Remote interviews can be conducted using the telephone (Glogowska, Young, & Lockyer, 2011), video conferencing technology such as Skype, (Deakin & Wakefield, 2014), chat or instant messaging applications (Voida, Mynatt, Erickson, & Kellogg, 2004), or over email (Hawkins, 2018). While many factors might influence the decision to choose between conducting f-t-f or remote interviews, remote interviews are currently recommended where possible, considering the ongoing impact of the Covid-19 pandemic (Open University Human Research Ethics Committee, 2021).

Like many other research methods, including those discussed in this section, interviews have both advantages and limitations which researchers need to be aware of when choosing which method to use. A summary of the advantages and limitations of interviews in general and remote interviews are shown in Table 3.1.

Table 3.1: Advantages and limitations of interviews as a research method

Interviews in general

Advantages	Limitations
Flexibility	Transcription and analysis of qualitative data can be time-intensive
Collect in-depth perceptions and experiences from participants	Analysis of responses can be complex
Interviewer can provide participant with more information if needed to minimise misunderstanding and can ask follow-up questions	The interviewer can introduce bias
Better engagement with individual participants and opportunity to build rapport	

Remote interviews

Lower cost compared to f-t-f interviews	Might be hard to build rapport given the lack of visual cues
Allow participants from any geographical location	Interviewer needs to concentrate and keep interview on course
Convenient for interviewer and participant	Participant may be distracted by things in their environment
Participants might find it easier to discuss sensitive topics	Potential for data of low quality
Allow access to specialist groups of participants	

(Based on (Hewson, 2009; Glogowska, Young, & Lockyer, 2011; Atkins & Wallace, 2012; Rea & Parker, 2014; Queirós, Faria, & Almeida, 2017))

While f-t-f interviews might better allow the researcher to develop rapport with the participant (Glogowska, Young, & Lockyer, 2011), interviews carried out remotely mean that participation is not limited by the need to meet in person. Thus, participants can take part from anywhere in the world (depending on the limits set by the research question(s) and research design). Semi-structured and unstructured interviews can be flexible in nature allowing follow-up questions, further information, and discussion of issues which come up during the interview (Queirós, Faria, & Almeida, 2017). Furthermore, they allow one to focus on and develop an understanding of complex issues and there is greater engagement with individual participants who can provide in-depth insights into their perceptions and experiences. Remote interviews can be carried out at a time and location that is convenient for both the researcher and the participant. They can also be cheaper to conduct. Furthermore, participants may find it easier to discuss sensitive topics in remote interviews (Hewson, 2009). Finally, the researcher can access specialist or hard-to-reach participant groups. Considering the disadvantages, it can be hard to build rapport between the participant and researcher with remote interviews. Furthermore, there is a chance of interviewer bias where participants and how they respond are influenced by the interviewer (Myers & Newman, 2007; Rea & Parker, 2014). Data transcription and analysis of qualitative interview data can be time consuming, and the data analysis can be complex and not necessarily straightforward (Atkins & Wallace, 2012). There is also a chance that either the participant or the researcher is distracted by their environment during the remote interview. Thus, there is potential for low-quality data (Hewson, 2009). The steps taken to mitigate some of these limitations are discussed in Section 3.4. This research used interviews (as part of a Delphi study technique) with the LA experts in Study 1 (refer to Chapter 4), and follow-up interviews in Study 2 (refer to Chapter 5) and Study 4 (refer to Chapter 6).

3.2.2   Surveys

Another research method that LA researchers have employed to determine students' perspectives of ethics and privacy in LA is surveys, for example (Arnold & Sclater, 2017; Vu, Adkins, & Henderson, 2019; Slade, Prinsloo, & Khalil, 2019; Ifenthaler & Schumacher, 2019; Nevaranta, Lempinen, & Kaila, 2020; Li, Sun, Schaub, & Brooks, 2021). In surveys, individuals are asked to respond to several questions and researchers analyse the collected data to observe trends (Boeren, 2018). Given the aim to observe trends and determine the extent to which they apply to a population of interest, how a sample of study participants is drawn is especially important in survey research (Müller, Sedley, & Ferrall-Nunge, 2014). Random sampling is ideal as every member of the population has an equal probability of being in the study sample. Random sampling can be hard to achieve given that it might be difficult to identify all the members of the

population, there is the chance that those invited will choose not to participate, and some participants could drop out of the study (Strunk & Mwavita, 2020).

Other sampling strategies are available; however, these face sampling bias (Cohen, Manion, & Morrison, 2011; Strunk & Mwavita, 2020). i) Representative or quota sampling where targets are set to recruit participants with specific characteristics, such as demographic variables, which are representative in each demographic category. ii) In snowball sampling the researcher starts with an initial group of participants who are then asked to recruit other participants who might qualify to take part. With snowball sampling, the challenge is that participants may be very similar to each other due to the use of social contacts. iii) Another strategy is purposive sampling where the researcher uses their knowledge or judgement to select a sample to participate in a study. Finally, iv) convenience sampling involves the researcher selecting participants who they have access to (and thus, at the researcher's convenience).

Equally important is the design of the questions to be asked. Krosnick (2018) provides helpful guidelines which were applied for the thesis research. This included avoiding technical words and leading questions that suggest that there is a desirable response, taking time to ease the participant in by having early questions as those that are easy to answer, and pilot testing the questionnaire before collecting data from the study participants. Further, related questions were grouped together in the survey instruments. There was a mix of open and closed questions in the surveys used in the thesis research. Open questions invited participants' statements in response, while closed questions limited participants' response options to those determined by the researcher and guided by the research questions. Further details about the surveys and questions used in the thesis research are provided in the respective chapters (Chapter 4 to Chapter 7). The advantages and limitations of surveys are summarised in Table 3.2.

Table 3.2: Advantages and limitations of surveys as a research method

Advantages	Limitations
Convenience for both researcher and participants	A challenge for participants to seek to clarify unclear questions
Can allow for trends to be observed; generalisable findings (tied to sampling strategy)	Little opportunity for follow-up questions by the researcher
Speed and ease of distribution for web-based surveys; low cost	May not obtain in-depth, detailed data from participants
Allow participants from any geographical location	May suffer from low response rates
	Participants may not provide accurate responses and have poor recall of events

(Based on (Gable, 1994; Glasow, 2005; Queirós, Faria, & Almeida, 2017))

While surveys can be administered using mail, telephone, or in-person, they are also distributed over the Internet and benefit from the subsequent low cost of data collection, ease and speed of administration, and the opportunity to access globally distributed populations. Where surveys are self-administered, the researcher does not have to be present as participants answer the questions independently. Despite this convenience, the researcher cannot clarify unclear questions for participants, and this could negatively impact the quality of responses collected. In addition, there is potential for bias if questions focus on patterns of use where participants must recall something.

Response rates might be increased where participants have an interest in the topic of research (Groves, Presser, & Dipko, 2004). Monetary incentives may also be offered to improve the response rate as well as the quality of responses provided (Evans & Mathur, 2005; Laguilles, Williams, & Saunders, 2011). While concerns have been raised about the potential for financial incentives to coerce participation in research, it is also acknowledged that there may be a need to provide participants with a token for their participation (The British Psychological Society, 2021). Non-monetary incentives include thanking participants, offering to send them a summary of the research results and any recommendations, or offering to make a donation to a charity of the participant's choice, however, they may not be as effective in comparison (Pedersen & Nielsen, 2016; Kelly, Margolis, McCormack, LeBaron, & Chowdhury, 2017).

Despite their advantages, surveys are not ideal if one wants to obtain detailed data from participants. It is also difficult to ask participants for further insights or to modify the questions once the data has been collected. The steps taken to mitigate some of these limitations of surveys are discussed in Section 3.4. Surveys were used in Study 2 (refer to Chapter 5) and Study 3 (refer to Chapter 6).

3.2.3   Focus Groups

With focus groups the researcher acts as a facilitator and asks questions of multiple participants at the same time. Interaction between the participants is emphasised over interaction between participants and the researcher. Focus group studies have been carried out in research on ethics and privacy in LA, for example, Roberts, Howell, Seaman, and Gibson (2016; 2017). Focus groups can also be carried out online, for example, as in the facilitated forum discussion used by Slade and Prinsloo (2014) with discussions lasting over an extended period of time (Lichtman, 2013). Online focus groups might encourage more interaction where participants perceive a higher level of anonymity. Further, there are opportunities to reduce costs, for example, where data transcription is not required, or participants do not need to travel to attend the focus group sessions (Woodyatt, Finneran, & Stephenson, 2016). Other advantages and limitations are summarised in Table 3.3.

Table 3.3: Advantages and limitations of focus groups as a research method

Focus Groups	
Advantages	**Limitations**
Discussion can concentrate fully on the topic of interest	The discussion could be contrived and unnatural
Help identify group attitudes and group thinking on the topic	Needs a skilled facilitator
Generate large amounts of data in a short time on the topic of interest	Data analysis can be complex
Allows larger sample sizes to provide qualitative data	The facilitator can introduce bias
Online Focus Groups	
Potential to reduce costs	
Participants might better discuss sensitive topics where they feel anonymous	

(Based on (Morgan & Spanish, 1984; Wilkinson, 1998; Fallon & Brown, 2002; Woodyatt, Finneran, & Stephenson, 2016; McKim, 2017; Chaudhari, 2021))

Focus group studies bring together a group of participants to interact and engage with each other in response to issues raised by the facilitator. One can obtain a broad range of views and as participants interact with each other, differences and similarities in their viewpoints can be identified. However, facilitators need to ensure that talkative participants do not dominate the conversation and they also need to encourage quieter participants to express their views. Facilitators may have to limit their questions to keep focus groups from taking too long.

3.2.4   Mixed Methods

Researchers in ethics and privacy in LA have also taken a mixed methods approach in their research. For example, Tsai, Whitelock-Wainwright, and Gašević (2020) combined a survey questionnaire with focus group studies with the latter providing them with rich data from participants. This is acknowledged as one of the strengths of mixed methods research (Creswell & Plano Clark, 2007; Halcomb & Hickman, 2015; Shannon-Baker, 2016). Schumacher and Ifenthaler (2018) used interviews and an online survey. In their work, they first carried out an exploratory study which they used to better understand the research problem and develop a quantitative survey instrument for use in the second study. The advantages and limitations of mixed methods research are summarised in Table 3.4.

Table 3.4: Advantages and limitations of mixed methods research

Advantages	Limitations
Deeper understanding of the research problem from different perspectives	Time intensive
Address weaknesses of qualitative and quantitative research when carried out on their own	Researcher needs training in both approaches
Can answer complex research questions	

(Based on (Sale, Lohfeld, & Brazil, 2002; Johnson & Onwuegbuzie, 2004; Creswell & Plano Clark, 2007; Halcomb & Hickman, 2015; Shannon-Baker, 2016))

Mixed methods research involves researchers using both qualitative and quantitative research methods, and thus, collecting both qualitative and quantitative data in a single study (Johnson & Onwuegbuzie, 2004; Creswell & Plano Clark, 2007; Halcomb & Hickman, 2015; Shannon-Baker, 2016). Where researchers combine only qualitative or only quantitative approaches in a single study, this is referred to as multi-method research (see, for example (Falcao, Ferreira, Rodrigues,

Diniz, & Gašević, 2019)) in the context of ethics and privacy in LA where researchers combined focus group studies and interviews).

With mixed methods research, one can examine a problem in greater detail than would be possible using qualitative or quantitative research methods on their own and to obtain richer and deeper insights (Creswell & Plano Clark, 2007; Halcomb & Hickman, 2015). Mixed methods research also allows the researcher to build on the strengths of the selected approaches and address the weaknesses of qualitative and quantitative research when carried out on their own (Creswell & Plano Clark, 2007). For example, the analysis of qualitative research involves interpretation by the researcher, and consequently, it is necessary to take steps to acknowledge and/or minimise any biases, while quantitative research can miss out on the context surrounding or motivating participants' responses. A further advantage of mixed methods research is that one can seek to answer research questions which cannot be answered by qualitative or quantitative research on its own (Creswell & Plano Clark, 2007). For example, qualitative data can help to explain quantitative results. The limitations of mixed methods research include that it can be time intensive collecting and analysing both qualitative and quantitative data. Additionally, the researcher needs training in both approaches (Creswell & Plano Clark, 2007).

Given the complexity inherent in the research questions, this thesis research took a mixed methods approach. At the start, there was a need to understand how privacy was thought about in LA and what the key privacy issues were from the perspective of LA experts (Study 1; RQ 1 and RQ 2). This examination favoured the use of qualitative research methods as the participants' perspectives were important, given that their conceptualisation of privacy was of interest to address the research question. In Study 2 (RQ 3 – RQ 5) and Study 3 (RQ 6 - RQ 8), there was a further need to enhance the quantitative data, and thus, qualitative data was obtained. For Study 2, the focus was on whether students were concerned about privacy in LA and there was a need to also understand what motivated their concern and/or lack of concern. Similarly, in Study 3, there was a need to understand the motivation for participants' data use preferences. Finally, Study 4 (RQ 9) sought in-depth insights from qualitative data to understand the quantitative results from Study 3. Thus, mixing research methods for the thesis research allowed in-depth insights to be obtained on privacy and utility tensions in LA.

The next section discusses the philosophical assumptions underpinning mixed methods research.

3.3   Philosophical Assumptions of Mixed Methods Research

A worldview (also referred to as a paradigm, approach, or stance) refers to the assumptions and beliefs which guide the research questions researchers choose to study and the research methods they choose to use (Creswell & Plano Clark, 2007; Shannon-Baker, 2016). These assumptions and

beliefs cover what researchers view as the nature of reality (ontology), how they think knowledge is obtained (epistemology), and the methods they think can be used to obtain knowledge (methodology) (Twining, Heller, Nussbaum, & Tsai, 2017).

As discussed in the previous section, mixed methods research was deemed suitable for the work reported in this thesis due to the complex nature of the research questions and consequently the need for deeper insights to better understand the research problem. There are several worldviews associated with mixed methods research (Christ, 2013; Shannon-Baker, 2016) including pragmatism, critical realism, transformative, and dialectical. These worldviews have associated assumptions and beliefs around ontology, epistemology, and methodology, which are summarised in Table 3.5.

Table 3.5: Summary of philosophical assumptions for mixed methods research

Worldview	Pragmatism	Critical Realism	Transformative	Dialectical
Ontology	Multiple forms of reality	Different levels of reality exist from objective and independent of human understanding to subjective and understood through meaning making	Our world is multifaceted, thus there are different realities; inequalities exist thus there is a need to expose power structures and give voice to those oppressed	Pluralism; subjective and objective realities and their interrelationships
Epistemology	Perceptions of the researcher and participants can coexist in a single study	There are levels of objective truth about what is studied; absolute truth is impossible	Knowledge is constructed within the context of power and privilege; Expose inequalities and represent the oppressed; participants are collaborators	Multiple perspectives brought together based on the research questions, ethical needs, and goals; should meet identified values, goals, outcomes, and social and local needs
Methodology	Blend qualitative and quantitative data	Processes to construct and interpret objective and subjective data`	Participatory and action-oriented research to improve the wellbeing of individuals and society; Privilege the voice of the oppressed	Examine value issues in relation to the means and process of a study and the desired ends; make the values explicit

(Based on (Christ, 2013; Cram & Mertens, 2015; Shannon-Baker, 2016))

The *transformative* worldview focuses on the experiences of those who have been marginalised in society (Cram & Mertens, 2015), while taking a *dialectical* stance in research means that researchers use two or more paradigms together, using any tensions to bring about new insights (Shannon-Baker, 2016). A *critical realist* stance acknowledges that there are structures and mechanisms which may not be observed, yet are real, and these influence events and experiences (Christ, 2013). Finally, *pragmatism* places primary importance on the research question (Plano Clark & Badiee, 2010) and supports combining qualitative and quantitative approaches to build on their strengths while minimising their weaknesses (Shannon-Baker, 2016). Furthermore, pragmatism views theories as fitting in a given context and transferrable to other situations (Shannon-Baker, 2016), a perspective which proved useful for the thesis research given that the theories applied to the research were from privacy research, yet the work was in the context of LA. Given that the research questions were deemed of primary importance, multiple research methods were used to address the research questions, and the research applied theories from one research field to another, this thesis research fits within a pragmatic worldview (Johnson & Onwuegbuzie, 2004; Creswell & Plano Clark, 2007). Emphasis was placed on selecting approaches to address the research questions based on what was deemed suitable to address the research questions.

The next section discusses the research design and outlines how the identified research questions are linked to the associated research methods in the four thesis research studies.

3.4   Research Design

As discussed in Section 3.2, there are a variety of suitable research methods to address research questions on students and privacy in LA, including surveys, focus groups, interviews, and laboratory studies (Lazar, Feng, & Hochheiser, 2010; Blandford, 2013; Ifenthaler & Schumacher, 2016). As argued previously, a pragmatic worldview allows one to choose a suitable research method(s) to answer the research questions (Creswell & Plano Clark, 2007; Hesse-Biber, 2015). Therefore, the research methods selected for this thesis research were those considered appropriate to answer the research questions. The chosen research methods are shown in Figure 3.2, corresponding to the research questions.
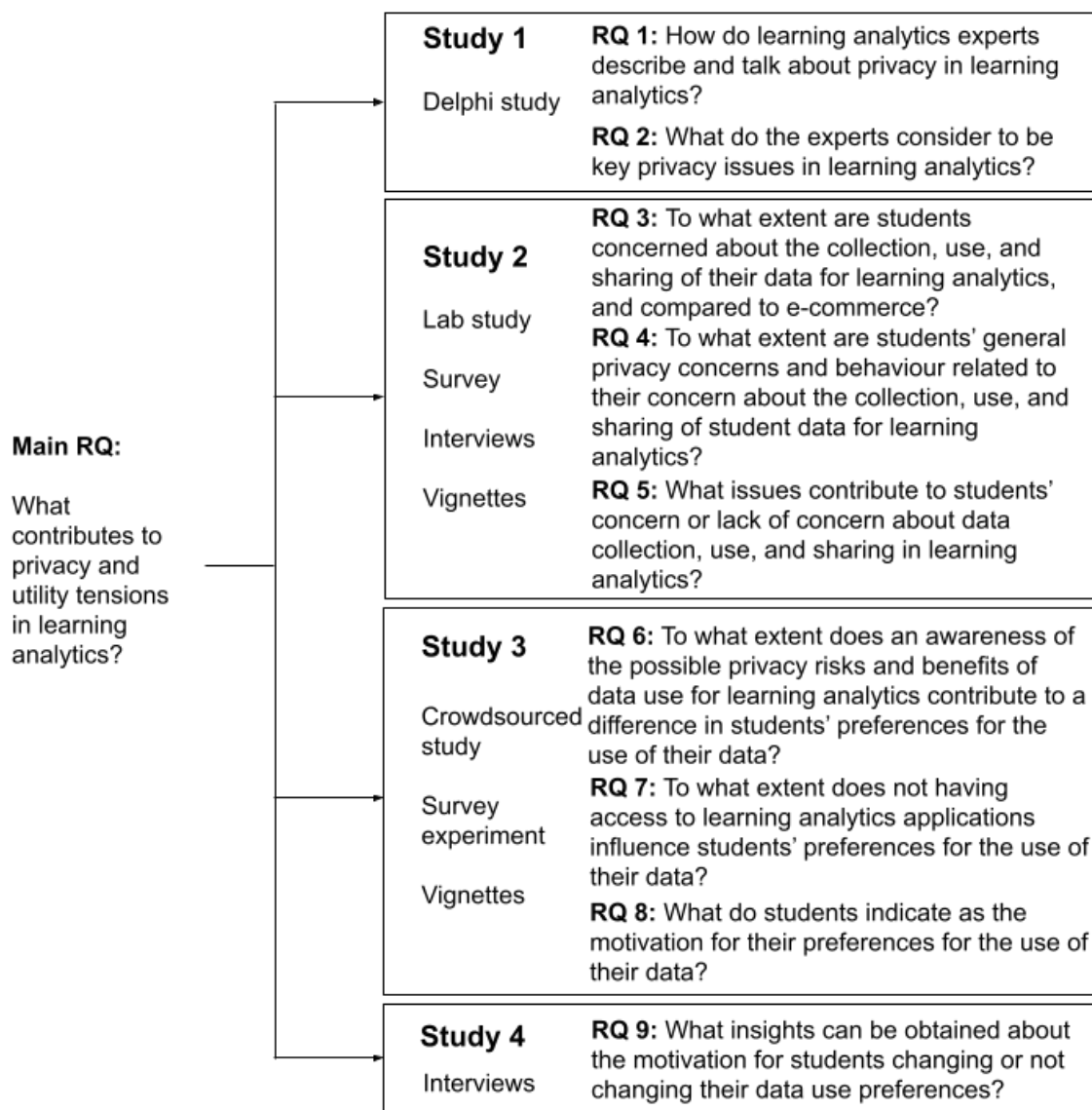
Figure 3.2: Research methods used in the thesis research corresponding to the study research questions

This section provides a summary of the research methods used in each of the four studies reported in this thesis. To minimise repetition of information, specific details of the research methods applied in each study are provided in each corresponding study chapter (from Chapters 4 through to 7).

Table 3.6 summarises the research methods that were used in each of the studies, the type of data collected, and the stakeholders who took part in each study.

Table 3.6: Summary of studies carried out highlighting the data collection techniques, type of data collected, and the stakeholders involved

Study	Research Method	Data Collection Technique	Qualitative / Quantitative Data	Stakeholder Focus
1	Delphi study	Semi-structured interviews	Qualitative	Experts
2	Laboratory study; Interviews; Vignettes	Survey questionnaire Semi-structured interviews	Qualitative + Quantitative	Students
3	Crowdsourced survey experiment; Vignettes	Experimental design; Survey questionnaire including open responses	Qualitative + Quantitative	
4	Interviews	Semi-structured interviews	Qualitative	

3.4.1   Study 1 – Examining Privacy Concepts

In Study 1 (reported in Chapter 4), Delphi study techniques were used to identify the privacy concepts held by LA experts, and what they considered as key privacy issues for LA. The Delphi technique was developed by Norman Dalkey and Olaf Helmer at the Rand corporation in the early 1950s (Williamson, 2002). The name 'delphi' referred to the famous Greek oracle who offered "visions of the future to those who sought advice" (Gupta & Clarke, 1996, p. 185). Originally applied to long-range forecasting (Gordon and Helmer-Hirschberg, 1964), the technique has developed over time and is now used for several purposes, such as to identify and prioritise issues, or to develop concepts and frameworks (Paré, Cameron, Poba-Nzaou, & Templier, 2013). In the field of educational technology, the Delphi technique has been applied in several areas, including to identify research areas in video-sharing technology (Snelson, Rice, & Wyzard, 2012) and to better understand the heuristics applied to instructional design (York & Ertmer, 2011).

Like any other research method, the Delphi study has its advantages and limitations (Murry & Hammons, 1995; Hasson, Keeney, & McKenna, 2000; Williamson, 2002; Keeney, Hasson, & McKenna, 2006; Hasson & Keeney, 2011; York & Ertmer, 2011; Plesch, Kaendler, Rummel, Wiedmann, & Spada, 2013). These are summarised in Table 3.7.

Table 3.7: Advantages and limitations of the Delphi study technique

Advantages	Limitations
Obtain experts' subjective judgements on an issue	Panel members must be chosen carefully
Anonymity provided to participants encourages non-conformity	Can face low response rates
Can extend a literature review by providing current knowledge on an issue	Risk that the researcher can introduce their views during analysis
	Can be time consuming

(Based on (Murry & Hammons, 1995; Hasson, Keeney, & McKenna, 2000; Williamson, 2002; Keeney, Hasson, & McKenna, 2006; Hasson & Keeney, 2011; York & Ertmer, 2011; Plesch, Kaendler, Rummel, Wiedmann, & Spada, 2013))

The Delphi technique allows one to collect experts' subjective judgements on novel or under-explored issues. The method encourages innovative thinking where it focuses on forecasting the future. Finally, as a group communication technique, the anonymity provided minimises pressure for the experts to conform to others' judgements or views. On the other hand, Delphi studies can be time-consuming. They can also suffer from low response-rates in the first and any subsequent rounds. One approach to mitigate this limitation is for the researcher to build rapport with the experts in the first round. Follow up emails can be sent as a reminder. While summarising the experts' responses, the researcher should take care to delineate their views from those of the participating experts. Finally, as the Delphi technique assumes that the experts have equivalent expertise and knowledge, the panel must be selected carefully. Given the method's strength at providing more up-to-date knowledge on an issue than can be obtained from a literature review (Delbecq, Ven, & Gustafson, 1975), the Delphi study technique was deemed suitable for Study 1.

Semi-structured interviews were carried out to collect data from the LA experts for Study 1 as part of the Delphi study. While interviews can be used to collect in-depth data, they take a lot of time and effort which limits the number of participants that can be recruited. Furthermore, interviewing is a skill that researchers need to develop. Steps were taken to mitigate some of the

limitations of interviews as a research method. First, the researcher carried out several practice rounds, interviewing members of the supervision team before meeting with the LA experts. Furthermore, participants were asked settling-in questions about themselves to help them be at ease at the start of the interview and to build rapport. Finally, a systematic approach was taken to analysing the qualitative data along with collaborative qualitative data analysis as discussed in Section 3.5.1.4.

3.4.2 Study 2 – Examining Privacy Concern

In Study 2 (reported in Chapter 5), a laboratory study was carried out as part of students' learning on organizational behaviour to understand whether they were concerned about privacy in LA, and to better understand the dimensions of their privacy concern or lack of privacy concern with the use of student data for LA. Laboratory studies and corresponding research as part of students' learning have been used in educational research (Rienties & Héliot, 2018) and LA studies in particular (Ifenthaler & Schumacher, 2016; Pijeira-Díaz, Drachsler, Järvelä, & Kirschner, 2016; Knight, et al., 2017; Mittelmeier, Rienties, Tempelaar, & Whitelock, 2018).

A survey questionnaire was selected as the data collection technique based on the aims of Study 2. Questionnaires used in existing research on privacy and privacy in LA were used, as detailed in Section 3.4.2.1. Additionally, semi-structured interviews were used to allow a deeper understanding of the motivation for participants' responses to the survey questions.

Participants were also provided with vignettes which detailed the way data can be collected and used for LA in a university context and in an e-commerce context. Vignettes depict situations in short story form to which study participants are invited to respond (Finch, 1987). They enable actions to be explored in a given context in a distanced and less personal way (Barter & Renold, 1999). Vignettes have been used extensively, both in privacy and human computer interaction research (Xu & Teo, 2004; Naeini, et al., 2017), and in education research (Mittelmeier, et al., 2018). The vignettes used in Study 2 allowed participants' responses to the two contexts to be compared.

The advantages and limitations of surveys as a research method have been detailed in Section 3.2.2. Additional limitations that were observed while carrying out Study 2 are that the questionnaires facilitated the collection of limited data and it was not possible to ask participants follow-up questions where interesting responses or trends were noted in the data. To mitigate some of the limitations, follow-up interviews were carried out with the aim of obtaining in-depth insights of the motivation for participants' responses. Additionally, participants were given the opportunity to provide open responses, thereby sharing the motivation for their responses. Finally, questions asking participants to recall events were avoided in the surveys.

3.4.2.1 Survey Questionnaires in Privacy Research

The process of designing survey questionnaires can be time consuming and care must be taken to ensure that the questionnaires measure what they are intended to measure (validity) and that there is consistency with repeated measurements in a study (reliability) (Gaskell & Bauer, 2000). It is good practice to use existing survey questionnaires, where possible, to increase the reliability and validity of the quantitative research (Sullivan & Calderwood, 2017). Using existing questionnaires also allows for comparison of the findings with other research studies. This thesis research used several existing questionnaires. In Study 2 the following three questionnaires were used: i) The privacy index questionnaire (Westin & Maurici, 1998); ii) the online privacy concern questionnaire (Buchanan, Paine, Joinson, & Reips, 2007); iii) the questionnaire on privacy concern in LA (Slade, Prinsloo, & Khalil, 2019). Study 3 used the Internet Users Information Privacy Concern Questionnaire (IUIPC) (Malhotra, Kim, & Agarwal, 2004). These questionnaires are detailed further alongside a discussion of the corresponding research studies in Chapter 4 to Chapter 6.

3.4.2.2 Translation of Survey Questionnaires

An existing questionnaire from research in ethics and privacy in LA was adapted for the thesis research (Ifenthaler & Schumacher, 2016). The original questionnaire was written in German and translated into English for the thesis research. Following recommendations in Bielick (2017), the translation process was a rigorous team effort with five translators working to translate the questions from German into English and from English back into German. The survey questions were comprehensively pilot tested to ensure that participants could understand the questions, and they fit into the educational context in the UK.

3.4.3 Study 3 – Examining Data Use Preferences

In Study 3 (reported in Chapter 6), there was a need to identify how students would respond to requests to use their data for LA when the potential privacy risks and benefits were made explicit. Additionally, Study 3 needed to investigate differences in students' acceptance of data use for LA given differing conditions. These two requirements made the *survey experiment* an ideal research method. A survey experiment investigates how a variable of interest influences another variable of interest using random assignment of participants into the experimental conditions (Gergle & Tan, 2014). Open responses from participants allowed insights into what motivated participants' data use preferences. A survey experiment combines, in one study, the advantages of surveys (such as being cost effective), with the advantages of experimental research (such as manipulating study conditions) (Yan, et al., 2018).

Experiments require many participants to allow for statistical analysis and generalisation of the findings to a given population. Survey experiments seek to establish generalisable causal relationships (Barabas & Jerit, 2010). While population-based survey samples are seen as ideal (Mullinix, Leeper, Druckman, & Freese, 2015), their costs can be prohibitive, hence the growing use of convenience samples such as those found on crowdsourcing platforms. Study 3 was conducted while the Covid-19 pandemic was ongoing. As a result, the study could not be conducted f-t-f due to health and safety concerns for the research participants and the researcher. This further promoted crowdsourcing as a suitable approach to allow recruitment of large groups of people to participate in the research. With crowdsourcing, many participants are recruited to complete small tasks for small payments or incentives (Egelman, Chi, & Dow, 2014). Research suggests that a sample of participants recruited from crowdsourcing platforms behaves in a similar way to a sample of participants recruited from a participant pool of undergraduate students (Behrend, Sharek, Meade, & Wiebe, 2011).

Participant recruitment and the study were conducted online using the Prolific platform, which is a crowdsourcing platform. Prolific acts as a subject pool for researchers carrying out experimental research (Palan & Schitter, 2018). In a study comparing participants from a university subject pool and two alternative crowdsourcing platforms (Amazon Mechanical Turk and Crowdflower), Peer and colleagues (2017) found that participants on Prolific had a higher response rate than the university subject pool. Additionally, participants on Prolific provided higher quality data than the university subject pool and participants on Crowdflower. Finally, participants on Prolific seemed to be less experienced as participants in experimental research compared to those on Amazon Mechanical Turk. They were also more diverse based on geographical location and ethnicity. Researchers using Prolific are encouraged to pay study participants the minimum wage (£7.52 per hour at the time of conducting Study 3). Prolific also meets the Open University's research data management standards as data are stored in the UK. These reasons demonstrated that Prolific was a suitable crowdsourcing platform for Study 3.

Prolific has been used for various relevant studies including to determine users' preferences for being notified about data use on their wearable devices (Murmann, Reinhardt, & Fischer-Hübner, 2019). Research conducted using crowdsourcing platforms is not a novelty in education, see for example (Follmer, Sperling, & Suen, 2017; Johnson & Borden, 2012) who discuss the use of Amazon Mechanical Turk for education research and use it for data collection. However, evidence for the use of Prolific Academic is sparse in the published literature in the education context, suggesting that it is slowly gaining ground.

Experimental research requires well-defined hypotheses that can be tested. There is a need to control confounding factors in experimental research, which can be challenging. In addition, the observed behaviour, for instance, in a laboratory study may not be users' typical behaviour. The unsupervised nature of tasks on crowdsourcing platforms requires researchers to implement techniques for quality control, such as attention check questions, to minimise the chance that participants will submit spurious data (Egelman, Chi, & Dow, 2014). Attention check questions were used in Study 3 as detailed in Chapter 6.

3.4.4   Study 4 – Examining Motivation for Data Use Preferences

Semi-structured interviews were carried out in Study 4 as they allowed in-depth insights to be sought from Study 3 participants to explain the quantitative results. In choosing to carry out semi-structured follow-up interviews, the researcher must choose what aspect of the quantitative data will be followed up on (Creswell & Plano Clark, 2007). As the variable of interest in Study 3 was students' data use preferences, and there was an expectation in designing Study 3 that the interventions would have an influence on participants' data use preferences, the decision was made to seek to better understand the motivation for any changes to participants' data use preferences.

The next section gives an overview of how the qualitative and quantitative data collected for the thesis research was analysed.

3.5   Data Analysis

Suitable qualitative and quantitative data analysis techniques were determined during the study design phase. This section details the qualitative data analysis techniques used in the thesis research.

3.5.1   Qualitative Data Analysis

Qualitative data analysis was carried out on the qualitative data collected for the four research studies, from Study 1 to Study 4. Furthermore, in Study 4 (reported in Chapter 7), data collection was carried out in two phases and data analysis was done concurrently, as recommended by Schutt (2017, p. 267).

Qualitative data can be analysed deductively or inductively (Burnard, Gill, Stewart, Treasure, & Chadwick, 2008). With the deductive approach, a pre-determined framework of themes is developed whether from theories or existing literature and this is used during the analysis process. While deductive analysis is quick, it can limit theory development by excluding themes arising from participants' responses. The inductive approach, in contrast, uses participants' responses to determine the themes and their structure. The inductive approach is comprehensive

and thus time-consuming; however, it serves to centre participants' perspectives. Consequently, the inductive approach was used to analyse the qualitative data collected in the thesis research.

An inductive approach to analyse the data was selected given that the codes are developed from participants' responses with the aim of mirroring what is in the data as opposed to the researcher's perspectives or views (Linneberg & Korsgaard, 2019). An additional strength lies in the ability to capture the diversity and complexity in the data. In contrast, the deductive approach uses a pre-existing list of theoretical codes or themes taken from literature. A deductive approach to qualitative data analysis has the advantage of reducing the complexity of the analysis and helps the researcher to remain focused (Linneberg & Korsgaard, 2019). A purely deductive approach was deemed unsuitable for the thesis research given the focus on identifying various factors which contribute to privacy and utility tensions. As a result, the aim was to be as expansive as possible, while a deductive approach risked restricting the research findings.

Saldaña (2016) discusses numerous qualitative data analysis methods which he divides into so-called first cycle and second cycle coding methods. In this thesis, both first and second cycle coding methods were used. The qualitative data analysis methods were chosen with the aim to use the best approach to understand the data and answer the research questions. This is in line with the pragmatic philosophy guiding the research as detailed in Section 3.3. The specific qualitative data analysis methods that were used are described below.

3.5.1.1   First Cycle Coding Methods

Descriptive coding techniques were used to analyse the qualitative data from Study 1 to Study 4 where the aim was to summarise, using a word or a phrase, the topic of a passage from participants' responses. Care was taken to summarise the topic of what a participant was referring to rather than summarise what they were saying.

In vivo coding was also used to analyse the qualitative data. Here, codes were selected based on terms participants used in their responses, that is the term chosen as a code could be found in a participant's response. This was important to do to keep the focus on participants and their responses, thereby reducing the researcher's influence on the work (Given, 2008).

Versus coding involves identifying somewhat binary situations arising from participants' statements which suggest situations of conflict. This data analysis technique was used to highlight situations where participants' statements highlighted or emphasised contrasting situations.

Themeing the data involves labelling participants' responses with a statement which captures the theme, rather than labelling the statement using a short code. A theme is "an extended phrase or sentence that identifies what a unit of data is about and/or what it means." (Saldaña, 2016, p.

74

199). With respect to participants' responses, a theme "brings meaning and identity to a recurrent experience and its variant manifestations" (DeSantis & Ugarriza, 2000) demonstrating that a theme is shared across participants' responses.

3.5.1.2   Methods that Fall Between First and Second Cycle Coding

Saldaña (2016, p. 212) identifies *eclectic coding* as a coding technique that falls in between first and second cycle methods. Eclectic coding combines first cycle coding methods as identified above with second cycle coding techniques, which are discussed in the next paragraph. Eclectic coding involves the researcher reflecting on participants' responses and the codes identified using first cycle coding methods, while applying methods from second cycle coding. In this way, the identified codes are brought together and unified to develop themes.

3.5.1.3   Second Cycle Coding Methods

The aim of second cycle coding methods is to further organise the codes and themes identified using the first cycle coding methods to identify themes which can explain what has been identified from participants' responses. The following two methods were applied in the thesis research as they were useful for directing the researcher towards identifying suitable themes from the codes and categories.

Focused coding involves combining codes that are similar to each other, perhaps touching on similar concepts. This also involves a process of determining which codes make the most sense.

Theoretical coding makes progress towards identifying the central or core category and consequently the main theme of the research.

3.5.1.4   Collaborative Qualitative Data Analysis

In collaborative qualitative data analysis, two or more researchers focus on and discuss a shared body of data to arrive at an agreed upon interpretation (Cornish, Gillespie, & Zittoun, 2013; Saldaña, 2016, p. 36). A key motive for carrying out collaborative qualitative data analysis was to tap into the different perspectives that the researcher and members of the supervision team contributed to the research project. These different perspectives came about due to the differing disciplinary backgrounds of the researcher and members of the supervision team, including in computing and human computer interaction, LA, educational technology, mathematics, and psychology. These different perspectives were sought in analysing the qualitative research reported in this thesis.

While there are various ways to enhance credibility in qualitative research, peer debriefing, which was originally proposed by Lincoln and Guba (1985), was used for the thesis research. In peer debriefing, the supervisors acted as a sounding board for the codes and themes identified in the

research. In this way, the credibility of the research results was enhanced (Spall, 1998). Details about the peer debriefing process and collaborative data analysis are provided in Chapter 4 through to Chapter 7.

Working with a team to collaboratively analyse qualitative data offers the advantage of broadening the scope and depth of the analysis as well as acting as a 'sense-check' through peer discussion (Saldaña, 2016). While calculating inter-coder reliability is one way for teams to analyse data collaboratively and enhance the reliability and trustworthiness of the results (Lazar, Feng, & Hochheiser, 2010), alternative approaches include inter-coder agreement and group consensus (Campbell, 2013). Both inter-coder reliability and inter-coder agreement were deemed suitable for the qualitative data collected for the thesis research due to the researcher's agreement with Saldaña's (2016) perspective of analysis as an interpretive activity. Multiple coders bring their own unique perspective to the data and consequently their interpretation of the data could differ (Williamson, Rooyen, Shuttleworth, Binnekade, & Scott, 2020). This serves to enhance the perspectives and therefore the richness of the analytic process (Cornish, Gillespie, & Zittoun, 2013), and further increases the trustworthiness of the results by involving more than one researcher in the data analysis process (Linneberg & Korsgaard, 2019). In line with the pragmatic stance taken for the thesis research (as detailed in Section 3.3), both approaches to seek agreement as well as multiple perspectives were deemed useful to address the thesis research questions and unpack the nuances and complexities in the data.

To demonstrate the credibility of the interpretation of the data in this thesis, the presentation of the data includes a demonstration of how the interpretation was reached. Additionally, the steps taken to analyse the data are continuously justified, following the approach proposed by Mason (2002). Examples of participants' responses alongside the identified themes and codes are provided in the corresponding study chapters to enhance the transparency of the data analysis process (Cornish, Gillespie, & Zittoun, 2013).

Mason argues against the suitability of the triangulation method where one uses different methods or data sources to investigate the same phenomenon with the aim of "getting an accurate reading or measurement of it" (Mason, 2002, p. 190). Instead, she indicates that different methods and data sources "shed light onto different social phenomena or research questions and provide different versions or 'levels' of an answer" (Mason, 2002, p. 190). This is the stance taken in this thesis research, that the different research methods, data collected, and analysis techniques employed allow for different perspectives to be obtained on privacy and utility tensions in LA.

3.5.2 Quantitative Data Analysis

Quantitative data analysis was carried out on quantitative data collected from Study 2 and Study 3. In Study 2, t-tests were used to answer RQ 3, comparing participants' concern across the e-commerce and LA contexts. In addition, correlation tests were used to answer RQ 4 as to how the study variables related to each other. In Study 3, correlation tests were carried out to assess whether and how the study variables influenced each other to answer RQ 6 and RQ 7. Study 3 investigated whether the privacy risk or benefit intervention influenced participants' data use preferences, their concern about data use for LA, as well as their perception of benefits from the use of student data for LA. Therefore, ANOVA tests were deemed suitable to examine whether there were any differences between the control and experimental groups (Strunk & Mwavita, 2020). As Study 3 implemented a pre-post test study design asking participants their data use preferences before and after the intervention, Chi-square tests were used to evaluate the statistical significance of changes students made to their data use preferences. To minimise repetition, and to enhance the clarity of the content discussed, details of the quantitative data analysis are provided in Chapter 5 and Chapter 6, alongside the research questions the studies sought to address.

The next section touches on the ethical issues that needed to be considered for the four thesis studies based on the stakeholders who took part.

3.6 Ethical Considerations

As discussed in Chapter 2, this thesis focused on two stakeholder groups: LA experts and students. The ethical considerations for research with these two stakeholder groups are discussed in this section.

3.6.1 Study with Learning Analytics Experts

Study 1 focused on identifying how LA experts (n=12) conceptualised privacy in LA and what they thought were key privacy issues. Ethical approval was sought from the university HREC. The project received a favourable opinion (HREC reference number - HREC/2826/Korir) shown in Appendix A.1.

The participants were people with considerable knowledge of ethics and privacy in LA or educational technology. They were either academic staff, researchers, or practitioners working in HEIs. As such, they were not considered vulnerable participants. In addition, Study 1 sought their views on ethics and privacy in LA. Since they are knowledgeable in the issues being discussed, it was unlikely that the topic under discussion would cause them any discomfort or harm.

The interviews were carried out remotely. Both the experts and the researcher were at a location of their choosing and convenience, therefore, personal safety was also not a concern. In addition, the participants could choose their preferred software to take part in the interview (Skype or Google Hangouts) and did not have to install new software which they were unfamiliar with or uncomfortable using. Participants were provided with an information sheet containing details of Study 1 and what it would focus on. They were asked to sign a consent form to participate in Study 1, and their additional consent was sought before the interviews were audio-recorded. One participant did not consent to an audio-recording of the interview and in this case, the researcher took notes during the discussion.

3.6.2   Studies with Students

The second stakeholder group the thesis research focused on was students in Study 2 (n=111), Study 3 (n=447), and Study 4 (n=15). These studies sought to identify and understand whether students were concerned about privacy in LA, how they conceptualised the privacy of their data in LA, as well as whether and how their data use preferences were influenced by an awareness of the potential privacy risks and benefits of LA, and the motivation for their data use preferences.

While in Study 2 the research sought to identify students' privacy conceptualisations, it is acknowledged that students may not have thought about the collection, use, and sharing of their data for LA. The study questions could have heightened their awareness of privacy issues and potentially made them uncomfortable if they thought that the privacy of their data was being violated. The studies (2, 3, and 4) could also have made students feel powerless to address any privacy concerns they may have had. On the other hand, recruiting participants for a study on privacy may have caused bias by raising their awareness prior to the study (Krol, Spring, Parkin, & Sasse, 2016).

These issues were addressed in several ways: Care was taken not to refer to privacy at the start of the studies (e.g., in recruitment material), and instead, participants were debriefed at the end. A participant information sheet was used to inform all participants about the studies, and that they were free to end their participation at any time if they felt uncomfortable, with no negative consequences for them. Furthermore, at the end of Study 2, participants were given more information on different privacy-preserving strategies that they could use if they wanted to protect their privacy and the privacy of their data.

Data for Study 2 was collected primarily f-t-f, while data for Study 3 and Study 4 was collected online. While initially plans were made for Study 3 and Study 4 to be conducted using f-t-f means, the onset of and recurring nature of the Covid-19 pandemic meant that this component of Study 3 had to be excluded. One advantage of conducting research studies online is that physical safety

for the researcher and students is not a concern, as both can participate in the study from a location of their choosing. The f-t-f element of Study 2 was carried out in a public place to address any physical safety concerns. Ethical approval for Study 2, 3, and 4 was obtained from the university's HREC (HREC/3033/Korir and HREC/3287/Korir). The ethical approval for Study 4 was obtained along with that for Study 3.

3.7   Researcher Reflexivity

In this work, reflexivity is taken as a critical reflection of the research practices (Nielsen, Randall, & Christensen, 2017) with the aim of identifying and acknowledging how the researcher contributed to the production of knowledge in this work (Kennett, O'Hagan, & Cezer, 2008). With its foundations in qualitative research (Finlay & Gough, 2008, p. 3), researcher reflexivity has also gained ground in mixed methods research and is seen as one way to enhance the quality of the work (Curry & Nunez-Smith, 2015).

My approach to conducting the research reported in this thesis was driven by my personal interests, background, training, and experience in privacy, computer science, and human computer interaction (HCI), working in academia and industry for over ten years. As a result, I am interested in understanding how users interact with various security and privacy products and services and seeing how these can be improved, while the application domain has been varied. In the context of this thesis research, the application domain is LA. Consequently, I am interested in identifying challenges users face in various interaction scenarios at the intersection of privacy and LA and designing solutions to address these challenges. This interest is also driven by a desire to amplify the users' voice (that is their preferences) in contexts where it ordinarily would be absent.

I take a pragmatic stance in this research, and I am driven to apply what works to solve the identified research problems. I am aware of tensions in bringing qualitative and quantitative research together for this work and I seek to identify these tensions and examine them for the lessons and insights they can provide to me and others. I also have an interest in my research bringing about positive change in the context under study. Thus, the work also has a focus on identifying and recommending action various stakeholders in HEIs can take considering the answers to the research questions.

3.8   Chapter Summary

This chapter has discussed the methodology driving the thesis research. The research design has also been detailed, identifying how the thesis research questions were addressed using four research studies. The next four chapters (Chapters 4 through to 7) detail the research methods

used to address the corresponding research questions and discuss the findings and their implications for privacy and utility tensions in LA.

4   Study 1 Methods and Results[5]

This thesis is comprised of four studies, and this chapter describes Study 1. Study 1 focused on understanding how privacy is conceptualised in LA, and identifying the key privacy issues, all from the perspective of experts in ethics, privacy, and LA. This chapter is divided into four sections. First, the rationale for Study 1 is described in Section 4.1, along with the research questions. This is followed in Section 4.2 by a detailed description of the methods used for Study 1, including identifying the experts who took part, how the study was carried out, the study instruments and the data analysis procedure. Section 4.3 reviews the findings of Study 1 in relation to the research questions. Finally, Section 4.4 discusses the implications of the findings, identifies the limitations of Study 1, and provides a rationale for Study 2.

4.1   Introduction

Ethics and privacy frameworks for LA were designed to fulfil several aims. One aim was to enhance understanding of the ethical and privacy issues arising from the use of student data (Slade & Prinsloo, 2013). Another aim was to guide HEIs to adopt LA in ethical ways which respect students' privacy (Drachsler & Greller, 2016; Pardo & Siemens, 2014). The ethics and privacy frameworks achieve these aims by systematically identifying ethics and privacy issues and detailing how HEIs can address them (Hoel & Chen, 2018; Drachsler & Greller, 2016; Sclater, 2016).

Many of the prominent and highly cited frameworks have largely focused on establishing an early understanding and consensus around broader ethical issues, where these are concerned with what is the right thing to do for privacy in the use of student data. However, there are only a handful of frameworks, (for example (Rubel & Jones, 2016)) which emphasise privacy alongside ethics issues. Thus, there was a need to understand how various stakeholders conceptualised privacy in the LA domain. Identifying and understanding stakeholders' thinking about privacy is important in contributing to the development of ethical LA. To add to this, the continued expansion of (inter)national legislation, for example, the UK General Data Protection Regulation (UK-GDPR)[6] has also led to a specific need for HEIs to better understand how privacy is thought about and understood in LA and what the key privacy issues are. Study 1 focused on addressing

---

[5] Sections of this chapter are adapted from the following article:

Korir, M. M., Slade, S., Holmes, W., and Rienties, B. (*Manuscript in preparation*). Privacy and learning analytics in the datafied university: Is there consensus among the experts?

[6] https://uk-gdpr.org/

these gaps, applying Delphi study techniques through remote semi-structured interviews involving twelve experts who had knowledge of ethics and privacy issues and had backgrounds in education, educational technology, and LA. Study 1 sought to answer the following research questions:

RQ 1: How do learning analytics experts describe and talk about privacy in learning analytics?

RQ 2: What do the experts consider to be key privacy issues in learning analytics?

The next section provides a detailed overview of the methods used in Study 1 to address these research questions.

4.2 Method

This section describes how the experts were identified and recruited to take part in Study 1, the protocol which guided how Study 1 was conducted, and how the data was analysed. The rationale for using Delphi study techniques for Study 1 are provided in Section 3.4. This study was considered low risk and received a favourable opinion from the Open University Human Research Ethics Committee – HREC Number: 2826 (see Appendix A.1 for the approval from the Ethics Committee).

4.2.1 Setting and Participants

To identify the experts to approach as potential study participants, a list of experts was created with authors of several ethics and privacy frameworks in LA. The initial list of experts was revised using background information about these experts obtained from different sources including Internet searches for publications and talks on ethics and privacy. However, several challenges were noted with using this approach to select expert participants. First, some of the researchers had Google scholar profiles or personal websites which listed their publications, while others did not have these available. This difference made it difficult to identify a standardised approach to evaluate an author's publications for a focus on ethics in LA. Second, LA was a developing field when the thesis research was carried out, with the first international conference being held in 2011. Consequently, several authors and researchers worked on LA in general while only a few carried out research primarily on ethics and privacy in LA. Third, a strict focus on academic publications or talks excluded practitioners who also had expertise in LA. Therefore, it was deemed more appropriate to supplement the initial list of experts to also include experts suggested by two members of the thesis supervision team. The thesis supervisors have published extensively on ethics, privacy, and LA and have attended conferences and workshops in this area, thereby allowing them to develop valuable knowledge to identify other experts (both researchers and practitioners) in the field.

A final list of twenty-four experts was developed. These experts were contacted by email and invited to take part in the study. A snowball sampling strategy (Oppenheim, 2000) was used where the experts were asked to forward the list to others who they thought might be interested in taking part. This strategy was implemented to broaden the pool of experts invited to participate. The invitation letter (shown in Appendix A.2) introduced the researcher and members of the thesis supervision team, detailed the purpose of Study 1, and invited the experts to respond by email to register their interest. Twelve experts confirmed their participation in the study (a 50% response rate). Those who opted not to participate indicated that they lacked the time to do so.

Table 4.1 provides background information about the twelve participants who took part in Study 1. It shows that they were from a total of seven countries, held different roles in industry and academia, and had different areas of expertise within education, educational technology and LA. Participants' expertise, as shown in Table 4.1, that was relevant for Study 1 was also due to whether they had published in LA or were part of a UK advisory group working with the Joint Information Systems Committee organisation (JISC) on a code of practice for LA (Sclater, 2016).

Table 4.1: Learning analytics experts who took part in Study 1

Name	Location	Job Title/Role	Area of Expertise Relevant for the Study	Experience in Ethics/Privacy	Host Institution during Study 1
Andrew Cormack	UK	Chief regulatory advisor	Learning analytics	Published on ethics and privacy in LA	JISC
Dai Griffiths	UK	Professor	Education/Educational cybernetics	Published on ethics and privacy in LA	University of Bolton
Jennifer Heath	Australia	Director	Student support and education analytics	Published on ethics and privacy in LA	University of Wollongong
Tore Hoel	Norway	Researcher	Learning technologies	Published on ethics and privacy in LA	Oslo Metropolitan University
Michael Kickmeier-Rust	Austria	Professor	Educational assessment	Published on ethics and privacy in LA	St. Gallen University of Teacher Education
Andrew Meikle	UK	Head of corporate information systems	Learning analytics	Participation in JISC code of practice	Lancaster University
Abelardo Pardo	Australia	Professor	Learning analytics	Published on ethics and privacy in LA	University of South Australia

Paul Prinsloo	South Africa	Professor / Educational researcher	Open and distance learning	Published on ethics and privacy in LA	University of South Africa
Maren Scheffel	Netherlands	Assistant professor	Learning analytics	Published on ethics and privacy in LA	Open University Netherlands
Yi-Shan Tsai	UK	Educational researcher	Learning analytics	Published on ethics and privacy in LA	University of Edinburgh
Rupert Ward	UK	Professor	Learning innovation	Participation in JISC code of practice	University of Huddersfield
James Willis III	USA	Consultant/Associate adjunct faculty	Digital badges	Published on ethics and privacy in LA	University of Indianapolis

4.2.2 Study Procedure

Remote semi-structured interviews were carried out with the participants. The interviews lasted between 30 and 45 minutes and were conducted using Skype or Google Hangouts. The experts were asked to describe their work in ethics, privacy, and/or LA, to share their thoughts on what student privacy in LA is, and what they thought were the key issues in privacy in LA.

The interviews were audio-recorded and transcribed for analysis. The experts' consent was obtained before the recording was made. One expert did not consent to an audio recording of the interview. In this case, notes of the key discussion points were taken during the interview.

4.2.3 Study Instruments

A questionnaire on privacy in LA was prepared and pilot tested with four experts with backgrounds in educational technology, ethics, privacy, and LA. The pilot test identified that the initial questions were too broad. This was addressed by refining the questions, for example, referring specifically to 'student privacy' in the higher education context rather than 'privacy' in general. In addition, the questions were rephrased to be more conversational to engage better with the experts during the interviews. The interview script used for Study 1 is shown in Appendix A.3.

4.2.4 Data Analysis

The transcripts were analysed using NVIVO version 11. Three transcripts were collaboratively coded with members of the thesis supervision team to take advantage of different approaches to analysing and interpreting the data, and of coding with others (Saldaña, 2016). Section 3.5.1.4 presents an analysis of the role of collaborative coding for enhancing credibility and trustworthiness in qualitative data analysis. The codes were reviewed for areas of agreement and disagreement. Discussions were held during the analysis process to clarify issues arising from the experts' responses and to identify new insights from the data. The remaining transcripts were then analysed by themeing the data and applying eclectic codes (Saldaña, 2016). The qualitative data analysis techniques are discussed in Section 3.5.1.

4.3 Results

This section presents the results of Study 1 in response to the research questions. First, privacy concepts were identified from the experts' responses to address the first research question about how they conceptualise privacy in LA. Second, key student privacy-related issues raised by the experts were identified to address the second research question.

4.3.1 Privacy Concepts

The privacy concepts discussed below illustrate how the experts who took part in Study 1 thought about student privacy in the context of LA and what they thought it to be. In particular, the experts referred to several established privacy concepts, including, privacy as a right, privacy as control over personal information (Westin, 1967), contextual integrity (Nissenbaum, 2010), and an understanding of privacy as linked to one's identity (Floridi, 2005). The latter two concepts are more recent theories of privacy (Heath, 2014), and they aim to give a better understanding of why students might view some actions as a breach of their privacy, and the extent to which privacy breaches are perceived to have a negative impact on students. The findings related to the experts' privacy concepts are discussed in detail in this section. In discussing these findings, emphasis was placed on identifying and elaborating how ethical LA can be developed from the experts' perspectives and to meet students' needs. This approach allowed avenues for future research to be identified, some of which were pursued in subsequent studies in this thesis.

4.3.1.1 Privacy as a Right

The first privacy concept from the experts was the view of privacy as a *human right*, as in the Universal Declaration of Human Rights (UN General Assembly, 1948):

> '...everyone has a right to privacy … and it's meant to protect an individual about … anything they think, feel, [or] write. It should not automatically be public knowledge whatever a person utters whether it is written or with words or by click streams, actions or anything'
>
> *Participant E5*

From the perspective of designing ethical LA applications, the view of privacy as a right suggests that privacy in LA should be provided to all students by default and that no student should be deprived of it. In this way, privacy is available to all students and not only to those who are willing or able to expend effort to ensure it. This expert's way of thinking about privacy in LA also suggests the need to protect the individual student by enabling them to learn without concern about being watched, or their actions being tracked.

Another expert described privacy as the *right to be left alone* as seen in the classic and highly influential work of Warren and Brandeis (1890). This way of thinking about privacy was linked to the possibility that students might desire to opt out of data collection and/or analysis for LA once they are informed about what data is collected and how it is used:

> 'I think that a student should have the right to be left alone if he or she chooses. I think that if a complete picture is drawn for the student by saying we are collecting your data in

order to send you a signal or let you know that something is askew so we can intervene
and help you, I think a student should have the moral and legal right to say I don't want to
look at that and I don't want to participate in that.'

Participant E7

This perspective of privacy as the right to be left alone suggests that students should have a say in whether to participate in LA. It also captures the need for transparency, where HEIs provide students with information about what data is being collected, how it is being used and why, and respect their informed decision (where this is implemented) not to participate. However, it is important to note that the practical implementation of an opt-out position could result in some students missing out on study opportunities, such as recommendations of study resources. HEIs are required to collect certain datasets relating to students, for operational and regulatory purposes (Griffiths, 2017). Any suggested restriction on data collection related to opt-out would need to carefully specify what data would be excluded and what would continue to be collected. In this respect, Cormack's (2016) data protection framework might be appropriate. It distinguishes between two stages of LA - analysis and intervention. Cormack suggests that in the analysis stage, HEIs can process student data in aggregate form, for various reasons including to improve course offerings and make other improvements within a course, department or at the institutional level. In the intervention stage, he suggests that informed consent should be sought from students, allowing them to opt-in to receive interventions that result from LA.

Finally, privacy was viewed as students' *right to have a say on whether and how their data is used*:

'… a student's right to choose how their data is used and not used.'

Participant E1

This particular privacy concept aligns with the view of privacy as control over personal information (Solove, 2009). It suggests giving students a way to indicate to the university how they think their data can and cannot be used and in this regard is closely related to the previous privacy concept in allowing students to opt-out of the use of their data for LA.

Overall, three of the experts conceptualised privacy as a right in the ways described in this section. In doing so, their perceptions differed, however, they promote a unified design for ethical LA, offering privacy by default, promoting transparency by HEIs towards students with regards to the use of their data, and suggesting that opt-out measures be investigated further.

4.3.1.2 The Context Matters for Privacy

The context surrounding data collection and use (for example, the time when student data is collected, what activity is ongoing during data collection, and when students are informed and/or their consent is sought) was important to determine students' expectations and to explain why students might consider certain uses of data as breaching their privacy:

> *'Privacy is a breach of contextual norms. It's when something happens to your data that you didn't expect, and the expectations are very dependent on [the] context within which the data process is taking place.'*
>
> *Participant E6*

Emphasising the context surrounding collection and use of student data is aligned to the view of privacy through the lens of contextual integrity (Nissenbaum, 2010). The context influences the expectations students might have around data collection and use. As such, students might view any data collection and use which does not meet these expectations to be a breach of their privacy.

Furthermore, it was thought that certain practices of data collection and use might be acceptable to students in one context, but not in another:

> *'On the other hand, we share more and more, and it depends on context. What I share on Facebook I don't necessarily want my lecturer to know and vice versa.'*
>
> *Participant E11*

This response demonstrates a differentiation between content for students' learning and content that they share or engage with for social purposes.

Overall, four of the experts referred, in different ways, to the importance of context in understanding privacy in LA, thereby making it the privacy concept most experts agreed on in Study 1. This result could be explained by the view that contextual integrity is relevant for understanding privacy in LA, for example, Heath (2014) demonstrates how it can be used to evaluate privacy considerations in the use of student data. Hoel, Mason and Chen (2015) additionally emphasise that context goes beyond thinking only about technology or industry, for example, but should extend to include social elements such as students' and other actors' perspectives.

4.3.1.3 Data as a Representation of Student Identity

The third privacy concept involved raising awareness of the inherent and possibly overlooked value of data. The experts indicated that data can be used to represent students in a way that

gives insight into who they appear to be as individuals. One expert expressed the view that student data might represent students' identity:

> '… my data is not something I can give away. Data is part of who I am… It is actually the student's whole persona that's captured in the data.'
>
> *Participant E11*

This understanding of privacy was proposed by Floridi (2005) in a bid to extend existing privacy definitions and further enhance understanding of privacy. By viewing privacy as identity, it becomes clear how unsettling data loss or unauthorised data use can be for students, even though no physical item belonging to the student is taken:

> '*I can see in a privacy sense that someone when they lose control of their data in the age that we live in… have lost control of something of themselves.'*
>
> *Participant E7*

Given that this perspective links to who students are and how they view themselves, the expectation that students will want control over the use of their data is unsurprising. By controlling who has access to their data, one can then control who has access to details about who they are as individuals.

Overall, two of the experts referred, in different ways, to the understanding of privacy in LA from the perspective of the insights offered into students' identity.

4.3.2   Key Privacy Issues

As indicated in Section 4.2.4, the experts' responses were analysed iteratively using thematic analysis. Three themes were identified from their responses. Each of these themes is discussed in turn in this section.

4.3.2.1   *Empowering the Student to Manage their Privacy*

The first key privacy issue identified focused on how HEIs can empower students to enable them to manage privacy, including the privacy of their data. Various approaches were mentioned by the experts, including through data ownership, control over various aspects related to their data, and informed consent for student data to be used in LA.

Data Ownership: There was an expectation that students should own the data:

> '*There's a perspective that the institution never own the data. Students keep sovereignty over the data and they allow access to the data, but the ownership of data and what can happen with the data belongs in the locus of control of the student.'*

The view was expressed that students also thought that they owned the data. This resulted from design decisions (motivated for instance by ethical practice or legislation) to seek students' consent before they use LA systems:

> *'But however, having said that, it doesn't mean that students don't think that they own the academic data. They think they own it… To students having the option to give consent in a way it is acknowledging that they have ownership of the data.'*

> *Participant E12*

The expectation that students would have ownership over their data was extended to include the perspective that they would own any meaning derived from the data:

> *'…It is their data. Yes, they study through the institution, but the data belongs to them. The meaning of that data belongs to them and not to us.'*

> *Participant E11*

Such a view is in keeping with the understanding that data is a representation of the student, and therefore, any meaning derived from the data also gives insight into who the student is. However, it was also acknowledged, that in some cases, HEIs or third parties which provide LA services would own the data:

> *'…as institutions in global south move online and make use of learning analytics, most of our student data will be owned by commercial companies and then these will most probably sell our student data to the first or the highest bidder.'*

> *Participant E11*

It is likely that institutions can own student data because data protection legislation is either absent, under development, or remains unenforced. As argued above, such practices can violate students' privacy as their data (viewed as part of who they are) can be sold for institutions' commercial gain.

Issues of data ownership are set out in legislation or in institutional (data use) policies. Corrin et al. (2019, p. 10) describe data ownership as the "possession of, control of, and responsibility for information." Pointing to the lack of clarity on this issue, they conclude, as did Greller and Drachsler in their ethics and privacy framework (2012, p. 50) that the owner of the data collection tool, often also the data client and beneficiary, owns the data. In contrast, Slade and Tait (2019, p. 7) suggest that the institution should act as a temporary steward of the data, subject to

institutional policies and legislation. Thus, different solutions are available in this case and therefore these guidelines need to be referenced to determine best practice for LA, and what, if any, changes should be made.

Student Control over Data: While control over personal information is one existing privacy concept (Westin, 1967), it was acknowledged that students have little say over how their data is collected, used, and shared:

> *'At the moment in an educational institution typically, there's very little control of the data by the student. The student generates the data, but they have no control over it.'*
>
> *Participant E9*

Technical features of LA systems would need to be set up to allow for fine-grained control over what data is collected, when, and how it is used:

> *'…we are giving you an opportunity to tell us who you want to share the data with, or whether there is anyone at all who you want to share it with, and ways you would like it to be used.'*
>
> *Participant E12*

While this is likely useful for students as demonstrated in work by Ifenthaler and Schumacher (2016), acceptance with various stakeholders, including students themselves should be examined. In addition, it is important to consider the overall impact and any practical issues that might arise from enabling student control over data use for LA. For example, Solove (2009) argues that while individuals' privacy preferences and considerations are important, they need to be balanced against societal needs. Additionally, giving students opportunities to control who has access to their data might lead to unexpected results. For example, one expert shared that their own research found that students might expect that their data is kept private from their teachers, yet doing so might complicate (or render impossible) the teachers' work:

> *'I do not want my teacher or my professor to be able to see the analysis or the data.'*
>
> *Participant E5*

This requirement has been expressed in other research (Slade & Prinsloo, 2014), and, as indicated by the authors, might be linked to a desire that students have to control aspects of their learning or study, or concerns that the information might bias teachers against them.

While enabling students to control who has access to their data and how it is used, it is also important to consider that HEIs need to use student data for operational purposes (Griffiths,

2017). Therefore, students might not be able to opt out of data collection entirely or might not be able to do so without some consequence for the operational functioning of the institution, especially if large numbers of students do so (Li, Brooks, & Schaub, 2019). As an alternative, the possibility of accommodating differing levels of engagement was raised, as well as an opportunity for students to correct any misunderstanding arising from the use of their data:

> *'…they can decide whether they want to use the learning analytics or not. Or whether they want to have it displayed prominently or not. And that when they feel misrepresented in whatever they see that they have a context that they can go to. So, they can basically adapt their data or at least say, look something must have gone wrong. This is not how it went.'*

> *Participant E5*

The feasibility and acceptance of such an approach would need to be tested with various stakeholders. The extent to which students are willing to engage with controlling the use of their data remains to be seen. Furthermore, there are the previously stated considerations of the impact of providing students with opt-out opportunities.

Informed Consent: As in Griffiths' work (2017), and as stated previously, there is a general view that HEIs must collect student data for operational purposes:

> *'…some of the data simply really is needed in order to process a student through university life.'*

> *Participant E5*

Certain personal information is required from students for them to access features related to their studies that are due to them, for example, course credits which cannot be issued to students if they are anonymous (Amigud, Arnedo-Moreno, Daradoumis, & Guerrero-Roldan, 2018). Thus, certain features related to learning require that students can be identified, and students' informed consent is needed to collect and process student data (Prinsloo & Slade, 2018).

Informed consent can be sought at different stages. For instance, students can be informed about what data is collected and how it is used at the point when they register as students at a HEI, or once they are registered as students, when they sign up for a particular course, module, or unit. Students' informed consent can also be sought before sharing insights from LA as argued in the ethics and privacy framework by Cormack (2016).

> *'The moment you personalise any intervention or make assumptions about me as an individual, I think that's where my concerns would kick in. So, in the contract students*

93

have with an institution … we give permission for a certain range of data to be harvested… the moment you adapt my learning environment or the amount of support I get, or my trajectory, the moment you start to change my environment, you need my further permission, and I may opt out of that.'

Participant E11

The point in time when students are asked for their consent is important. During registration, students might be overwhelmed by other registration requirements and as a result might not give full attention to what they are asked to do:

'But also, partly because when they enter the university it is part of the enrolment process. So, most people would just consider it as one of the tasks that they are trying to finish. Tick, tick, tick. So, they know that they gave consent, but they did not know exactly what they gave consent to.'

Participant E12

It is unclear if students know where to go to get information about what they have consented to in their consent agreements or what to do to update them (Prinsloo & Slade, 2018). Such information could aid in steps to achieving informed consent.

Asking for students' consent to share the findings of LA s is not without its challenges. By asking for consent after analysis has taken place, students are made aware that certain information about them exists which will remain unknown to them unless they consent to receiving it. Therefore, such an approach raises questions whether the consent obtained is free, that is, that students do not feel coerced to give it.

4.3.2.2   Expecting Institutions to Act Ethically

The second key privacy issue raised by the experts focused on their expectation of ethical behaviour from HEIs. More specifically, the experts wanted the HEIs to protect students' data, comply with existing legislation, and be transparent with students about LA practices.

Data Protection: HEIs were expected to use different approaches to protect students' data, including by adopting a data minimisation strategy where institutions would first identify what data was available and whether it was sufficient for their purposes, before attempting to collect more data:

'One is to reconsider what data we already have and how we use that data. So, I have a suspicion that data are scattered across institutions and used by various people for various purposes. And the data is stored in different formats. So, before we think of adding data or

> *looking for more data, I think it's critical that we consider what data we already have*
> *access to.'*

> *Participant E11*

There were also limits set that certain data was not collected or that data would be allowed to deteriorate over time after it had served the purposes for which it was collected:

> *'There's the data hygiene point of view of what we should and should not collect… privacy*
> *is about conscious design of what data is gathered and how that data deteriorates. So, I*
> *can see in the example I've given there's probably a reason why you need to collect that*
> *data in order to debug the system when it won't let somebody in but there's no*
> *operational reason why you need to keep that for more than a week. Or a day, or half an*
> *hour.'*

> *Participant E9*

Such an approach might require revising LA away from an approach which seeks to collect and use as much data as possible, even for potential future uses.

Legislation: At the time when these semi-structured interviews were conducted with the experts, the impact of the European Union's and subsequently the UK's General Data Protection Regulation (UK-GDPR) was as yet unclear, but it was expected to change practice in LA:

> *'There's things like GDPR which are going to impact significantly on institutions in terms of*
> *what consent is sought… So, you know, explicit rather than implicit… And as we've said*
> *with GDPR it's more about opting in.'*

> *Participant E1*

The experts expressed mixed reactions to the GDPR. They viewed it positively as being a safer position for HEIs:

> *'…the GDPR and the data protection regulation of the EU… is [a] very good opportunity in*
> *order to ask ourselves what does that mean for the design of learning analytics solutions.*
> *And I don't think that we have really had that discussion in this community yet.'*

> *Participant E2*

On the other hand, the GDPR was also viewed as a burden, adding onto requirements for students and HEIs to meet:

> *'I think it's challenging because it's going to make it harder for learning analytics because it will be, you know, there's clearly a barrier there that people have to actively consent to.'*
>
> *Participant E1*

GDPR was also considered as a challenge for LA, especially in smaller institutions which were expected to struggle to meet the requirements:

> *'And if we are facing these new EU data protection regulations, everything becomes very difficult. It in fact is a showstopper for learning analytics, for collecting data. Specifically, if we are talking about smaller organisations. The requirements that are put on smaller organisations, they are very strict… They actually would need lawyers, which they cannot pay, to make sure that whatever they are doing, digital data, is in line with these data protection regulations.'*
>
> *Participant E4*

Transparency: The experts held the view that HEIs can increase awareness of LA and communicate with students and other stakeholders about the practices surrounding data collection and use.

One issue that was raised focused on students' awareness of the use of their data for LA as the experts thought that students were unaware of university practices around data collection and use in LA:

> *'…students do not know what the university is doing even though the university may be doing everything legally within a legitimate framework.'*
>
> *Participant E12*

This concern is shared in related work (Roberts, Howell, Seaman, & Gibson, 2016; Jones, et al., 2020). Therefore, transparency is useful to increase student awareness of data collection and use in LA (Pardo & Siemens, 2014):

> *'I think the main issue really is awareness or transparency so that many times students don't know what is collected about them and who has access to that data. And this just needs to be made more transparent.'*
>
> *Participant E5*

It was suggested that background processes be made transparent to enhance students' understanding:

> *'…make these processes of aggregating data as transparent as possible to the users, specifically the students and to make them understand what is going on to allow them to understand how their assessments have been made.'*
>
> *Participant E4*

However, in ensuring that HEIs are transparent about LA, care should also be taken not to overwhelm students with too much information.

A second issue that was raised had to do with the way LA initiatives were communicated to students. The experts focused on specific ways through which students could be made aware of LA, that is, how HEIs can be transparent. Reference was made to using institutional policies to inform students:

> *'…how we communicate what we are doing with student data in the university. Especially through policy. So, every university would have the data protection policy… They might or they might not have the policy dedicated to learning analytics, but they would have the policy which deals with how we use student data.'*
>
> *Participant E12*

This refers to one of the roles that institutional (LA) data use policies can play, which is to make clear to students what institutions are doing with their data. Policies, which can be used in HEIs to inform students of data practices, have several challenges, including whether they are designed to be understood by students (Jones, 2019). A suggestion was made to develop separate policies with one specifically designed to be clear to students:

> *'Often when we are developing a policy, the policy is not developed for students to read… I think the audience we have in our minds might be academics, mainly staff. So, I think that might be something we should consider when we are developing a policy. How can we better communicate this to students? ... Should we have a separate set of policies which communicate this main policy in a different way?'*
>
> *Participant E12*

With respect to the content of the policies, it was argued that the information should be clear and should be presented in a way that is easy to understand:

> *'They would appreciate information about this in a succinct way. So, it could be written succinctly. But they also appreciate short meetings where they could have the opportunity to ask questions.'*

Finally, the need to communicate effectively with students, and to additionally offer them advice and guidance on how to interpret output from LA, was emphasised:

> *'So, it's how do you communicate in an effective way, and how do you put tools and techniques in place that enable the learner to act on it in an effective way.'*

> *Participant E1*

4.3.2.3   Acknowledging Conflicting Objectives

The third key privacy issue focused on acknowledging the potentially conflicting objectives held by students and HEIs. Through this issue, the experts recognised that both these stakeholders have objectives that may not be aligned and therefore, there is a need to identify these objectives and consider whether they can be met to the satisfaction of both parties.

This echoes the work of Parrish and Richman (2020, p. 11) who identify the need for shared goals between administrative and faculty members in implementing LA in higher education.

Balance Expectations: It was thought that some students might not want to participate in LA. A recurring narrative in Study 1 was that opting out of data collection in HEIs might not be possible because the university uses the data to provide students with a service. However, students might be able to opt out of the analysis or intervention phase of LA as proposed in the ethics and privacy framework by Cormack (2016). Excluding students' data from a data set was viewed as challenging:

> *'It is difficult to completely eradicate a student from data logs and to really give them that right to privacy.'*

> *Participant E5*

It was noted that excluding students' data from a dataset largely depended on the HEI's offering to students. A potential compromise was identified where HEIs could use aggregated data for LA:

> *'If an institution considers that certain data is crucial to maintain then they should explore compromises … for example, population measures …'*

> *Participant E10*

While arrangements could be made to allow students to opt out of data collection, this could potentially impact the types of LA HEIs could offer to students. Rather than a personalised offering, HEIs could provide a more generalised offering to students. However, Slade and Prinsloo (2014) demonstrated that students desire more relevant interactions with HEIs which take their

circumstances into consideration. This further demonstrates the differing expectations that need to be accounted for.

4.4   Study 1 Discussion

4.4.1   Implications of Findings

Study 1 sought to identify how privacy is conceptualised in LA, and what the key privacy issues are. To answer the research questions, Study 1 used Delphi study techniques which involved remote semi-structured interviews carried out with 12 education, educational technology, and LA experts from seven countries. The diverse and rich backgrounds of these experts contributed valuable insights into Study 1's research questions and enabled the identification of other research questions for subsequent studies in this thesis as well as for future work.

Study 1 aimed to contribute to knowledge about privacy and how it is conceptualised in the application of LA. Furthermore, Study 1 aimed to use the experts' extensive knowledge of ethics and privacy issues in LA to critically analyse, build on, and extend the issues expressed in existing literature. As a result, Study 1 helped to direct the focus of subsequent thesis studies.

The experts who took part in Study 1 were observed to hold different conceptualisations of privacy. Several concepts of privacy from prominent privacy literature (as discussed in Chapter 2) were identified, including, privacy as a right, privacy as control over personal information, privacy as contextual integrity, and privacy as identity. While the experts did share some of the privacy concepts, they had different perspectives, which also suggested different implications for the design of ethical LA. For example, while the concept of privacy as a right was shared by four experts, one expert (E5) viewed it as a human right, another (E7) as the right to be left alone, while two experts (E1 and E9) expressed it as students' right to have a say in how their data is used. This observation raises questions about how privacy in LA can be achieved if there is little consensus on its conceptualisation in this context.

It may be the case that institutional stakeholders involved in the implementation of LA applications hold differing privacy concepts, and differing expectations of whether privacy should be attended to, and how. Therefore, it is necessary to determine which privacy concept(s) will guide the process to implement LA applications in HEIs. For example, as demonstrated by Heath (2014), if contextual integrity (Nissenbaum, 2010) is selected as the privacy concept for use, then the norms, actors, and transmission principles guiding the flow of information have to be identified in order to guide steps towards minimising privacy violations.

Whichever privacy concept(s) is (are) selected to guide the implementation of LA applications, it is also important to include students' perspectives. Despite holding different privacy concepts, the

99

experts agreed that there was a need to enhance the students' role and engage them more actively to manage their privacy. Prinsloo and Slade (2015) make a similar call for HEIs to provide students with opportunities to opt-in or opt-out of the use of their data. They also emphasise the HEIs' responsibility of duty and care to its students. Students in the study by Sun, Mhaidli, Watel, Brooks and Schaub (2019) suggested different ways to enhance privacy in the use of a LA dashboard, for example, through students themselves managing who has access to their data. Further research is necessary to determine the feasibility of this approach in the educational context, given the scope to use such work to improve student data literacy (Knox, 2017).

The experts also identified several key privacy issues. A narrative which was observed across the three themes was that the HEIs bear the responsibility to realise these issues. For example, they can make available opportunities for students to manage their privacy in the LA context. While this is done, it also important to minimise the effort students have to expend so as not to burden them with activities requiring a high amount of effort.

4.4.2 Limitations

This study used one elicitation round where the experts provided their views on privacy and LA. Subsequent rounds seeking consensus were deemed unsuitable for this study, given that the research aims were met which included updating current knowledge on privacy in LA and obtaining insight into the different privacy concepts held by the LA experts.

One limitation of Delphi studies is that the researcher can add their own views and perspectives of the questions being researched while summarising the experts' views. To mitigate the risk of this happening, the data analysis process was conducted methodically, ensuring that a link was retained between the concepts and themes and the corresponding experts' responses.

Another limitation was that there might be alternative ways to categorise the data reported in this chapter apart from the categories used to detail the themes discussed in Section 4.3. However, effort was made to present the issues as raised by the experts who took part in Study 1, thereby addressing the research questions.

4.4.3 Links to Other Studies in This Thesis

One of the key privacy issues raised by the experts who took part in Study 1 was the need to empower students to better manage their privacy. Three issues were identified under this theme: data ownership, student control over data, and informed consent. Issues related to data ownership require legal and institutional interventions and were therefore not considered for further research. However, the remaining issues of student control over data and informed consent emphasised the need to engage with students more on how their data is used for LA. This

finding motivated the focus on students in subsequent studies in this thesis. Study 2, which is reported in the next chapter, focused on identifying whether students are concerned about the collection, use, and sharing of data in LA and understanding the dimensions and motivations for privacy concern and lack of privacy concern. Study 3, which is reported in Chapter 6 investigated the impact of providing students with the opportunity to control the data that is used for LA. Finally, Study 4 which is reported in Chapter 7 focused on understanding the motivation for students' data use preferences.

5   Study 2 Methods and Results[7]

This chapter presents the second study in this thesis which focused on investigating students' concerns about the collection, use, and sharing of student data for LA. Study 2 was motivated by mixed results in the research literature as to whether students are concerned about the collection, use, and sharing of their data in the LA context as indicated in Section 2.4. In this chapter, Section 5.1 introduces Study 2 and the research questions which guided the conduct of the study. Section 5.2 discusses how Study 2 was carried out and provides information about the setting and participants, the study design and procedure, the study instruments, and the data analysis procedure. Additionally, it gives details about follow-up interviews which were also carried out as part of Study 2. Section 5.3 provides an overview of the findings of Study 2 in relation to the research questions. Finally, Section 5.4 reviews the implications of the findings of Study 2, identifies limitations and links the work carried out to other studies reported in this thesis.

5.1   Introduction

One key privacy issue identified by LA experts in Study 1 (discussed in Chapter 4) was the need to empower students to manage the privacy of their data in the LA context. This finding guided the thesis research to further engage with students to understand their perspectives of privacy in LA. Additionally, as discussed in Section 2.4, there are mixed results in the findings of empirical research on students and privacy in LA as to whether students are concerned about privacy and the use of their data in LA. Consequently, Study 2 was designed to address this lack of clarity and seek deeper insights to better understand findings in the literature. To this end, Study 2 focused on identifying whether and how students' general privacy attitudes and behaviours are related to their concern about the collection, use, and sharing of their data in LA, whether and how the different phases from collection to use and sharing might contribute to students' privacy concerns, and the motivation for any concern or lack of concern students might have. Research on students and privacy in LA, as seen in Section 2.4, is yet to comprehensively address these gaps.

Contextual integrity (Nissenbaum, 2010) indicates that concern about collection and use of data can vary in different contexts. Research suggests that students have high levels of trust in the university (Slade, Prinsloo, & Khalil, 2019) compared to e-commerce companies (Jones, et al.,

---

[7] This chapter is adapted from the following submitted journal article:
Korir, M. M., Slade, S., Holmes, W., Héliot, Y., and Rienties, B. (Submitted). Investigating the dimensions of students' privacy concern in the collection, use, and sharing of data for Learning Analytics. Computers in Human Behavior Reports.

2020). There is also an expectation that universities will use student data in appropriate ways that will not harm students (Jones, Rubel, & LeClere, 2020). Therefore, two different contexts were used in Study 2 – collection and use of data in an e-commerce context and the LA context – to develop a clearer understanding of students' expectations and concerns (or the lack thereof). Study 2 sought to address the following research questions:

RQ 3: To what extent are students concerned about the collection, use, and sharing of their data for learning analytics, and compared to e-commerce?

RQ 4: To what extent are students' general privacy concerns and behaviour related to their concern about the collection, use, and sharing of student data for learning analytics?

RQ 5: What issues contribute to students' concern or lack of concern about data collection, use, and sharing in learning analytics?

The next section provides an overview of the methods used in Study 2 to answer the research questions.

5.2   Method

This section describes how Study 2 was carried out with students in a laboratory session and in follow-up interviews. The suitability of a mixed methods approach for Study 2 was discussed in Section 3.3, and the suitability of the research methods was discussed in Section 3.4. In this chapter, Section 5.2.1 describes the setting where Study 2 was carried out and the participants who were recruited. This is followed, in Section 5.2.2, by a discussion of the study design and procedure. The instruments used to collect study data are described in Section 5.2.3 and the design of the follow-up interviews is discussed in Section 5.2.4. Finally, Section 5.2.5 discusses how the Study 2 data was analysed. Study 2 received ethical approval from the Open University HREC (Number: 3033).

5.2.1   Setting and Participants

Study 2 was conducted at the business school of a UK university during the autumn semester of 2018. Students were studying a Masters' module in Organisational Behaviour. A total of 143 students were registered for the Organizational Behaviour module and 111 took part in the laboratory session. Of these, the majority were female (n = 90, 81%). Two students did not indicate their gender. The average age was 23.1 (SD = 1.9), and the ages ranged from 21 to 31 years. Three students did not indicate their age.

The GLOBE country cluster system (House, Hanges, Javidan, Dorfman, & Gupta, 2004) was used to categorise students according to their region of origin as there were several countries with only

one or two students (see Table 5.1). Most of the students were from the Confucian Asian (73 –
65.7%), Anglo (16 – 14.4%), and Southern Asian (10 – 9.01%) clusters. The large proportion of
students with international backgrounds was typical for postgraduate courses in the field of
business and management[8].

Table 5.1: Descriptive statistics of students' cultural backgrounds and nationalities

Cluster	No. of students	Percentage	Countries and no. of students for each
Confucian Asian	73	65.7	China (67), Taiwan (5), Hong Kong (1)
Anglo	16	14.4	UK (13), USA (3)
Southern Asia	10	9.01	India (5), Malaysia (2), Thailand (2), Vietnam (1)
Eastern Europe	6	5.41	Greece (5), Slovak (1)
Germanic Europe	2	1.8	Austria (1), Netherlands (1)
Sub-Saharan Africa	2	1.8	Nigeria (1), Tanzania (1)
Latin Europe	1	0.9	Italy (1)
Middle East	1	0.9	Turkey (1)

5.2.2   Study Design and Procedure

Masters' students studying Organisational Behaviour (n=111) took part in a laboratory session and
follow-up semi-structured interviews. During the laboratory session, students answered questions
about their general privacy behaviour and privacy concerns (the study protocol is described in
Section 5.2.2.2). In addition, they answered several questions based on two vignettes. The study
questions are detailed in Section 5.2.2.1. After answering the privacy questionnaires and
questions focusing on the vignettes, students participated in group discussions to enhance their

---

[8] Data obtained from the UK Higher Education Statistics Agency - https://www.hesa.ac.uk/data-and-analysis/students/what-study

learning on the topic being studied as seen in educational research (Rienties & Héliot, 2018) and LA studies (Pijeira-Díaz, Drachsler, Järvelä, & Kirschner, 2016; Knight, et al., 2017; Mittelmeier, Rienties, Tempelaar, & Whitelock, 2018). After a four-month break (due to holidays and students' exams), semi-structured follow-up interviews were carried out. The design of Study 2 is shown in Figure 5.1.
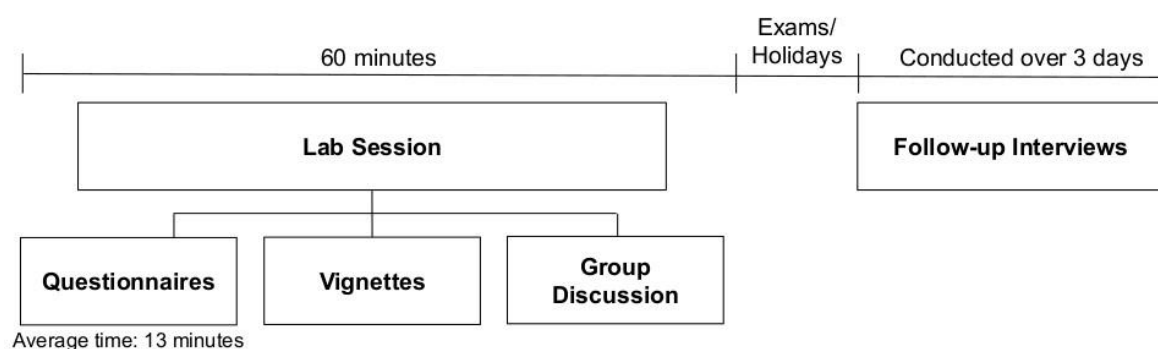
Figure 5.1: The design of Study 2

Vignettes were used to guide the discussion with students. This section first provides a description of the vignettes followed by the study protocol.

5.2.2.1   Vignettes

Students were shown two vignettes to explore whether they were concerned about the collection, use, and sharing of their data, comparing e-commerce and LA contexts. The first vignette shown to participants was based on Amazon[9], an American technology company offering its services in many countries around the world. This study focused only on its e-commerce services. The first vignette, which was read out loud to students in the laboratory session, is shown below:

> *Amazon is an e-commerce company that a number of you might be familiar with. It provides a personalised user experience, suggesting potentially relevant purchases based on your browsing and purchasing history. Please answer the questions that follow about the Amazon vignette.*

The second vignette was based on a student-facing LA dashboard (SFLAD). It described a hypothetical situation where student-facing LA was introduced to students at the university. Students were shown screenshots based on OU Analyse, a predictive LA system which has been adopted on a large scale at the Open University, UK. While OU Analyse focuses on students at risk of failing or of not submitting the next assessment, this study focused on the student activity

---

[9] www.amazon.co.uk

recommender feature which recommends resources that students are yet to interact with and that will help them prepare for the next assessment (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015). The text of the second vignette, which was also read out loud to students in the laboratory session, is shown below:

> *The University plans to roll out dashboards to help students keep track of their learning progress in individual modules and courses of study. The dashboards will be created using individual student data, data from their peers, and data from students who took the module in the past. Individual student data will include their performance on various assessments, their attendance to the classes, as well as their personal data provided at registration. You will now see a screenshot of the proposed system. Please review the screenshot, imagining you are the student referred to, and answer the following questions.*

Both vignettes were designed to be realistic, relevant, and easy for the students to relate to. Amazon was considered familiar to students as it offers incentives that are specifically relevant for students including reductions on book prices as well as free subscription to a next-day delivery service for one year. This assumption was later verified in Study 2 as all students stated that they had an account with Amazon. Similarly, the SFLAD vignette was relevant to the students as it described how they could keep track of their learning progress and receive personalised recommendations for learning resources.

The two vignettes shared similar characteristics as they focused on the provision of personalised services for students. They were of differing lengths, and while the Amazon vignette was realistic, the SFLAD vignette was hypothetical, since at the time of carrying out this research there were no publicised plans to unilaterally introduce LA at this particular university.

5.2.2.2 Study Protocol

Students in Study 2 attended a one-hour interactive laboratory session. They were briefed on the study and provided with an information sheet (See Appendix B.2) to review before the laboratory session began. The information sheet contained details about Study 2, and informed students about their rights, including that they could withdraw from the study with no negative effect on their course participation or grades.

Students provided their consent to participate in the study. They then filled out the privacy index (Kumaraguru & Cranor, 2005) and online privacy concern questionnaires (Buchanan, Paine,

Joinson, & Reips, 2007) using a survey tool from JISC[10]. The questions are discussed in Section 5.2.3. Students then engaged with the Amazon and SFLAD vignettes and were prompted to answer several questions to assess their privacy concerns with various uses of data in each context. Screenshots of the Amazon and SFLAD interfaces used in Study 2 are shown in Appendix B.1.

The free version of the PollEV software[11] was used to collect data during the interactive part of the laboratory session. PollEV has been used successfully to improve student engagement in lectures and classrooms (Kappers & Cutler, 2014) and was therefore appropriate to use in Study 2. As students engaged with the vignettes, their responses to the different questions were displayed as graphs on the screen. Students then had a brief discussion session with their peers to: (i) share their thoughts on collection and use of data as described in the vignettes, (ii) reflect on why they thought the way they did and find out what members of their group thought and why, and (iii) explore whether they and their peers had similar or different personality profiles. The latter two steps were linked to students' learning for the Organisational Behaviour module. Finally, the students were debriefed, and further discussions were held to relate the work carried out in the laboratory session to their learning on personality and organisational data practices for the Organisational Behaviour module (Rienties & Héliot, 2018).

Shortly after the laboratory session, students' responses to the privacy index (Kumaraguru & Cranor, 2005) and online privacy concern questionnaires (Buchanan, Paine, Joinson, & Reips, 2007) were used to generate a personalised privacy profile which was sent as feedback to each student who took part in the study. The privacy profile was especially useful as students were shown their scores in response to the questions and were provided with additional resources on privacy, thereby providing further opportunities for learning. A sample of the feedback given to students is shown in Appendix B.3. During the laboratory session, students were invited to register to take part in follow-up semi-structured interviews. Fifty students indicated an interest in participating, suggesting that they were interested in the topic and found it relevant to their learning.

5.2.3   Study Instruments

The privacy index questionnaire (Kumaraguru & Cranor, 2005) was used to segment students into three categories according to their level of privacy concern. Definitions of the three categories are provided below, summarised from work of Kumaraguru and Cranor (2005) and Iachello and Hong

---

[10] www.onlinesurveys.ac.uk
[11] www.pollev.com

(2007): **Privacy fundamentalists:** These users rank high in their privacy concern. They think that privacy has a high value and that organisations do not need as much personal data as they seek to collect from users; **Privacy pragmatists:** These users rank medium in their privacy concern. They acknowledge that there are risks present in organisations collecting and using customer data but think that adequate safeguards are present; **Privacy unconcerned:** These users rank low in their privacy concern. They are not worried about privacy and have little concern about providing their personal information to organisations. The questionnaire consists of three items and has been used in numerous studies on privacy concern, for example, (Olson, Grudin, & Horvitz, 2005; Iachello & Hong, 2007; Preibusch, 2013; Woodruff, Pihur, Consolvo, Brandimarte, & Acquisti, 2014). The response scale was modified to include a neutral option following best practice in questionnaire design (Krosnick, 2018). A limitation of the scale was noted from related work (Woodruff, Pihur, Consolvo, Brandimarte, & Acquisti, 2014) with respect to the validity of the scale and this was later identified in the Study 2 results. It was observed that the privacy index scale had a very poor reliability score at .177. For this reason, the privacy segments were only used to indicate students' privacy profiles as in related work, for example (Taylor, 2003; Kobsa, 2007), and were not used for further analysis in Study 2.

The online privacy concern questionnaires (Buchanan, Paine, Joinson, & Reips, 2007) was used to determine students' privacy behaviour and general privacy concern. The questionnaire is divided into three scales: general caution, technical protection, and privacy concern. It has also been used in numerous studies, see for example (Coles-Kemp & Kani-Zabihi, 2010; Woodruff, Pihur, Consolvo, Brandimarte, & Acquisti, 2014; Lee, Wong, Oh, & Chang, 2019).

The questions for the interactive laboratory session were adapted from work by Slade, Prinsloo, and Khalil (2019). These questions focus on students' perspectives of data collection, use, and sharing for LA. Examples of the questions used in Study 2 are shown in Table 5.2, while all the questions are shown in Appendix B.4.

Table 5.2: Examples of questions used in Study 2

Scale	N items	Example item	Response scale	M	SD	Alpha
Privacy index	3	Consumers have lost all control over how personal information is collected and used by companies	[1] Strongly disagree - [5] Strongly agree	3.34	.714	.177
General caution	6	Do you shred/burn your personal documents when you are disposing of them?	[1] Never - [5] Always	2.82	.86	.755
Technical protection	5	Do you watch for ways to control what people send you online (such as check boxes that allow you to opt-in or opt-out of certain offers)?	[1] Never - [5] Always	2.95	.8	.665
Privacy concern	10	In general, how concerned are you about your privacy while you are using the Internet?	[1] Never - [5] Always	3.58	.66	.836
Concern about data collection and use - Amazon	6	I feel comfortable that Amazon can offer me a better service (e.g., offers based on my buying or search patterns) by collecting my personal data?	[1] Totally disagree - [5] Totally agree	2.61	.457	.590
Concern about data collection and use - University	6	I feel comfortable that the University shares my personal and online activity data, in a personally identifiable way, with third parties?	[1] Totally disagree - [5] Totally agree	2.77	.485	.668

5.2.4   Follow-up Interviews

After data from the laboratory session was analysed, 41 students (out of the 50 who had volunteered to participate) were contacted for the follow-up interviews. These 41 students were selected to meet two criteria: first, that they had responded to most of the study questions, and

second, that they represented different privacy segments based on their responses to the privacy index questionnaire. Of those students contacted to take part in the follow-up interviews, four were classified as privacy fundamentalists, thirty-two as privacy pragmatists, and five as privacy unconcerned. The aim of the follow-up interviews was to gain deeper insights into the motivation for students' individual responses to the questions in the laboratory session. The interview schedule is shown in Appendix B.5.

It was not possible to interview all students as some did not respond to the invitation or were no longer able to participate due to end of year holidays followed by an examination period. In total, four students were interviewed. Three were privacy pragmatists and one was privacy unconcerned. The follow up interviews did not aim to obtain a representative sample, instead the focus was on obtaining insights into students' motivations with representation from different privacy segments. The findings, which are discussed in Chapter 5.3.4, offer insights from two out of three of the identified privacy segments. However, given that only four students took part, the insights from the follow-up interviews are preliminary, and point out areas for further investigation with a larger group of participants.

5.2.5   Data Analysis

T-tests were used to analyse the data to answer RQ 3, comparing participants' concern across the e-commerce and LA contexts. In addition, correlation tests were used to analyse the data to answer RQ 4 as to how the study variables related to each other. Finally, qualitative data analysis techniques, as described in Section 3.5.1 were used to analyse the data from the follow-up interviews to answer RQ 5. Multiple rounds of qualitative data analysis were carried out. First, the open responses were analysed using *eclectic coding* and by *themeing the data*, after which *versus coding* was used to better highlight nuances in students' responses. Finally, the data was analysed with each participant considered as a case study and their responses analysed in detail and compared to other participants' responses across similar and differing privacy segments.

5.2.5.1   Missing Data

A total of 111 students attended the laboratory session. All 111 students filled out the privacy index and Buchanan questionnaires and provided their demographic data. However, some data was not collected during the interactive laboratory session. Several issues contributed to the missing data. First, the free version of the PollEV software allowed a maximum of 40 participants per session. As the first two groups of students had slightly over 40 participants, some students were unable to take part in the poll. Second, since the laboratory session was scheduled for one hour, there was little opportunity to wait for extended periods of time after each question for all students to respond, and it was challenging to keep track of who was yet to respond to a

question. Third, there was data loss during the data download process. Finally, some data was lost as students used different identifiers across the two data collection tools (PollEV and JISC online surveys), and thus their data across the two data sets could not be combined for analysis. On average, 37 students did not respond to the Amazon and university questions. Missing values were replaced with the mean value calculated from participants' responses to a question following best practice recommendations (Groves, et al., 2009).

5.3   Results

Most of the students (80 – 72%) were privacy pragmatists, and thus ranked medium in their privacy concern. Of the remaining students, 16 (14.4%) were privacy unconcerned and 15 (13.5%) were privacy fundamentalists, indicating that they had a high level of privacy concern. These results are largely similar to those reported in other literature (Kobsa, 2007; Taylor, 2003). Other findings from Study 2 are presented in this section.

5.3.1.1   *Factor Analysis*

Principal components analysis with direct oblimin rotation was carried out on the questions focused on collection, use, and sharing of data in the university context. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .659. Bartlett's test of sphericity was significant (chi-square = 101.713; df = 15 p<.001). Two components were identified, explaining 57% of the variance. The first component had an eigenvalue of 2.28 (corresponding to 38.1% of the variance), the second component had an eigenvalue of 1.17 (corresponding to 19.5% of the variance). The first factor was related to comfort with data use and data sharing and the second factor to comfort with benefits for tracking. The factors and components are shown in Table 5.3.

Table 5.3: Factor analysis results for comfort with collection and use of data - University

Factor	Items and loading	Proportion variance	Alpha	Mean	SD
Factor 1 – comfort with data use and data sharing	… can offer me a better service by collecting my personal data (.80); … shares my data in a personally identifiable way (.73); … shares my data in an anonymised format (.69)	38.1%	.61	2.94	.54
Factor 2 – comfort with benefits for tracking	… offers benefits for tracking (.84); offers specific benefits for tracking and data shared (.82); offers specific benefits for tracking and data not shared (.49)	19.5%	.61	2.59	.63

Similarly, factor analysis using principal component analysis with direct oblimin rotation was carried out on the questions focused on collection and use of data in the Amazon context. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .594. Bartlett's test of sphericity was significant (chi-square = 88.802; df = 15; p<.001). Three components were identified, explaining 71.5% of the variance. The first component had an eigenvalue of 2.06 (corresponding to 34.3% of the variance), the second component had an eigenvalue of 1.16 (corresponding to 19.3% of the variance), while the third component had an eigenvalue of 1.08 (corresponding to 18% of the variance). The first factor was related to students' comfort with benefits for tracking with no data sharing, the second to comfort with benefits for tracking and identifiable data sharing, and the third to comfort with data use and anonymised data sharing. The factors and components are shown in Table 5.4.

Table 5.4: Factor analysis results for comfort with collection and use of data - Amazon

Factor	Items and loading	Proportion variance	Alpha	Mean	SD
Factor 1 – comfort with benefits for tracking and no data sharing	… offers benefits for tracking (.808); offers specific benefits for tracking and data not shared (.918)	34.3%	.70	2.86	.60
Factor 2 – comfort with benefits for tracking and identifiable data sharing	… shares my data in a personally identifiable way (.842); offers specific benefits for tracking and data shared (.598)	19.3%	.39	1.93	.56
Factor 3 – comfort with data use and anonymised data sharing	… can offer me a better service by collecting my personal data (.549); shares my data in an anonymised format (.934)	18%	.42	2.71	.65

The factor analysis of the questions across the two contexts identified different factors. This pointed to the need to further validate the scale and is highlighted as one of the limitations of the work (refer to Section 5.4.2). In presenting the remaining results in this section, only the factors identified from the questions focused on collection and use of data in the university context will be used further.

5.3.2 The Extent of Students' Concern about the Collection and Use of Data

RQ 3 was "to what extent are students concerned about the collection, use, and sharing of their data for learning analytics, and compared to e-commerce?" The mean values from participants' responses to the questions on collection, use, and sharing of their data in both the Amazon and university scenarios were obtained, as seen in Table 5.5.

Table 5.5: Mean and standard deviation results for individual Items in the Amazon and University scales

I feel comfortable that ….	Amazon		University	
	Mean	Std. dev.	Mean	Std. dev.
… can offer me a better service (e.g., offers based on my buying or search patterns) by collecting my personal data	3.11	.84	3.93	.64
… shares my personal and online activity data, in a personally identifiable way, with third parties	1.82	.72	2.08	.76
… shares my personal and online activity data, in an anonymised format, with third parties	2.31	.8	2.81	.77
… offers me specific benefits in exchange for tracking me online	2.68	.9	2.71	.93
… offers me specific benefits in exchange for tracking me online and assures me that my data will not be shared with third parties	3.68	.82	3.14	.95
… offers me specific benefits in exchange for tracking me online on condition that my data will be shared with third parties	2.04	.71	1.93	.63

Students seemed to be more comfortable with the university rather than Amazon carrying out the following activities: the collection of their personal data (mean = 3.93), sharing of their personal and online activity data with third parties in an anonymised format (mean = 2.81), and in a personally identifiable way (mean = 2.08), and being offered specific benefits in exchange for being tracked online (mean = 2.71). This could result from the collection of personal data in the educational context being more familiar to them. However, students were observed to be less comfortable with their data being shared with third parties by the university compared to Amazon (mean = 1.93). This result may be because students are unaware who the third parties are and may think that they are influential entities, such as future employers.

In the Amazon context, students were more comfortable with Amazon offering them specific benefits in exchange for being tracked on the condition that their data was not shared with third

parties (mean = 3.68) and least comfortable with Amazon sharing their personal and online activity data in a personally identifiable way with third parties (mean = 1.82). There was a small difference between the two contexts with respect to students' comfort with specific benefits in exchange for being tracked online, thus their comfort levels in both contexts were comparable in this instance.

A single scale was derived from participants' responses separately for the Amazon questions (Cronbach's alpha = 0.59; Mean = 2.61; SD = .46) and the university questions (Cronbach's alpha = 0.67; Mean = 2.77; SD = .49). A Shapiro-Wilk's test revealed that the data was not normally distributed (W = .940 Amazon; W = .936 University; p < .000). Therefore, a Wilcoxon's signed-rank test was carried out to determine if there were any differences in the median values between participants' responses to the Amazon and university questions. A statistically significant difference was observed (Z = -3.463; p < .001), suggesting that overall, participants were more comfortable with the collection and use of data in the university context than in the Amazon context. This result might be because students have greater trust in the university than Amazon, or they are more familiar or comfortable with the practice of the collection and use of their data in the university context.

Furthermore, students indicated that they were comfortable with sharing their data with their tutors so that the tutors could support them better. While 18 students (17.1%) disagreed to their data being shared with their tutors, 60 students (57.2%) indicated that they were comfortable with this practice, while 26 students (24.8%) were neutral. This finding aligns to other research (Vu, Adkins, & Henderson, 2019) which has shown that students are comfortable with sharing their data where the recipient and the purposes are known and related to their learning.

5.3.3 Students' General Privacy Concerns and Behaviour and Concern about Collection and Use of Data

RQ 4 was "to what extent are students' general privacy concerns and behaviour related to their concern about the collection and use of student data for learning analytics?"

'Hiding a bank PIN when using cash machines/making purchases' and 'shredding personal documents when disposing of them' were the most practiced activities on the general caution scale (mean = 4.05 and 3.23, respectively), while 'reading a website's privacy policy' and 'reading licence agreements fully before agreeing to them' was the least practiced activity (mean =2.07 and 1.99, respectively). The most practiced technical protection activities for participants in Study 2 were 'watching for ways to control what one is sent online' and 'using pop-up window blockers' (mean = 3.39 and 3.19, respectively), while the least practiced technical protection activities were 'checking one's computer for spyware' and 'removing cookies' (mean = 2.65 and 2.59,

respectively). In both responses, it might be the case that participants did not know what cookies or spyware were or did not know how they could be removed. Finally, with the privacy concern scale, the activities leading to the highest privacy concerns related to 'someone intercepting a credit card while one is buying something on the Internet', or 'one being mischarged when buying something on the Internet using the credit card' (mean = 3.96 and 3.92, respectively), and the activities with the least concern involved 'information about one being found on an old computer' and 'someone gaining access to the student's electronic medical records' (mean = 3.32 and 3.05, respectively). Participants' responses to the general caution, technical protection, and privacy concern scales seemed to be as expected.

Shapiro-Wilk's tests showed that the general caution, technical protection, and privacy concern scales were normally distributed while the scales identified from a factor analysis of the university-related questions - comfort with data use and data sharing (Factor 1) and comfort with benefits for tracking (Factor 2) - were not. Therefore, Spearman correlations were used on all the scales. The results alongside mean, standard deviation and normality results for the different Study 2 scales are shown in Table 5.6.

Table 5.6: Mean, standard deviation, normality results and Spearman correlations for Study 2 scales

	Mean	SD	Shapiro-Wilk	Sig.	1	2	3	4	5
General caution	2.82	.819	.980	.086	1				
Technical protection	2.95	.795	.979	.080	.449**	1			
Privacy concern	3.58	.657	.989	.483	.200*	.117	1		
Comfort with data use and sharing_F1	2.94	.635	.920	.000	.067	.128	-.017	1	
Comfort with benefits for tracking_F2	2.59	.485	.935	.000	-.090	-.136	-.095	.361**	1

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed)

The results suggested that students who carried out the technical protection activities also carried out activities related to general caution (r = .449; p < .01). Those students who had high privacy concerns carried out more general caution activities (r = .200; p < .05). Finally, those who were comfortable with data use and data sharing for LA were also comfortable receiving benefits in exchange for tracking in the university context (r = .361; p < .01).

5.3.4 Issues Contributing to Students' Concern about Collection and Use of Data

Follow-up semi-structured interviews were carried out with 4 students who took part in the initial study, with the aim to answer RQ 5: "What issues contribute to students' concern or lack of concern about data collection, use, and sharing of student data for learning analytics?" The students had an average age of 23.5. Further demographic details about the 4 participants along with their responses to some of the study questions (A/U1 – A/U6) are shown in Table 5.7.

Table 5.7: Demographic information of students who participated in the follow-up interviews and their responses to a selection of relevant study questions

PID	Age	Gender	Nationality	Privacy segment	A1	A2	A3	A4	A5	A6	U1	U2	U3	U4	U5	U6
1	24	Male	Greek	Privacy pragmatist	Totally Agree	Neutral	Neutral	Disagree	Disagree	Disagree	Agree	Disagree	Disagree	Disagree	Disagree	Totally Disagree
2	22	Female	Indian	Privacy unconcerned	Agree	Disagree	Agree	Neutral	Agree	Totally Disagree	Totally Agree	Disagree	Agree	Agree	Agree	Disagree
3	24	Male	Greek	Privacy pragmatist	Agree	Neutral	Agree	Totally Disagree	Disagree	Disagree	Totally Agree	Agree	Agree	Agree	Agree	Neutral
4	24	Female	Chinese	Privacy pragmatist	Agree	Totally Disagree	Totally Disagree	Disagree	Agree	Mis.	Agree	Mis.	Mis.	Agree	Agree	Disagree

A1 - I feel comfortable that Amazon can offer me a better service (e.g. offers based on my buying or search patterns) by collecting my personal data; A2 - I feel comfortable that Amazon shares my personal and online activity data, in a personally identifiable way, with third parties; A3 - I feel comfortable that Amazon shares my personal and online activity data, in an anonymised format, with third parties; A4 - I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online; A5 - I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online and assures me that my data will not be shared with third parties; A6 - I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online on condition that my data will be shared with third parties; S1 - I feel comfortable that the University can offer me a better service (e.g. alerts on potential problems or recommendations of learning resources) by collecting my personal data for use in the student learning dashboard; S2 - I feel comfortable that the University shares my personal and online activity data, in a personally identifiable way, with third parties; S3 - I feel comfortable that the University shares my personal and online activity data, in an anonymised format, with third parties; S4 - I feel comfortable that the University offers me specific benefits in exchange for tracking me online; S5 - I feel comfortable that the University offers me specific benefits in exchange for tracking me online and assures me that my data will not be shared with third parties; S6 - I feel comfortable that the University offers me specific benefits in exchange for tracking me online on condition that my data will be shared with third parties; Mis – Missing data

This section discusses the preliminary insights identified from these four students' responses to the follow-up questions. While this section contains a comparative analysis of the views held by the four participants, other views might be held by students who were not able to attend the follow-up interviews. Consequently, effort is made to present the themes that were shared by the students, rather than those stated by only one student.

5.3.4.1 Relationship with the University and Corresponding Trust or Lack of Trust

Students' relationship with the university influenced how they perceived institutional use of student data. Both Participant 1 and Participant 2 were willing to share data based on their *relationship* with the university. Participant 1 expected that the university would have and therefore would use students' data by virtue of the student-university relationship. In fact, for this student, this seemed to be a foregone conclusion:

> "And for the [University name]" part, I mean, I'm their student. They are supposed to have my data. I don't have a problem with that at all."

> [Participant 1, Privacy pragmatist]

The relationship between the student and the university was noted to contribute to the student developing trust in the university. Participant 2 was more supportive of personalised services from the university compared to Amazon:

> "I think it's because [University name] is something that is really close to me right now. Amazon, I'm not. So, I would want to believe that I can rely on my institute more. And obviously when it comes to my privacy and everything. But Amazon is not something that I'm connected to."

> [Participant 2, Privacy unconcerned]

Participant 3 and 4 also stated that they *trusted* the university to handle their data appropriately and not to students' detriment. Consequently, Participant 3 stated that they were more comfortable being tracked online by the university than by Amazon. However, the student expressed that there were limits to the influence they expected the university to have:

> "So, when I think about tracking is, I don't know, maybe on my location, or what I do, what I search on Amazon, so yeah, it's like I'm being watched or something. That's what it comes to my mind. So that's why I totally disagree with Amazon. While, within university, I'm not sure how it will university can track me. Yeah, maybe what I search on the library database or so for educational purposes. Yes, I agree with the university to offer me

something based on what I do. Not of course, intervene me in my personal life. So yes, again, I trust the university more on that and not obviously the Amazon platform."

[Participant 3, Privacy pragmatist]

In contrast, Participant 1 expressed mistrust that the university would handle student data appropriately, given that students rarely read the data use policies where details of data use would ordinarily be provided:

"Because sometimes I don't think that they might share data anonymously. I don't think they do that. Even though they might be anonymous, at the end, they might know that this data are for me. So, there's no anonymity at the end… I think that's most of the cases when we agree on the terms and conditions, we never read about it. We never do it. So, you don't know what you have signed. And in most cases, the small prints are the ones that they say, we might sell your data anonymously with third parties. So that's why I say that disagree. Because I never do that. I never read the terms and conditions, but I accept them. So, I mean, in the backstage, you don't know what really happens. They might tell you something, but it might be otherwise."

[Participant 1, Privacy pragmatist]

5.3.4.2 Data Access and Control

Participants expressed an interest in having control over third parties' access to their data. Participant 2, for example, wanted control over which third parties could access her data, rather than have the university make this decision. The participant's stance could have been motivated by a lack of information about the third party's identity and how it would use students' data:

"… because I would not want that to go to the third party, because I'm, like I said, it's restricted to one particular Institute, and I would want it that way. If I really want access to another third party, I will go there. I wouldn't want someone else to give my information there… if it's a third party, I don't know the party. I don't know what my information is going to be used [for]."
[Participant 2, Privacy unconcerned]

5.3.4.3 Benefits and Trade-offs

While the potential benefits of the collection, use and sharing of student data for LA were observed to play a role in enabling students' acceptance, students also indicated an awareness of the need to provide their data to access these benefits. Both Participant 3 and Participant 4 referred to the benefits they stood to gain from sharing their data with the university. While Participant 4 anticipated that other interaction would become more convenient as a result,

120

Participant 3 felt more comfortable sharing data with the university as he perceived that it would provide more functional benefits:

> *"I think Amazon could have more data from me, but the only thing that they can do is suggest me things to buy. So, to give more money. While the university can offer me a different kind of service, more quality of my studies, or yeah. So I think if I was to give information to these two platforms, or something, the University could use them more widely to offer me something better. While Amazon wants my money actually. So yeah, so I will be more comfortable to give more information to the university."*

> *[Participant 3, Privacy pragmatist]*

Finally, students' experiences in other situations outside of the learning context may have contributed to the perception that they shared their data in return for some service from the university, as suggested by Participant 2:

> *"So I think that works for me, because how else will I ever work in an institution? I mean, no, no company or nobody in person is ever going to be able to help me without giving some input like that that particular company should know what are the things that I'm looking for? What are the searches that I have? What are the things I'm like I want right now. And if I give that data, only, then they will be able to help me [with] what I want... So, I think the trade-off would if it benefits me, I will be okay with that trade off. Because it's not only going to give me what I really want, but it is also going to provide me with benefits."*
> *[Participant 2, Privacy unconcerned]*

5.3.4.4 Privacy conceptualisations

Each of the students interviewed expressed privacy in different ways, as summarised in Table 5.8 along with the corresponding responses from participants. These responses are provided to make clear how the different privacy concepts were identified. As can be seen in Table 5.8, each student held a different conceptualisation of privacy. A comparison of these conceptualisations of privacy to those identified in Study 1 (reported in Chapter 4) highlights the differences and similarities between how the experts and the students think about privacy in LA. Both the conceptualisation of privacy as contextual integrity and as identity were raised by the experts and the students. The students raised other unique conceptualisations of privacy which were not shared with the experts. Yet another difference between the conceptualisations of privacy is that the experts used more technical language whereas the students were more descriptive in highlighting their expectations.

Table 5.8: Privacy conceptualisations identified in the follow-up interviews

Participant 1	Participant 2	Participant 3	Participant 4
Conceptualised **privacy as** data not being shared with third parties and not being used in ways the student is unaware of. "…privacy is exactly the thing that they keep your data for themselves, to promote themselves to make recommendations for you by themselves, and not selling your personalized preferences to others, so others can take advantage of your data without even knowing it."	Conceptualised privacy as **secrecy**, and the **freedom** to act as they desire with respect to their learning without feeling pressured to act in a certain way. A second conceptualisation of privacy included elements of data being linked to one's **identity**. "…privacy is wherein my information is not shared with anyone else or is not seen by anyone else… So, if at all my, my search history is being tracked, I will be always conscious about making certain searches …. if somebody seeing this, I would, I will have a constant pressure of completing that sooner, and I will not, I will not go with my pace. So that's a problem. … privacy	Conceptualised privacy as involving decisions about **whether one shares data.** "...online privacy is more like the data, yeah, share data or not."	Conceptualised privacy as **personal information** that she provides to the institution, as well as how the institution **uses the data** and its **impact** on her life in the future. "Like when I input my basic information, just like the name, the birth, I think, is a kind of privacy, privacy. And, and things about things about my, just like my career preferences or the address, yeah, those important details and the email or the phone number, I think. And I think all of this information I input into my [university name] account is by private information is about the privacy. So sometimes I think most of

means that I would want to do things my way and not, without having the pressure of somebody seeing me. Or constantly being worried that how am I going to be judged on the basis of whatever I do on online." "…my personal data is something because this, that's my identity. And my identity is going to lead to judgment, which is something I'm really not okay with."		the time, I think, the [university name] is about my university, and it relates to my university life, maybe it relates to my, to my work life in the future. So, when I include this information, [university name], it could be used, it could be shared with other parties, because the parties make the relationship work relationship with the [university name], it must be highly reliable, so it could be shared with them."

The differing privacy conceptualisations emphasise the need for HEIs to make clear which privacy conceptualisation(s) they consider in their LA implementations. Given that students' privacy conceptualisations might not be part of what HEIs consider, there is a possibility that the LA implications do not meet students' privacy expectations or do so only partially.

5.4 Study 2 Discussion

5.4.1 Implications of Findings

Regarding RQ 3, the results of Study 2 suggest that students were more comfortable with the use of their data in the LA context than in the Amazon context. Furthermore, students in the study were less comfortable with the university sharing their data with third parties compared to Amazon. At first glance, this appears somewhat counterintuitive. As discussed below, students have suggested higher trust in their university than in external bodies, yet express greater concern in their university's potential data sharing practices. This may be explained by referring to contextual integrity (Nissenbaum, 2010). For example, it may be that many are already aware of data sharing in a commercial context, and less aware, and potentially therefore more disturbed by, data sharing in an educational context. Additionally, students might have been concerned that they did not know who would have access to their educational practices and data records. They may have considered that the third parties were potential employers, and as such, may have wanted to know what was shared with them, given the potential to influence their future employment prospects. As these concerns could be due to students' lack of knowledge about the details of LA implementations, it is recommended that universities make this information clear to students. Based on the work of Vu, Adkins and Henderson (2019) this level of transparency can be expected to have a positive impact on students' willingness to share data for LA.

Regarding RQ 4, a key finding of this study was that students were significantly more comfortable with the collection and use of data in the university context rather than the Amazon context. Thus, these findings point to students' lack of concern with the use of their data for LA. This result, as explained previously, could be due to students' higher levels of trust in the university.

Regarding RQ 5, the qualitative data suggested that the relationship between the student and the university could lead them to trust the university to use student data for students' benefit. Where mistrust was expressed, this was not necessarily due to the university's inaction, rather it was expressed due to students' lack of awareness (for instance, by not engaging with the content of the privacy policies). This emphasises the need for students to take up available opportunities to be made aware about how their data is used. At the same time, HEIs need to ensure that this information can easily be accessed and understood by students. Finally, students' perceptions of

what they stood to gain from the use of their data for LA also seemed to play a role in minimising their concern.

Overall, the findings from Study 2 shed further light on the dimensions of privacy and students' specific privacy concerns around the collection, use, and sharing of their data for LA. These findings are aligned to contextual integrity (Nissenbaum, 2010) that comfort with data use and data sharing might be influenced by the context in which the data is collected, used, and shared. The results demonstrated students' comfort with the university using their data for LA compared to Amazon using their data. Additionally, students' discomfort with their data being shared with third parties and being shared in an identifiable way can be seen as examples of their norms of appropriateness and flow, that is how they expect their data to be used, and who they expect to have access to their data. These findings emphasise that the context where data is collected and where it is used is an important component in understanding students' data use preferences and what practices might stand out to them as unusual or unacceptable, and thus what they might perceive as violating their privacy.

Concerns about who has access to students' data, for example, third parties in general and future employers as a specific example highlights an informational norm that there is an expectation that data can or cannot be shared, or that it is shared under certain constraints. This highlights a possible need for opportunities for these informational norms to be shared with or captured by the HEI. Thus, it is important to consider how HEIs can identify these informational norms from students, and how these can be used in the design and development of LA, while considering personnel and other resource constraints that HEIs operate under.

The findings from the follow-up interviews demonstrate relevant nuances in the collection and use of student data for LA. The students who were interviewed expressed a lack of trust where organisations wanted to share student data with third parties. Participants also had a desire to control who had access to their data. On the other hand, they indicated that they trusted the university to use the data for students' benefit and seemed to expect to exchange their data in return for a service from the university.

Following on from this work, it is recommended that HEI data use transparency initiatives include information whether student data is shared with third parties, and what this means. For example, it may be the case that only anonymised data is shared and informing students about this can help ease their concerns, or as is usual, that HEIs share student data as part of a service agreement, such as marketing. Furthermore, the results from Study 2 emphasise the need to unpack privacy as a concept into specific dimensions for study, in this way bringing greater clarity to research findings on privacy concern in LA.

125

5.4.2 Limitations

While fifty students signed up to participate in the follow-up interviews, demonstrating their interest in the issues under investigation, only four students were available to attend, providing preliminary insights from the follow-up interviews. Consequently, further research with more students is needed to identify further insights and to determine if these insights are shared by students depending on their stance on privacy. However, the responses from the four students have enabled relevant and noteworthy insights to be identified.

Study 2 faced additional limitations. As students engaged with the vignettes, their responses to the different questions were displayed as graphs on a screen visible to all students in the laboratory. While the PollEV software used to collect data from students displays these responses anonymously (that is, they are not linked on the display to individual participating students), having aggregated responses on the screen could have influenced students' responses to later questions.

An error during the design of the questionnaire resulted in an "Agree" option being displayed on the response scale instead of a "Strongly Agree" option. The loss of this so-called 'multiplying adverb' or 'intensifier' (Lietz, 2010) might have resulted in an overlap in participants' responses (Worcester & Burns, 1975).

While validated questionnaires (Westin & Maurici, 1998; Buchanan, Paine, Joinson, & Reips, 2007) were used to examine students' general privacy concerns and behaviour, the questionnaires used to examine privacy concern in LA, despite being taken from existing research (Slade, Prinsloo, & Khalil, 2019) were observed to load onto multiple factors and different factors across the e-commerce and university contexts. This demonstrates the need for further work to develop validated questionnaires to examine privacy issues in LA, such as the students' expectations of learning analytics questionnaire (Whitelock-Wainwright, Gašević, Tejeiro, Tsai, & Bennett, 2019), or questionnaires used in the work of Ifenthaler and Schumacher (2016).

Study 2 only had input from students from a single HEI. Furthermore, they were all postgraduate students pursuing a single programme of study. As a lesson learned from this limitation, Study 3 (discussed in Chapter 6) was carried out with students from a variety of HEIs and programmes of study in the UK to gather a broader perspective.

5.4.3 Links to Other Studies in This Thesis

The interview data from Study 2 introduced the notion of students having control over the collection, use, and sharing of their data for LA. In particular, one student suggested a preference to determine which third parties could access her data. It was thought that the participant's

perspective was influenced by a lack of information about and control over the use of her data. Additionally, there was an expectation of an exchange and a trade-off – that students would provide their data in return for a service from the HEI. Three of the students who took part in the interviews seemed to appreciate the idea of exchanging their data in return for some potential benefits. Furthermore, their interactions with entities outside the university seemed to contribute to these expectations. These findings motivated the focus of Study 3, which is described in the next chapter.

6 Study 3 Methods and Results[12]

This chapter presents the third study in this thesis. Study 3 investigated whether and how students' awareness of the potential benefits and privacy risks of data use for LA influence their preferences for the use of their data. Section 6.1 introduces the study and the research questions along with the theoretical framing for Study 3. This is followed in Section 6.2 by a detailed discussion of the methods used in Study 3. Section 6.3 contains the results of Study 3 in line with the research questions. Finally, in Section 6.4 the implications of the study findings are identified, along with the limitations of Study 3, and links to the final study in the thesis.

6.1 Introduction

Empirical research on students and privacy in LA has largely focused on whether students (i) are concerned about privacy in the use of their data for LA, (ii) accept institutional use of their data for LA, and (iii) are willing to share their data for LA. With respect to students' willingness to share data, several authors have framed LA as an exchange (Ferguson, 2019; Wintrup, 2017). More specifically, LA is structured and designed such that students have to accept institutional use of their data so as to be provided with LA applications. Students' consent might be sought explicitly or implicitly, for example, through their participation in learning activities. Ferguson (2019, p. 27) argues that the exchange is beneficial to both the student and the HEI.

An element of this transactional process that has not previously been explored, as identified in the research gaps discussed in Section 2.5, is what happens when students can control whether and how their data is used, and they are informed of the potential privacy risks and benefits of data use for LA. There have been calls to equip students with tools to facilitate them managing the privacy of their data (Kelly, McCormack, Reeves, Brooks, & O'Brien, 2021). The authors argue that doing so will more easily enable HEIs to comply with privacy regulations, better track the data that is collected and the purposes, gain clarity on students' awareness of HEI privacy policies, and finally, allow students to manage the data HEIs can collect and use.

Students' data can be used for purposes such as providing students with learning recommendations (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015) and intervening early to recommend remedial action with the aim of improving students' performance (Jayaprakash, Moody, Lauría, Regan, & Baron, 2014; Ho, 2017). LA has had some effect to improve student

---

[12] Sections of this chapter were published in the following book chapter:
Korir, M. M., Slade, S., Holmes, W., and Rienties, B. (2021). Eliciting students' data use preferences in learning analytics: A crowdsourced approach. In: Rienties, B., Hampel, R., Scanlon, E., and Whitelock, D. eds. *Open World Learning: Research, Innovation and the Challenges of High-Quality Education*. Routledge.

grades and retention (Sønderlund, Hughes, & Smith, 2019). However, Wintrup (2017) highlights the potential of LA to negatively impact students' drive for learning. Furthermore, use of student data in these ways also has the potential to harm student privacy. Possible harms to privacy include profiling and categorising students, where there are risks to mis-categorise students and, contrary to expectations of anonymity, identify the individual whose data is used (Yeung, 2018; Reidenberg & Schaub, 2018; Solove, 2009).

Thus, an area identified for further research was that of students' perceptions of the exchange of their data for a service in the context of LA. In particular, little is known about higher education students' perceptions of the use of their data in LA and their preferences for such use of their data. What is also unclear is whether and how students' data use preferences might change if they can control the use of their data and are made aware of the LA applications they can and cannot access based on their data use preferences. Consequently, Study 3 focused on the following research questions:

RQ 6: To what extent does an awareness of the possible privacy risks and benefits of data use for learning analytics contribute to a difference in students' preferences for the use of their data?

RQ 7: To what extent does not having access to learning analytics applications influence students' preferences for the use of their data?

RQ 8: What do students indicate as the motivation for their preferences for use of their data?

The findings of empirical research studies on students and privacy in LA (refer to Section 2.4) suggest that numerous factors influence students' preferences for the use of their data. Furthermore, the findings of Study 2 (discussed in Section 5.3) suggested that students expected to exchange their data in return for some service from the HEI. Thus, to determine the relevant factors for Study 3, the privacy calculus theory (Laufer & Wolfe, 1977; Dinev & Hart, 2006) was used as the theoretical foundation. As discussed in Section 2.3.5, the privacy calculus theory suggests that individuals weigh multiple factors, including the privacy risks and benefits of sharing personal information, to determine whether to share personal data. The next section provides a detailed overview of the methods used in Study 3 to answer the research questions.

6.2   Methods

This section describes how Study 3 was carried out using an online survey experiment with participants recruited through a crowdsourcing platform. As discussed in Section 3.4, a survey experiment was a suitable research method because of the need to randomly assign participants to a control and experimental groups to investigate whether the independent variable of awareness of potential privacy risks and benefits had an influence on the dependent variables

which were students' responses to the various study questionnaires. Furthermore, crowdsourcing allowed recruitment of large groups of participants, despite the restrictions the Covid-19 pandemic placed on face-to-face interactions at the time when Study 3 was carried out.

In this chapter, Section 6.2.1 describes the setting where Study 3 was carried out and the participants who were recruited. This is followed in Chapter 6.2.2 by a discussion of the study design and procedure. The instruments used to collect the study data are introduced in Section 6.2.3 and the materials used for the study are described in Section 6.2.4. This is followed by a discussion of the pilot testing done in Section 6.2.5 and then how the study was conducted on the crowdsourcing platform in Section 6.2.6. Follow-up interviews were planned to follow the mixed methods research design and these are discussed in Section 6.2.7. Finally, Section 6.2.8 describes and justifies the quantitative and qualitative data analysis techniques that were used. Study 3 received ethical approval from the Open University Human Research Ethics Committee (HREC number: 3287).

6.2.1 Setting and Participants

The crowdsourcing platform Prolific Academic[13] was used to recruit participants for Study 3. The decision to recruit participants using Prolific Academic was partly motivated by the Covid-19 pandemic which necessitated that research was carried out remotely. Further details on recruiting participants using crowdsourcing platforms are provided in Chapter 3.

The sample was drawn from UK-based students who were registered on Prolific. The inclusion criteria were that (i) participants were between the ages of 18 and 24 (representing the age of most undergraduate students in universities in the UK[14]), (ii) had completed their A-levels and (iii) indicated that they were students. It was not clear at the time of recruitment whether participants were in further or higher education. Effort was made to recruit an equal number of male and female participants, following the method suggested by Prolific Academic[15] to distribute two separate yet identical questionnaires, one for male participants, and another for female participants.

The sample size was calculated a priori for ANOVA tests (as Study 3 aimed to examine differences between a control group and the experimental groups) with an anticipated effect size of 0.25, a

---

[13] https://www.prolific.co
[14] Universities UK, Higher Education in Facts and Figures, 2019. https://www.universitiesuk.ac.uk/facts-and-stats/data-and-analysis/Documents/higher-education-facts-and-figures-2019.pdf
[15] Prolific guidelines on recruiting an equal number of male and female participants are available on this page: https://researcher-help.prolific.co/hc/en-gb/articles/360009221213-Demographic-balancing-e-g-ensuring-50-males-and-50-females

desired power level of 0.95, and a probability level of 0.05. Based on these assumptions, the minimum sample size was calculated to be 400 participants. The number of participants was identified using G*Power version 3.1.9.4 for the power analysis (see Appendix C.1). As detailed in Section 6.2.2, participants were randomly assigned to four groups, thus each group was expected to have approximately 100 participants. Participants were compensated £1.88 each and participants took between 10 and 20 minutes to complete the study.

As noted in the inclusion criteria, participants recruited for Study 3 were students in further or HEIs in the UK. The Further Educational (FE) institutions included colleges where participants studied academic, vocational, and professional courses, the Higher Education (HE) institutions were largely universities. A total of 447 participants took part in the study. There were 216 male participants (48.3%) and 231 female participants (51.7%). The mean age was 20.6 (SD = 1.86). Most of the participants (409 - 91.5%) were studying in HE and the remainder in FE (38 – 8.5%). The GLOBE country cluster system (House, Hanges, Javidan, Dorfman, & Gupta, 2004) was used to categorise students according to their region of origin as there were several countries with only one or two students (Table 6.1). Most of the participants were from the Anglo (365 – 81.6%), Southern Asia (26 – 5.8%), Eastern Europe (14 – 3.1%), and Latin Europe (12 – 2.6%) clusters.

131

Table 6.1: Descriptive statistics of students' cultural backgrounds and nationalities

Cluster	No. of students	Percentage	Countries and no. of students for each
Anglo	365	81.6	USA (2), Australia (1), British (241), English (100), Ireland (1), Northern Ireland (1), Scotland (11), Wales (8)
Southern Asia	26	5.8	Bangladesh (8), Philippines (2), India (5), Indonesia (1), Malaysia (1), Pakistan (3), Spain (3), Sri Lanka (2), Cambodia (1)
Eastern Europe	14	3.1	Bulgaria (2), Czech Republic (1), Poland (7), Slovak (3), Slovenia (1)
Latin Europe	12	2.6	Cyprus (1), France (1), Italy (5), Portugal (5)
Sub-Saharan Africa	6	1.3	Nigeria (5), Uganda (1)
Confucian Asia	4	0.8	China (3), South Korea (1)
Middle East	4	0.8	Iraq (1), Jordan (1), Morocco (1), Turkey (1)
Nordic Europe	4	0.8	Finland (1), Sweden (3)
Central Asia	3	0.7	Iran (2), Afghan (1)
Germanic Europe	2	0.4	Austria (1), Germany (1)
Latin America	1	0.2	Venezuela (1)
	6	1.3	I'd rather not say (6)

6.2.2 Study Design and Procedure

The aim of Study 3 was to examine the differences in students' data use and data sharing preferences based on whether they were exposed to a "risk", "benefit", or "risk and benefit" intervention. The actual interventions used for Study 3 are shown in Section 6.2.4.

A further aim of Study 3 was to identify the motivation for students' data use and data sharing preferences. Given the need to answer different research questions, a mixed-method study was deemed suitable, collecting both quantitative and qualitative data. Further detailed discussion on the suitability of mixed methods research design for the thesis research is provided in Section 3.3. The quantitative data for Study 3 was collected using a survey tool from JISC[16], while the qualitative data was collected by asking participants to explain, using open responses, the motivation for their data use and data sharing preferences.

Additionally, participants were invited to take part in semi-structured follow up interviews. These follow-up interviews constitute the fourth study of this thesis. Accordingly, the details and the results are discussed in Chapter 7.

6.2.2.1 Procedure

A between-subjects design was used where each participant was randomly assigned to one of four groups: one control group and three experimental groups: (i) privacy risks, (ii) benefits, and (iii) privacy risks and benefits. After providing consent to take part in Study 3, participants were asked to indicate whether they would want personal data used for LA, and if so what category of data they would want to be used, where the categories were data about themselves and/or data about their activities on the online learning platform. The study proceeded as follows for all participants:

1. They were briefly informed about what LA is.
2. They viewed a sample LA dashboard.
3. They were asked to assess the usefulness of the LA dashboard features and to indicate whether they were concerned about the use of their data for LA.
4. Participants were then shown specific data items that can be used for LA and asked to indicate if there were any data items that they would be willing to share for LA.
5. They then indicated their data use preferences again (as at the start of Study 3) and were then informed of potential loss of benefits based on their data use preferences and asked whether they wanted to include missing features.

---

[16] www.onlinesurveys.ac.uk

6. Participants then answered questions to indicate their level of privacy concern (in general), and lastly, provided demographic information.

As indicated in Section 6.1, the theoretical framework for Study 3 was the privacy calculus theory (Laufer & Wolfe, 1977; Dinev & Hart, 2006) which suggests that users in effect carry out a privacy risk-benefit analysis to determine whether to disclose their personal data. Therefore, based on the privacy calculus theory, participants in the experimental groups additionally received privacy risk and benefit interventions in addition to that mentioned in the procedure described above.

- Participants in the *privacy risks* group were given a privacy risks intervention (detailed in Section 6.2.4), which was a description of the potential privacy risks arising from the use of student data for LA. They were asked to indicate their level of concern for the stated privacy risks.

- Participants in the *benefits group* were given a benefits intervention (also detailed in Section 6.2.4), which was a description of the potential benefits arising from the use of student data for LA and were asked to indicate how beneficial they thought these were.

- Finally, participants in the *privacy risks and benefits* group were given both the privacy risks and benefits interventions and were asked to indicate their level of concern for the stated privacy risks and their perception of the benefits.

After expressing their data use preferences, participants in the four groups saw the impact of their choices on a sample LA dashboard (described as one of the study materials in Section 6.2.4). Specifically, components designed to display the output of LA were added to or removed from the LA dashboard depending on participants' choices. This design decision sought to mimic a trade-off between data use (and privacy risks) and benefits (or features) of LA. Participants then saw the available and missing features and were given a choice to include them again. The design of Study 3 is summarised in Figure 6.1, which also distinguishes between the study instruments and the study materials. These are discussed in the next section.
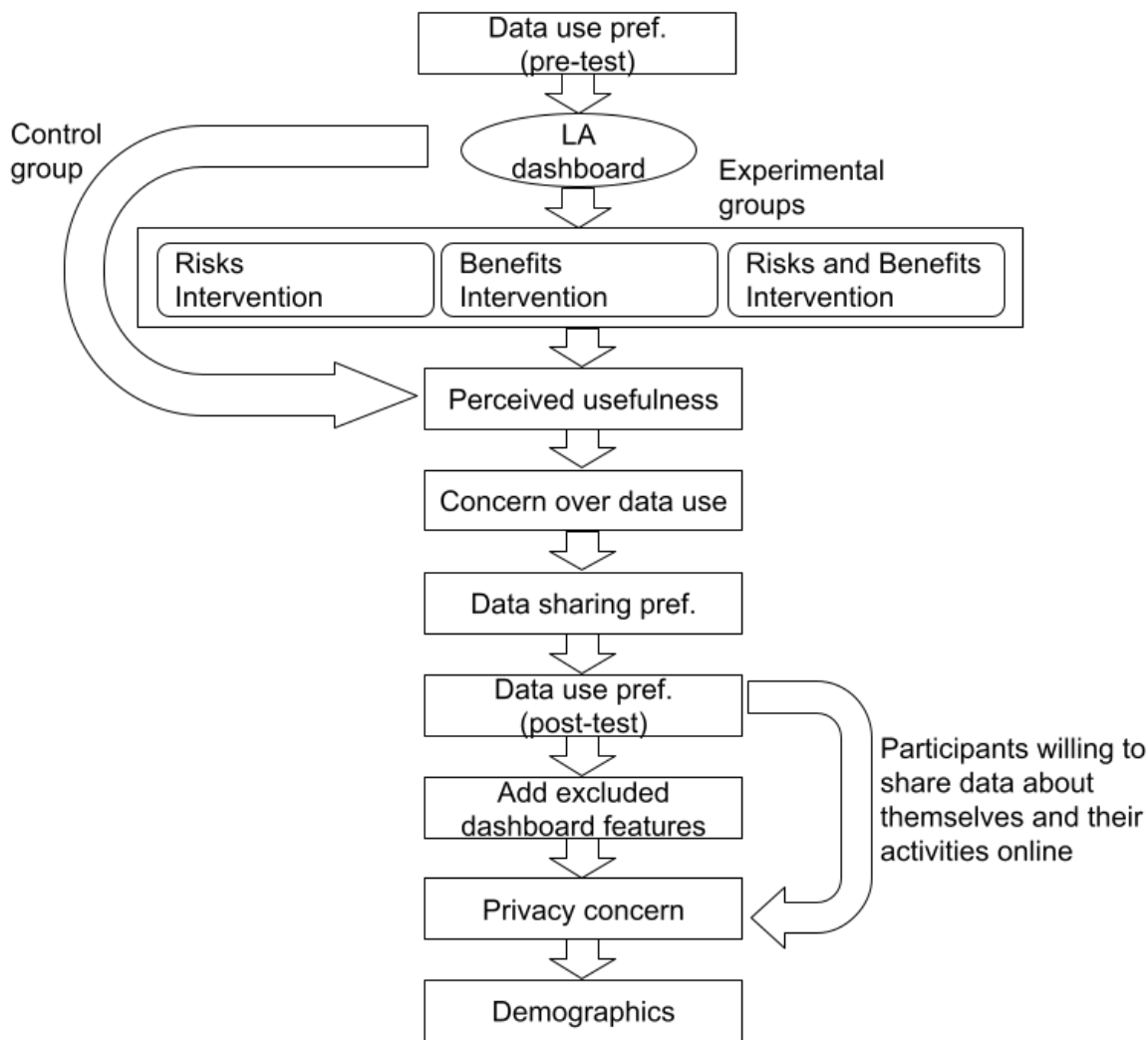
Figure 6.1: Study 3 procedure

6.2.3   Study Instruments

This section gives further information on the measures used in Study 3. As indicated in Table 6.2, three measures were created for Study 3 while a further five measures were drawn from published research as recommended in best practice for questionnaire design (Groves, et al., 2009). The instruments are shown in Table 6.2 in the order in which participants encountered them in Study 3. Sample questions for each of these measures are also shown while all questions are provided in Appendix C.2.

Table 6.2: Examples of questions used in Study 3

Scale	Source	N items	Example item	Response scale	M	SD	Alpha
Usefulness of learning analytics features	Ponciano, Barbosa, Brasileiro, Brito and Andrade (2017)	3	How useful are the following features for your studies: the feature to see your predicted performance	[1] Not at all useful – [5] Very useful	4.03	.77	.42
Perceived usefulness	Arbaugh (2000)	4	Using learning analytics and the learning analytics dashboard would enhance my effectiveness in my studies	[1] Strongly disagree – [5] Strongly agree	4.02	.80	.88
Concern about data use	Developed for this study	3	Please indicate to what extent do you agree or disagree with the following statements? I would be concerned about my learning institution using student data for learning analytics	[1] Strongly disagree – [5] Strongly agree	2.59	.94	.71
Sharing of data	Ifenthaler and Schumacher (2016)	28	The following list represents the types of data that might be used for LA and to create the LA dashboard. Please indicate whether you would agree to the use of the following data. (i) name	[1] Do not agree – [2] Agree	1.37	.20	.87

Data use preferences	Developed for this study	1	Please indicate how you would prefer your learning institution use your data to improve teaching and learning, based on the categories and examples shown in the image above	I prefer that no data is used, I prefer that only data about myself is used, I prefer that only data about my activities on the learning platform is used, I prefer that data about myself and my activities on the learning platform is used, I need more information to make a decision	3.03 (pre); 3.00 (post)	.89 (pre); .93 (post)	-
Concern about privacy risks	Developed for this study	4	Please indicate whether you are concerned with the possible risks associated with sharing data for learning analytics:	[1] Not at all concerned – [5] Extremely concerned	2.94	1.13	.84

			(i) We will monitor what you and other students are doing on the online learning platform				
Benefit perception	Naeini, Bhagavatula, Habib, Degeling, Bauer, Cranor and Sadeh (2017)	2	The following use of my data would be beneficial for me: (i) Personalised support to help me complete the course	[1] Strongly disagree – [5] Strongly agree	4.27	.88	.86
Privacy concern	Malhotra, Kim, and Agarwal (2004)	10	Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared	[1] Strongly disagree – [5] Strongly agree	4.35	.58	.76

Three measures, namely data use preferences, concern about data use, and concern about privacy risks were created specifically for Study 3. It was necessary to create these three measures as there was little research on data use preferences in the LA context, therefore there were few opportunities to identify questions from related research as recommended in best practice for questionnaire design. Given that these questions were new, extensive pilot testing was carried out (as detailed in Section 6.2.5) to enhance the clarity of the questions for the study participants. In a pre-post test design, participants' data use preferences were obtained at the start of the study to establish a baseline measure as well as during Study 3 to determine if there were any changes. This allowed a comparison and an assessment of the influence of the study interventions.

The measure for usefulness of LA features was adapted from Ponciano, Barbosa, Brasileiro, Brito and Andrade (2017) who developed this instrument in a study conducted in Brazil, on privacy concerns in Internet of Things systems. The scale was amended to refer to LA features where the original scale referred to the features of Internet of Things systems. The authors did not report the Cronbach alpha values obtained for this instrument. In Study 3, however, the scale was found to have low reliability with a Cronbach alpha value of 0.45. Consequently, the scale on the usefulness of LA features was excluded from further use in the data analysis.

The measure for perceived usefulness of LA features was adapted from Arbaugh (2000) who developed it in a study on student satisfaction with MBA courses with 114 students. The scale had an acceptable Cronbach alpha value of 0.82 and has been used in related work such as with 295 e-learners (Sun, Tsai, Finger, Chen, & Yeh, 2008) with even higher Cronbach alpha values reported of 0.91. The scale was amended to include reference to LA and the LA dashboard used in Study 3.

The sharing of data scale was used in work by Ifenthaler and Schumacher (2015; 2016; 2019) with over 300 students. The scale has acceptable reliability with a Cronbach alpha value of 0.74. The original scale was administered to students in Germany and therefore was in German. The sharing of data scale lists different data items that can be used for LA, for example, name, address, and posts on university discussion forums. The scale was selected for use in Study 3 because of its high reliability scores and extensive use in prior related research (Ifenthaler & Schumacher, 2019; Ifenthaler & Schumacher, 2016). Amendments to the scale involved translating it into English and adapting it to the higher education context in the UK. To make the adaptations, five individuals who were born in Germany or in German-speaking countries were asked to provide translations from German to English (three individuals) and from the English translation back to German (two individuals). Brief email discussions or video call meetings were held to discuss the translations

provided before a final set of questions was prepared for Study 3. The translations provided and the final list of questions are shown in Appendix C.3.

The measure for perception of benefit from data use for LA was adapted from Naeini et al., (2017) who conducted a crowdsourced study with 1,014 participants focusing on privacy preferences in the Internet of Things. The authors did not report the Cronbach alpha values obtained for this instrument. The scale had acceptable Cronbach alpha values of 0.86. The questions were adapted for Study 3 by indicating the specific use of data referred to in the LA context.

Finally, the Internet Users Information Privacy Concern (IUIPC) scale (Malhotra, Kim, & Agarwal, 2004) was developed to measure users' privacy concerns, focusing on three dimensions of privacy, namely, control, collection of data, and awareness of privacy practices. The scale was originally developed in two studies with over 700 participants (Malhotra, Kim, & Agarwal, 2004) and gives insights into users' privacy concerns.

Best practice recommendations for questionnaire design and crowdsourced studies were implemented in the design of Study 3. All the scales used in Study 3 were modified to include a "not applicable" option so that participants could provide a response to a question if it did not apply to them (Krosnick, 2018; Aldridge & Levine, 2001). Additionally, attention check questions were used to ensure that participants did not answer questions at random during the study (Egelman, Chi, & Dow, 2014). An example of an attention check question that was used in Study 3 is "Please answer 'no' for us to test the reliability of the responses".

6.2.4   Study Materials

This section describes the materials required for the online survey experiment. It was not expected that participants were familiar with or had been exposed to LA. Thus, all participants were first given a brief description of LA as shown in Figure 6.2.

Background

Imagine that you are about to begin an online course at your learning institution

Your learning institution is testing a new initiative – **learning analytics** and students' use of a **learning analytics dashboard**. We would like to **obtain your permission to use your data**.

Learning analytics refers to the use of data about students and their use of online learning resources in several ways. We plan to use student data in two ways: to predict students' performance and to provide study recommendations that can improve students' learning.

Predicting students' performance

We will **predict how likely you are to submit an assignment and how likely you are to pass or fail the course**. These predictions can be shared with your tutors or lecturers, academics or course team, and/or other members of the student support team.

Providing recommendations to improve students' learning

We will **suggest additional study materials or learning activities** that are likely to improve your understanding of the course material and, as a result, your performance. The predictions and recommendations will be shown to you using a learning analytics dashboard.

Submit and continue >

Figure 6.2: Background information on learning analytics for Study 3 participants

Data Use Preference Prototype and Learning Analytics Dashboard

In determining which LA dashboard to use as an example for Study 3, a collection of LA dashboards from a 2018 workshop on student-facing LA [17] was examined. This was done to identify different ways that LA dashboards present information to students and the information that can be presented. Two principles guided the development of the LA dashboard: simplicity and modularity. It was deemed important to use a simple design to maintain focus on Study 3's research questions rather than the interface design components. Therefore, it was considered appropriate to design a LA dashboard with limited but relevant content. Consequently, preference was given to textual content rather than graphs and charts. Furthermore, a modular design was applied to facilitate studying the impact of the independent variables on the dependent variable. Specifically, the design facilitated the addition and removal of features from

---

[17] https://sflaprinciples.wordpress.com/

the dashboard to investigate how the possible loss or inaccessibility of LA features would influence participants' data use preferences.

The data use preference prototype indicated to participants some of the data that can be used for LA. Sclater and Peasgood (2016) highlight the different types of data that can be used for different types of LA. As the LA dashboard visualised predictive and prescriptive LA, two types of data were selected: data about the student and data about the student's activities on the online learning platform. These two types of data are also used in the Open University's predictive analytics tool, OU Analyse (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015), the interface selected as a template for Study 3, and which informed the design of the prototype.

The data use preferences that participants expressed changed the content on the LA dashboard to allow an investigation about how students responded to these trade-offs. The LA dashboard shown to all participants during Study 3 and those who chose to share both data about themselves and data about their activities on the learning platform is shown in Figure 6.3. The LA dashboard shown to participants who chose to share no data is shown in Figure 6.4. The LA dashboard shown to participants who chose to share data about themselves is shown in Figure 6.5. Finally, the LA dashboard shown to participants who chose to share data about their activities on the learning platform is shown in Figure 6.6.
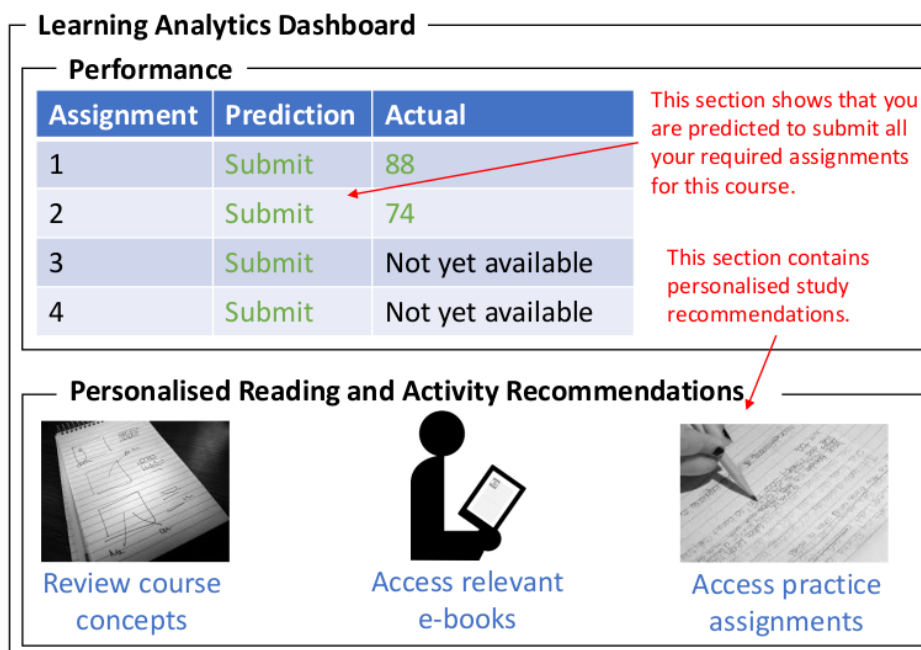
Figure 6.3: The prototype learning analytics dashboard designed for Study 3

Learning Analytics Dashboard

Performance

Assignment	Actual
1	88
2	74
3	Not yet available
4	Not yet available

This section shows the scores on your submitted assignments for this course.
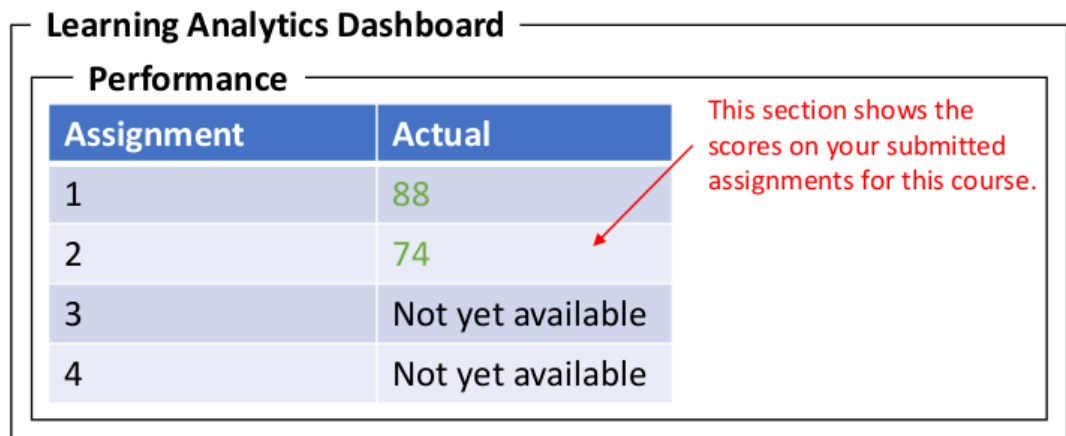
Figure 6.4: The learning analytics dashboard shown to Study 3 participants who chose to share no data

Learning Analytics Dashboard

Performance

Assignment	Prediction	Actual
1	Submit	88
2	Submit	74
3	Submit	Not yet available
4	Submit	Not yet available

This section shows that you are predicted to submit all your required assignments for this course.

This section contains general study recommendations.

General Reading and Activity Recommendations

Non-personalised recommendations

Non-personalised recommendations

Non-personalised recommendations

Figure 6.5: The learning analytics dashboard shown to Study 3 participants who chose to share data about themselves
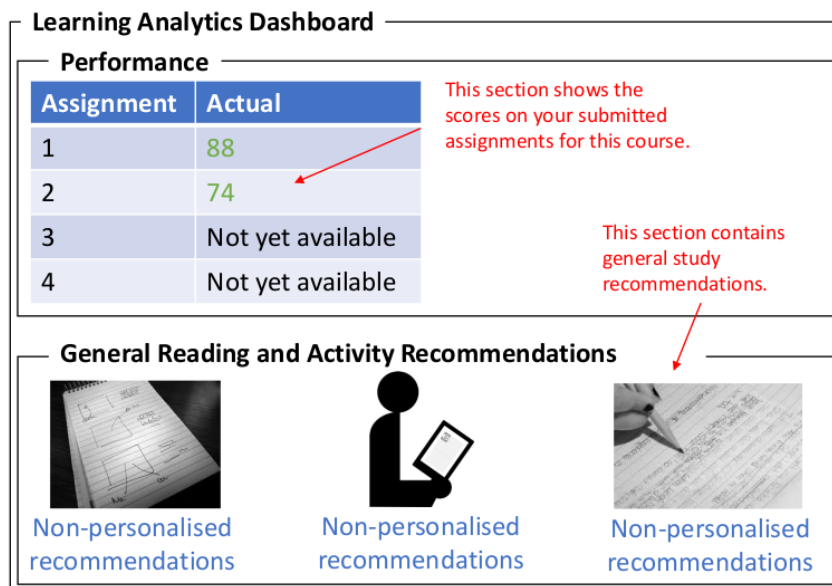
Figure 6.6: The learning analytics dashboard shown to Study 3 participants who chose to share data about their activities
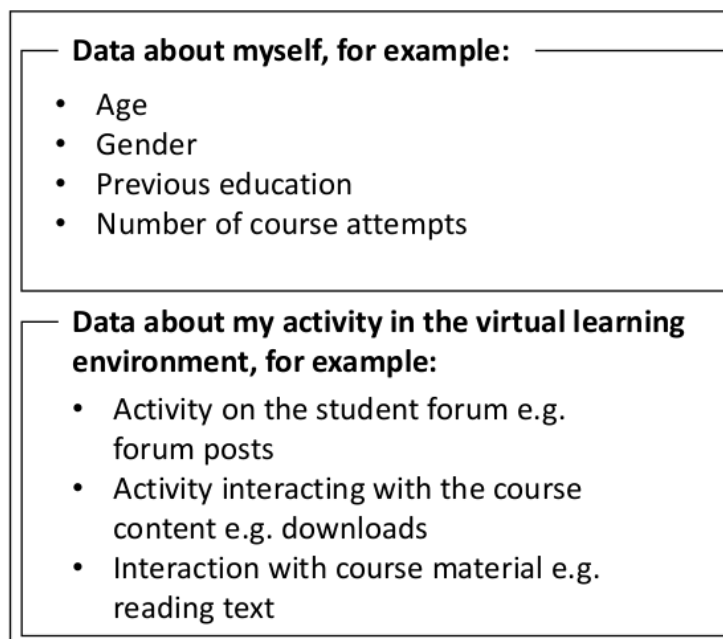
Figure 6.7: Examples of the data used for learning analytics for Study 3 participants

Figure 6.7 contains examples of data that can be used for LA, namely, age, gender, previous education, and number of course attempts. Other data from this broad category that can be used includes ethnicity, disability, and presumed income (based, for example, on postcodes). A potential critique of this selection is that it could be argued that participants were shown data that can be considered 'safe', for example, age and gender, rather than including data such as ethnicity, disability, and presumed income (based on postcodes) whose use in LA has generated

144

debate (Prinsloo & Slade, 2018). This choice of 'safe' data types might reflect varying data use practices across different countries as the scale was developed and used in Germany (Ifenthaler & Schumacher, 2016) and was adapted for Study 3 in the UK. While a broad range of data items are used in LA systems to identify students deemed as 'at-risk' (Feathers, 2021), they were not highlighted in Study 3 as they were deemed as likely to elicit a strong negative response from participants while the benefits were not considered as likely to evoke a strong positive emotion to provide a suitable counterbalance.

Privacy Risks Intervention

The findings of related research in the context of the privacy calculus theory (Laufer & Wolfe, 1977; Dinev & Hart, 2006) suggest that there is a relationship between awareness of privacy risk and willingness to share personal information. Specifically, where there is a high perception of privacy risk, users are less willing to transact with their personal information (Dinev & Hart, 2006). Therefore, Study 3 sought to make some of the participating students aware of possible privacy risks of data use for LA. A *privacy risk intervention* was designed using Solove's taxonomy of privacy harms (2009), in particular the categories of information collection and information processing, to raise students' awareness of possible privacy risks of data use. With information collection, students could be observed overtly or covertly, or they could be pressured to provide information. With information processing, data is aggregated to form a partial picture of a user, this picture can be tied to the person in the real world, data can be poorly stored such that it is insecure, personal data could be re-used without students' consent, or students can be excluded from information about their records. Further details about the taxonomy of privacy harms are provided in Section 2.3.1.4.

The privacy risk intervention used a scenario to describe some of the possible privacy risks arising from the use of student data for LA. In particular, the privacy risks intervention was developed using two categories of the privacy harms framework and corresponding privacy risks: the *information collection* category which corresponds to the risk of *surveillance*, and the *information processing* category which corresponds to the risks of aggregation and identification. These two categories and the corresponding risks were selected in alignment with work by Reidenberg and Schaub (2018, p. 6) who argue that "some of the most salient privacy harms arise in the processing and dissemination of student information, spurred by extensive information collection."

Accordingly, participants were shown the following description of privacy risks:

We will monitor what you and other students are doing on the online learning platform [1]. Data that you and other students have provided to separate information systems at your learning institution (for example during registration) will be combined to form a digital profile [2]. The digital profile can be linked to the individual student [3], and this information will be used to make decisions about you and other students, such as, predicting your performance and giving you study recommendations [1].

The first risk [1] is referred to at the beginning and end of the description. It relates to information collection and the risk of surveillance[18]. The second risk [2] falls under the information processing category, and the risk of aggregation. The third risk [3] is also in the information processing category, under the risk of identification.

Benefits Intervention

The findings of related research in the context of the privacy calculus theory (Laufer & Wolfe, 1977; Dinev & Hart, 2006) also suggest that there is a relationship between awareness of benefits and willingness to share personal information. Specifically, where users stand to receive benefits, they are observed to share more data (Li, Rathindra, & Xu, 2010). Therefore, Study 3 sought to make students in conditions 2 and 3 aware of possible benefits of data use for LA.

A *benefits intervention* was designed which informed some of the participants in conditions 2 and 3 of the potential benefits of the use of student data for LA. The intervention focused on predictive LA and learning recommendations based on related research (Herodotou, Rienties, Boroowa, Zdrahal, & Hlosta, 2019; Sclater, Peasgood, & Mullan, 2016). Participants in Study 3 were shown the following description:

We can offer you personalised support to help you complete the course, including nudging to submit assignments or follow up from the student support team. We can also provide you with personalised recommendations of learning materials that can improve your understanding of the course material.

The next section discusses how the study instruments and materials were tested prior to Study 3 being carried out.

6.2.5   Pilot Testing

Pilot tests of Study 3 were carried out between May and July 2020 in two stages. Participants recruited for the pilot tests were doctoral students registered at the Open University. They had

---

[18] The description of the risk of surveillance [1] was split into two parts to enhance the presentation and clarity for participants.

mixed educational subject backgrounds (as shown in Appendix C.4). A convenience sample such as this that is selected because it is accessible to the researcher is acceptable for pilot testing purposes (Bryman, 2016). The pilot tests aimed to identify problems with the design of Study 3, including the questions for participants, so that improvements could be made before the main study was carried out. The pilot tests were carried out remotely as face-to-face meetings could not be carried out, due to the Covid-19 pandemic.

The first stage of the pilot tests involved a remote lab study. Five doctoral students were presented with the study material and invited to talk aloud as they went through the study, following recommendations in Lazar, Feng, and Hochheiser (2010). These doctoral students (4 female and 1 male) had backgrounds in education and educational technology but did not have considerable experience with LA. They were asked to point out any areas where they perceived that the questions were unclear. The students highlighted minor issues related to the wording of the questions, which were later resolved.

The second stage of the pilot tests was then carried out with sixteen PhD students. There were 14 female and 2 male participants. The students also had mixed educational backgrounds, including education, educational technology, and literature, and one participant had a background in LA. These pilot test participants were randomly distributed across the four experimental conditions (control, risks, benefits, and risks and benefits). They filled out the study questionnaire and provided feedback on issues they experienced during the pilot test. The second pilot test served to evaluate the design of Study 3, clarify how long it would take (participants took an average of 18 minutes to complete the study), and facilitate the development of the quantitative data analysis plan. The issues raised by participants in the second round of the pilot test were addressed before Study 3 was conducted.

6.2.6 Administering the Online Survey Experiment on Prolific

The questionnaire was distributed on the Prolific platform[19] between September and October 2020. Separate questionnaires had also been created for each of the four conditions. It was important to ensure that participants did not fill in the questionnaire multiple times. Therefore, on the advice of the Prolific researcher support team, the questionnaires were distributed one at a time rather than all at once. Once a questionnaire was filled by the desired number of participants, the data was reviewed. Data from participants who failed the attention check questions were excluded. Following guidelines from the Prolific platform, participants who failed the attention check questions were sent a brief message explaining the reasons for their data

---

[19] www.prolific.co

being excluded. Participants whose data was accepted were then compensated, after which their Prolific identification numbers were added to an exclusion list before the next questionnaire was distributed on the platform. This allowed exclusion of those participants who had already taken part.

6.2.7   Follow-up Interviews

At the end of the questionnaire, participants were asked to indicate their interest in taking part in remote semi-structured follow-up interviews. They were offered £5 as compensation for their time. One hundred and seventy-five (175) participants across the four groups expressed an interest in taking part. Details about the follow-up interviews are provided in Chapter 7 to allow for a detailed analysis and discussion of the findings.

6.2.8   Data Analysis

The independent variables in Study 3 were students' awareness of the potential privacy risks and/or benefits of LA. The dependent variables were participants' responses to the various questions indicated in Table 6.2. This section describes and justifies the quantitative and qualitative data analysis techniques that were used to analyse the data and answer the study research questions.

6.2.8.1   Quantitative Data Analysis

Dealing with Missing Data, 'Not Applicable', and 'Need More Information' Responses: There was missing data as six participants did not fully respond to all questions. One participant in the control group did not indicate their age. Furthermore, a total of three participants in the privacy risks and benefits, control, and privacy risks groups did not indicate their pre-test data sharing preference, one participant in the privacy risks and benefits group did not answer one question on the risk perception scale, and one participant in the benefits group did not answer one question on the privacy concern scale.

Missing data was replaced with the mean value calculated from participants' responses to a given question following best practice guidelines (Groves, et al., 2009). Additionally, all the scales included a 'not applicable' or 'I need more information' option. This data did not clearly indicate participants' response to the questions and was therefore also treated as missing data during the analysis (and consequently replaced with the mean value of participants' response to the question).

Differences Between Groups Based on Demographic Characteristics: ANOVA and chi-square tests were used to check whether the groups were the same based on participants' demographic

characteristics (age, gender, nationality, education) and additionally, based on participants' responses to the data use preferences questions (pre- and post-test).

The null hypothesis for the ANOVA tests was that the groups were the same, and the alternate hypothesis was that the groups were different. A one-way ANOVA showed that there was no significant difference based on age ($F(3, 443) = .907$, $p = .438$). Therefore, the null hypothesis was retained, to conclude that the groups were the same based on age.

The null hypothesis for the chi-square tests was that there was no association between pre-test score, post-test score, gender, nationality, and education and the groups, and the alternative hypothesis was that there was an association between these variables and the groups. The chi-square tests showed that there was no association between post-test score ($X_2(9, N = 447) = 10.279$, $p = .328$), gender ($X_2(3, N = 447) = .109$, $p = .991$), nationality ($X_2(123, N = 447) = 108.805$, $p = .816$), education ($X_2(3, N = 447) = 1.286$, $p = .732$), and group. Therefore, the null hypothesis was retained to conclude that there were no differences between the groups based on nationality and education. However, an association was identified between pre-test score and group ($X_2(9, N = 447) = 26.151$, $p = .002$, $V = .140$). Therefore, the null hypothesis was rejected, to conclude that the groups differed based on the pre-test scores.

6.2.8.2   Qualitative Data Analysis

Participants' open responses were analysed to answer the following research question:

RQ 8: What do students indicate as the motivation for their preferences for use of their data?

As discussed in Section 6.2.2, the data use preference question differed from the data sharing questions. The data use preference question asked participants to indicate the *categories* of data use preference namely, whether they preferred to share no data, data about themselves, data about their activities on the online learning platform, or both data about themselves and their activities on the online learning platform. The data sharing preference question asked participants to indicate specific data items they were (not) willing to share for LA. Additionally, participants were asked to indicate the motivation for (not) choosing to add excluded features to the learning dashboard. Thus, there were three sub-questions investigated:

RQ 8a: What motivated participants' (post-test) data use preferences?

RQ 8b: What motivated participants' data sharing preferences?

RQ 8c: Why did participants (not) choose to add excluded features to the learning dashboard?

Participants gave open responses to explain how they determined (i) whether to share their data with LA applications (post-test data use preferences), (ii) which data items to (not) share (sharing

of data), and (iii) whether to include missing LA features. Thematic analysis was used to analyse participants' responses (Saldaña, 2016).

During the qualitative data analysis process, participants' open responses indicating the motivation for their data use preferences and their data sharing preferences were analysed in a stepwise manner as follows: Responses from participants in each group were analysed iteratively, with resulting codes noted and refined as new codes were identified. As the data analysis process continued across responses from the different groups, previously identified codes were grouped to form categories containing related codes. The codes and categories were reviewed iteratively to identify themes. While themes can play three roles in qualitative data analysis, namely, summarise what is going on from participants' responses, provide explanations, or suggest why one has observed what the data contains (Saldaña, 2016), for the purposes of Study 3, the identified themes helped to explain the quantitative results.

Where participants' responses did not address the question they were asked, or suggested that they had not understood the question, or participants did not provide enough information to allow for an interpretation of their response, these were coded as 'other' during the data analysis process. For example, one participant, when asked to explain why they had indicated a preference for no data to be shared stated 'I do not want my details regarding this study to be shared' which was interpreted as a misunderstanding of the study question (given their reference to 'this study'). Subsequently the response was coded as 'other'.

A subset of the data was analysed collaboratively with the thesis supervisors. The numerous benefits of collaborative qualitative data analysis are elaborated on in Section 3.5.1.4. First, the data was analysed inductively after which the researcher discussed the codes and themes with members of the supervision team. The researcher shared a random selection of responses from 60 participants across the 4 groups with two members of the supervision team, along with a code book containing the identified themes and codes. The data selected contained approximately 10% of the codes identified by the researcher. The supervisors were invited to inductively code the data. Areas of agreement and disagreement were reviewed, and further improvements were made to the codebook and the coding. A second round of collaborative data analysis was then carried out where a smaller set of codes was shared with members of the supervision team. The researcher then calculated the level of agreement across the codes. For the data use preferences, Cohen's kappa was 0.309 in the first round, and 0.741 in the second round. For the data sharing preferences, Cohen's kappa was 0.91 in the first round and 1 in the second round. Both values for Cohen's kappa from the second round were considered to demonstrate satisfactory agreement (Lazar, Feng, & Hochheiser, 2010). Therefore, a code book was generated based on these

discussions and the responses reviewed, aligning existing codes to those discussed with the supervision team. The code book for participants' data use preferences is provided in Appendix C.6 while the code book for participants' data sharing preferences is provided in Appendix C.7. Finally, the themes and codes corresponding to participants' reasons for (not) adding excluded features to the LA dashboard are provided in Appendix C.8.

Note that Study 3 was conducted soon after it was revealed that UK high school students' results had been determined by an algorithm which substantially reduced students' grades from those predicted by their teachers (BBC, 2020). Some participants in Study 3 did comment on the issue in their responses. However, it is noted from the quantitative data that participants did not outrightly accept or reject the use of their data for LA. The latter would be expected if the reported challenges resulting from the use of the algorithm had negatively influenced their perspective. Rather, as discussed in the quantitative results in Section 6.3.1, participants' responses suggested that they sought to balance the use of their data with the benefits they expected to receive.

6.3   Results

The findings of Study 3 are presented in this section, first with a focus on the quantitative results, followed by the qualitative results.

6.3.1   Quantitative Results

This section presents the quantitative results of Study 3, corresponding to the research questions.

6.3.1.1   Awareness of Privacy Risks and Benefits and Participants' Data Use Preferences

Participants' data use preferences were obtained at the start of Study 3 (pre-test) and after participants were shown the privacy risks and/or benefits interventions (post-test). The mean and standard deviation values for participants' data use preferences are shown in Table 6.3.

Table 6.3: Descriptive statistics of students' pre-test and post-test data use preferences by experimental group

Descriptive statistics	Condition	Mean	Std. Deviation
	Control	3.09	0.97
	Privacy risks	2.93	0.90
Data use preference pre-test	Benefits	3.04	0.85
	Privacy risks and benefits	3.05	0.82
	Total	3.03	0.89
	Control	2.97	0.9
	Privacy risks	2.89	0.93
Data use preference post-test	Benefits	3.07	0.90
	Privacy risks and benefits	3.05	0.92
	Total	3	0.93

As seen in Table 6.3, the pre-test mean values suggest that participants started Study 3 with a high willingness to share data. This might be due to the data collection method, using a crowdsourcing platform, where participants could be used to providing data about themselves for various studies. This issue is related to the study methodology and is discussed further in Section 3.4.

There were also differences between the groups on the pre-test mean values, with the control group having the highest mean, the risk group having the lowest mean, while minor differences were observed between the mean values for the benefits group and the privacy risks and benefits group.

Compared to the pre-test mean values, there was a decrease observed in the post-test mean values for the control group and the privacy risks group, and an increase in the post-test mean values for the benefits group, while the post-test mean values for the privacy risks and benefits group remained unchanged. In other words, the results suggest that the awareness intervention had an influence on participants' data use preferences in the control, privacy risks, and benefits group, but had no influence on the data use preferences of participants in the privacy risks and benefits group. However, a one-way ANOVA revealed that the differences among the means of the four groups (control, privacy risks, benefits, privacy risks and benefits) were not significant on pre-test data use preferences ($F(3, 443) = 0.64$, $p = .59$), and post-test data use preferences ($F(3, 443) = 0.79$, $p = .50$).

A slight decrease was observed between the overall post-test and pre-test mean scores (post-test mean = 3.00; pre-test mean = 3.03). A paired samples t-test revealed that these differences were

not statistically significant (p = 0.341). Therefore, these results suggest that other factors had a stronger influence on participants' data use preferences than the privacy risks and/or benefits interventions.

6.3.1.2 Awareness of Privacy Risks and Benefits and Participants' Data Sharing Preferences

Along with the analysis of whether the interventions influenced participants' data use preferences, a related question of interest was whether the privacy risks and benefits interventions had an influence on participants' data sharing preferences. The descriptive statistics for participants' data sharing preferences in terms of the mean and standard deviation are shown in Table 6.4.

Table 6.4: Descriptive statistics of students' data sharing preferences by experimental group

	Descriptive statistics		
	Condition	Mean	Std. Deviation
	Control	1.23	0.42
	Privacy risks	1.24	0.42
Data sharing preference	Benefits	1.3	0.46
	Privacy risks and benefits	1.27	0.44

As can be seen in Table 6.4, there were slight differences between the groups on the mean values for participants' data sharing preferences, with the benefits group having the highest mean followed by the privacy risks and benefits group. While it was expected that participants in the privacy risks group would be less willing to share data, the mean value of participants' data sharing preferences was only slightly different from that of participants in the control group. Therefore, participants in the treatment groups were observed to have a higher willingness to share their data for LA, compared to the control group. Thus, this finding suggests that other factors had an influence on participants' data sharing preferences, or the potential risks described in the study did not concern them.

Participants' responses to the sharing data scale demonstrated their willingness to share different data items for LA. The results are shown in Figure 6.8.
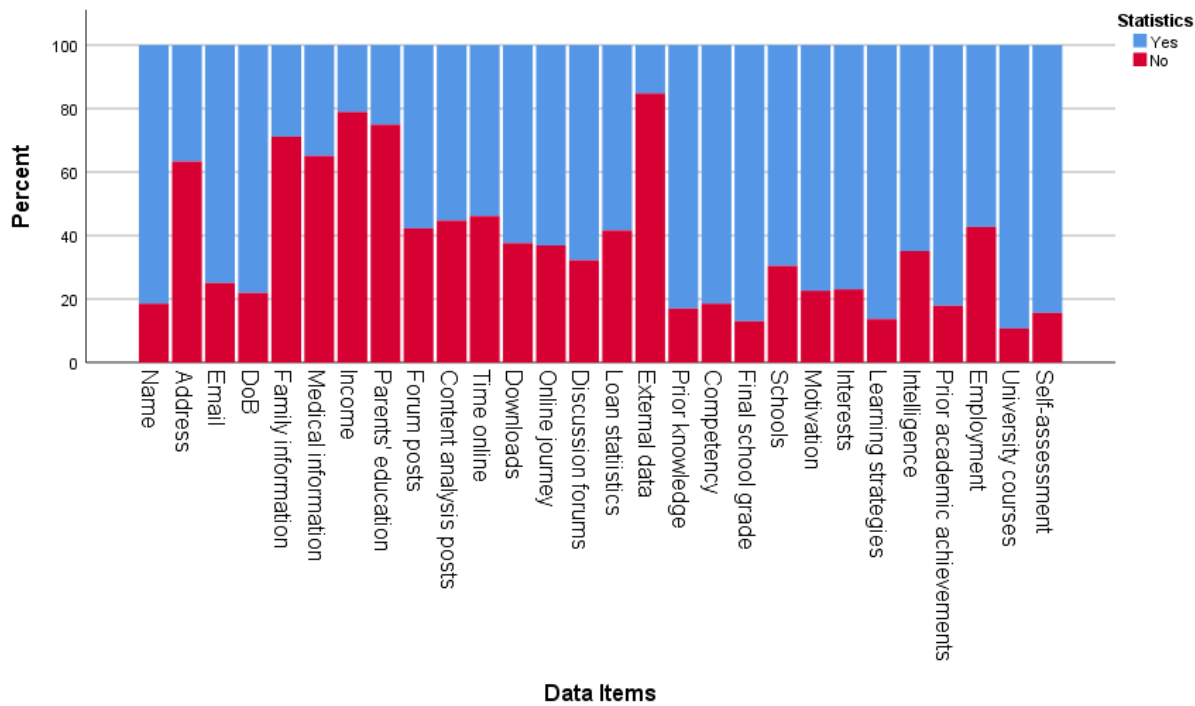
Figure 6.8: Participants' willingness to share various data items in response to the sharing data scale

Participants were most willing to share data related to their studies, namely, information about their chosen university modules/courses (89.3%), final school grade (87%), learning strategies (86.4%), results of self-assessment tests (84.3%), prior academic achievements (82.1%), and their name (81.4%). In contrast, they were most reluctant to share data that to them did not seem directly related to their studies. For example, few participants were willing to share search terms and personal social media profiles (15.2%), income (21%), parents' education (25.1%), information about their family (28.9%), medical information (34.9%), and their address (36.7%).

Factor Analysis: Twenty-eight items relating to participants' willingness to share data were factor analysed using principal components analysis with direct oblimin rotation. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.84, above the commonly recommended value of 0.6 (Williams, Onsman, & Brown, 2010), and Bartlett's test of sphericity was significant (X_2(378) = 4304.758, p < .001). The scree plot (shown in Appendix C.5) indicated the presence of 2 or 3 factors. Three factors were selected (as shown in the pattern matrix in Appendix C.5):

Factor 1 was labelled *'learning and academic history'* because of the high loadings by the following items: learning strategies, motivation, final school grade, interests, prior academic achievements, self-assessment tests on set reading, information about chosen university modules/courses, competency test results, intelligence, tests of individual prior knowledge, time spent in school and types of schools, and information about employment during studies. This first factor explained 23.3% of the variance.

154

Factor 2 was labelled *'personal data'* because of the high loadings by the following items: information about family, medical information, address, income, date of birth, parents' educational level, email, and name. This second factor explained 9.5% of the variance.

Factor 3 was labelled *'digital exhaust data'* because of the high loadings by the following items: forum posts, content analysis of posts, time online, downloads, online user journey, university discussion forums, library loan statistics, and external data e.g., social media. This third factor explained 7.3% of the variance.

These three factors explained 40% of the variance in the data. The three factors identified were similar to those reported in the work of Ifenthaler and Schumacher (2016). They were used to better understand participants' data use preferences in the quantitative data as well as the qualitative data indicating the motivation for participants' data sharing preferences (discussed in Section 6.3.2.2).

Correlation tests were carried out to assess whether and how the study variables influenced each other. Study 3 investigated whether the privacy risk or benefit intervention influenced participants' data use and data sharing preferences, their concern about data use LA, as well as their perception of benefits from the use of student data for LA. Therefore, ANOVA tests were deemed suitable to examine whether there were any differences between the control and experimental groups (Strunk & Mwavita, 2020).

Study 3 implemented a pre-post test study design asking participants their data use preferences before and after the intervention. Thus, Chi-square tests were used to evaluate the statistical significance of changes students made to their data use preferences

Consequently, using the factors identified from the factor analysis, the results suggest that students were more willing to share data related to learning and academic history, followed by their digital exhaust data, and were reluctant to share personal data. This result is in line with the findings reported by Ifenthaler and Schumacher (2016) in their use of the sharing data scale at a German university.

A one-way ANOVA was used to test whether there were significant mean differences in data sharing preferences across the awareness intervention groups (control, privacy risks, benefits, privacy risks and benefits) following the intervention. The test revealed no significant differences among the means of the four groups (control, privacy risks, benefits, privacy risks and benefits) on participants' data sharing preferences ($F(3, 443) = 0.49$, $p = .69$).

6.3.1.3 Awareness of Privacy Risks and Benefits and Other Study Variables

Tests were also carried out to determine the influence of the privacy risks and/or benefits interventions on other dependent variables in Study 3. This involved assessing whether the interventions had an influence on participants' (i) concern about data use, (ii) concern about privacy risk, (iii) perceived usefulness, and (iv) benefit perception. The results are discussed in the following sections.

Concern about Data Use: The mean and standard deviation values for concern about data use for LA are shown in Table 6.5.

Table 6.5: Descriptive statistics of students' concern about data use for learning analytics by experimental group

	Descriptive statistics		
	Condition	Mean	Std. Deviation
	Control	2.55	1.01
Average concern about learning analytics	Privacy risks	2.62	0.92
	Benefits	2.57	0.90
	Privacy risks and benefits	2.62	0.90

Table 6.5 shows that overall, most participants were not necessarily worried about data uses for LA as presented in Study 3, as previously found in Study 2 (reported in Chapter 5). However, there were slight differences in the mean values for concern about data use for LA across the four groups. The mean values for the privacy risks and privacy risks and benefits groups were the highest and the same and the control group had the lowest mean. A chi-square test revealed that these differences in the mean scores were not statistically significant (X_2(12, N=447) = 13.73, p = 0.32). This suggests that an awareness of the possible privacy risks and benefits of data use did not influence participants' concern about the use of student data for LA.

Concern about Privacy Risk: Descriptive statistics of participants' concern about privacy risk were also obtained for participants in the privacy risks and privacy risks and benefits groups. As can be seen in Table 6.6, there was a slight difference in the mean value comparing the privacy risks and benefits group to the privacy risks group.

Table 6.6: Descriptive statistics of students' concern about privacy risk by experimental group

Descriptive statistics	Condition	Mean	Std. Deviation
Average risk perception	Privacy risks	2.9	1.11
	Privacy risks and benefits	2.97	1.14

There were no statistically significant differences between the groups based on students' risk perception.

Perceived Usefulness: The perceived usefulness variable was available from participants across the four experimental conditions. The mean and standard deviation values for perceived usefulness are shown in Table 6.7.

Table 6.7: Descriptive statistics of students' perceived usefulness of data use for learning analytics by experimental group

Descriptive statistics	Condition	Mean	Std. Deviation
	Control	4.05	0.73
	Privacy risks	4.13	0.77
Average perceived usefulness	Benefits	4.04	0.83
	Privacy risks and benefits	3.86	0.86

There were slight differences in the mean values for perceived usefulness across the four groups. The highest mean score was seen in the privacy risks group and the lowest in the privacy risks and benefits group. There was a slight difference in the mean scores between the control and benefits groups.

A chi-square test revealed that these differences in the mean scores were not statistically significant ($X_2(12, N=447) = 13.01$, $p = 0.37$). Thus, this finding suggests that an awareness of the possible privacy risks and benefits of data use did not influence participants' perceived usefulness of data use for LA.

Benefit Perception: The benefit perception variable was only available from participants in the benefits and privacy risks and benefits groups. The mean and standard deviation values for the benefit perception variable are shown in Table 6.8.

Table 6.8: Descriptive statistics of students' benefit perception by experimental group

	Descriptive statistics		
	Condition	Mean	Std. Deviation
Average benefit perception	Benefits	4.31	0.88
	Privacy risks and benefits	4.24	0.89

There was a higher mean value in the benefits group than in the privacy risks and benefits group and no differences between the groups based on students' benefit perception.

6.3.1.4   Potential Loss of Benefits and Participants' Data Use Preferences

To determine if the potential loss of benefits influenced participants' data use preferences (**RQ 7**), the change in data use preferences was assessed when the potential loss of benefits was made salient (that is, when participants were informed of the LA features that were no longer available based on their post-test data use preferences) comparing the treatment and control groups. Figure 6.9 represents the participants who indicated that they would (not) want to add more features to their learning dashboard, once the possible loss of benefits was made salient to them.

	No data		Self		Activities	
	Yes	No	Yes	No	Yes	No
■ Control	5.4%	10.9%	15.1%	21.9%	20.5%	26.0%
■ Privacy risks	1.4%	14.9%	5.9%	16.4%	26.8%	34.3%
■ Benefits	5.0%	10.0%	5.0%	15.0%	31.6%	33.3%
■ Privacy risks and benefits	1.5%	13.6%	4.5%	16.6%	31.8%	31.8%
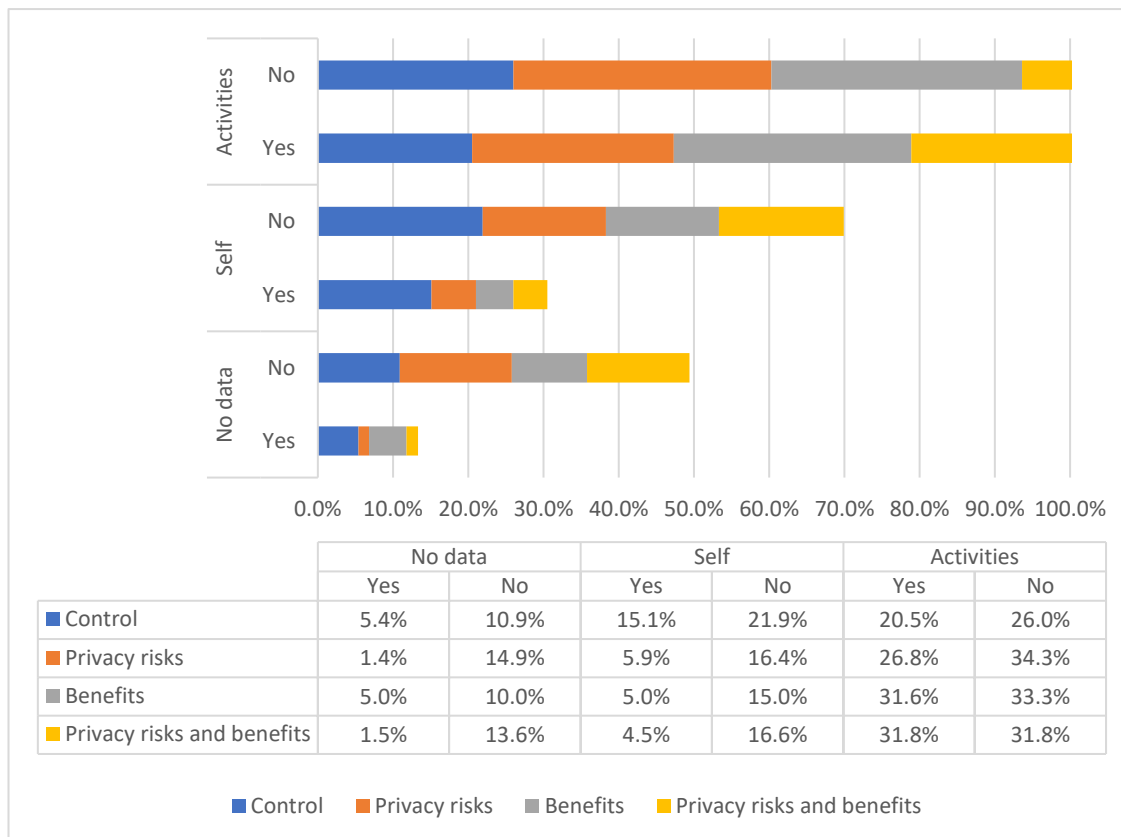
Figure 6.9: Number of participants who did (not) opt to add more features to the learning analytics dashboard

Figure 6.9 excludes those participants who, at the start of the study, indicated a preference to share both data about themselves and their activities on the online learning platform, since they were already sharing as much data as was possible. Additionally, the figure excludes those participants who stated that they needed more information.

Across the four groups, the highest proportion of participants willing and not willing to add more features were those who had chosen to share data about their activities on the online learning platform. Most of the participants who opted not to share their data, or to only share data about themselves retained the same preference even when the possible loss of benefits was made salient to them. These results suggest that the potential loss of benefits (LA features) did not influence participants to change their data use preferences and opt to share their data for LA. Once participants had made up their minds, they were unwilling to express a different data use preference, even when they were made aware of potentially deleterious effects to them.

Chi-square tests revealed no statistically significant differences given the changes observed in Figure 6.9, for those who opted to share no data ($X_2(3, N = 42) = 3.54, p = .32$), for those who opted to share data about themselves ($X_2(3, N = 68) = 2.14, p = .55$), and for those who opted to share data about their activities ($X_2(3, N = 156) = .47, p = .93$).

6.3.1.5 *Summary of Quantitative Results*

The quantitative results demonstrated that the privacy risks and benefits interventions did not influence participants' data use preferences, data sharing preferences, concern about data use for LA, perceived usefulness, benefit perception and concern about privacy risk. Additionally, the potential loss of benefits did not cause participants to significantly change their data use preferences. These findings suggest that participants' data use preferences are pre-determined and strongly held, and thus not easily changed.

In analysing the qualitative data, effort was made to understand the motivation for participants' data use and data sharing preferences. The results from the qualitative data are presented in the next section.

6.3.2 Qualitative Results

The qualitative results discussed in this section will focus on issues common to many participants in their responses to the Study 3 questions, as well as those which are clearly illustrative of the theme and are a key part to answering the question (**RQ 8**) of what motivated participants' data use and data sharing preferences (Mason, 2002). Examples of statements from participants are provided to give clarity to the theme and corresponding codes identified during the data analysis. Where these examples are provided, the following additional information is given: participant

number, experimental group participant was randomly assigned to, gender, data use preference, and whether participants changed their data use preference.

6.3.2.1 *Motivation for Participants' Post-test Data Use Preferences*

Participants were asked the reason for their post-test data use preferences. Across the four groups, participants provided a total of 416 open responses, corresponding to 10,972 words in total (26 words in length on average). The unit of analysis was each participant's response, and each response could be assigned multiple codes. Five hundred and sixty-one (561) codes were identified leading to two main themes: (i) support for institutional use of student data (42.8% of codes, n=240) and (ii) reservations about institutional use of student data (45.3% of codes, n = 254). Participants' responses were also coded to identify further information participants needed to determine their data use preferences (6.9% of codes, n = 39), as well as cases where participants expressed a change in opinion likely due to the study (2.9% of codes, n = 16). Finally, some participants' responses were either brief and consequently difficult to code, or did not adequately answer the question. These responses were coded as 'other' (2.1% of codes, n = 12). This section discusses the two themes and the observations made with respect to further information participants wanted as well as what they expressed in indicating a change in data use preference. The themes are defined and a summary of the codes across the four groups shown in Table 6.9 and the codebook is provided in Appendix C.6.

Table 6.9: Summary and definition of themes for motivation for participants' post-test data use preferences with corresponding number of codes per theme and experimental group

Theme	Definition of theme	Data use preference	%/no. codes from control group (*n* = 119)	%/no. codes from privacy risks group (*n* = 94)	%/no. codes from benefits group (*n* = 97)	%/no. codes from privacy risks and benefits group (*n* = 106)	Total
Theme 1: Support for institutional use of student data (42.8% of codes, *n*=240)	Statements where participants expressed support for institutional use of student data, while stating conditions they thought need to be considered alongside this use. The statements also demonstrated that some participants made a trade-off in the use of their data in exchange for a service from the institution.	No data	0.18 (1)	0.18 (1)	0.18 (1)	0.36 (2)	**0.9**
		Self	2.14 (12)	0.89 (5)	1.07 (6)	2.32 (13)	**6.42**
		Activities	2.50 (14)	2.85 (16)	1.78 (10)	2.67 (15)	**9.80**
		Self and activities	8.02 (45)	4.99 (28)	5.70 (32)	6.95 (39)	**25.66**
		Total	**12.84**	**8.91**	**8.73**	**12.30**	**42.78**
Theme 2: Reservations about institutional use of student data (45.3% of codes, *n* = 254)	Statements where participants were hesitant about the use of data for LA due to ethical and privacy considerations. Participants also focused on whether and how student data is (not) used and expressed a desire for boundaries in the use of student data.	No data	2.85 (16)	3.21 (18)	1.96 (11)	3.39 (19)	**11.41**
		Self	5.35 (30)	2.32 (13)	2.14 (12)	1.96 (11)	**11.77**
		Activities	3.74 (21)	3.39 (19)	6.60 (37)	4.28 (24)	**18.01**
		Self and activities	0.71 (4)	0.53 (3)	1.25 (7)	1.60 (9)	**4.09**

		Total	12.65	9.45	11.95	11.23	45.28
Further information participants needed	Statements where participants indicated what further information they would need to determine their data use preference		23.07 (9)	35.89 (14)	28.2 (11)	12.82 (5)	**6.9%**
Change in opinion	This category contains cases where participants indicated that they had had a change of their data use preference or opinion having gone through the study. The change was either negative or positive, however, the negative changes were dominant.		**50% (8)**	**12.5 % (2)**	**25% (4)**	**12.5% (2)**	**2.9%**
Other	Statements unrelated to the motivation for participants' data use preferences		**33.3% (4)**	**25% (3)**	**16.6% (2)**	**25% (3)**	**2.1%**

The highest percentage of codes across the control group and the three experimental groups related to the second theme – reservations about institutional use of student data. This second theme had a high proportion of codes from participants who were willing to share data about their activities (18.0%), followed by codes from participants who were willing to share data about themselves (11.8%) and codes from participants unwilling to share any data (11.4%). The lowest proportion of codes was identified from participants who were willing to share both data about themselves and their activities (4.1%) suggesting that these participants had fewer reservations compared to other participants.

With respect to the first theme – support for institutional use of student data – the highest proportion of codes was identified from those participants willing to share data about themselves and their activities (25.7%). There were fewer codes identified from participants who were willing to share data about their activities (9.8%), data about themselves (6.4%) and, as would be expected in this theme, very few codes from those not willing to share data (0.9%).

Comparing the codes across the four study groups revealed that the highest proportion of codes came from participants in the control group in support of (12.8%) and reserved about (12.7%) institutional use of student data. This was followed by codes from participants in the privacy risks and benefits group who supported the institutional use of student data (12.3%). An interesting observation was the high proportion of codes from participants in the benefits group who were reserved about institutional use of student data (11.9%). Further insights to explain these observations are discussed in this section.

Theme 1: Support for Institutional Use of Student Data

Participants' responses indicated their *support for institutional use of student data* to improve teaching and learning (42.8% of codes, *n*=240). The *data used, tools and students' experience* were observed as a motivator for participants' data use preferences (79.6% of codes, *n* = 191 (out of 240 codes in this theme)). At the same time, they indicated their expectations for *purpose and scope limitation and appropriate support* (8.8% of codes, *n* = 21 (out of 240 codes)). However, participants were observed to make *trade-offs in data use for benefits* even while supporting the use of student data (4.2% of codes, *n* = 10 (out of 240 codes)). It is noted that not all participants shared all three concerns equally.

Data Used, Tools, and Students' Experience

Participants' reasons for their data use preferences were observed to focus on the *data* (n = 86 (out of 191 codes))*,* on the *learning analytics tools* (n = 58 (out of 191 codes))*,* and on the *students' experience* (n = 38 (out of 191 codes))*.*

Participants made numerous statements indicating that the data shared should be *sufficient* or *appropriate for the stated purposes* (19.4% of codes, *n* = 37 (out of 191 codes)). Their perception of the data being sufficient took on several forms, for example, they wanted to share what was *most relevant* (7.3% of codes, *n* = 14 (out of 191 codes))*, was least invasive* (3.7% of codes, *n* = 7 (out of 191 codes)), *felt comfortable or safe for them to share* (3.7% of codes, *n* = 7 (out of 191 codes)), or *what they thought showed their engagement with their studies* (2.1% of codes, *n* = 4 (out of 191 codes)). As one participant stated:

> *"Because that is directly related to my learning and doesn't take into consideration other factors which may not assess academic performance."*
>
> *[P161, Risks group, Female, willing to share data about activities, no change]*

Participants perceived an opportunity to *improve the learning experience for themselves and for other students* (19.9% of codes, *n* = 38 (out of 191 codes))**.** One noteworthy perspective was offered by a participant who framed the process of being monitored as motivating:

> *"I think it will be beneficial if I allow some of my data to be used because then I can receive the best support and recommendations. I also think it will ensure I stay productive and motivated knowing that my data is recorded and monitored to some extent."*
>
> *[P266, Privacy risks and benefits group, Female, willing to share data about self and activities, no change]*

This particular response frames monitoring or tracking of students' activities on the online learning platform positively, and as having a motivational effect on this student, in contrast to the perspective held by other participants in the study who did not want to be tracked and others who were concerned that they would adapt their behaviour and only carry out those activities that they knew were being measured.

Yet another reason for participants' data use preferences was that they saw *no harm in sharing (some) data* (4.2% of codes, n = 8 (out of 191 codes)), perhaps because the data use indicated in Study 3 met their expectations. As one participant said:

> *"I don't mind if data about myself is used, and it makes sense that my activities would be analysed. I don't see any potential drawbacks."*
>
> *[P25, Privacy risks group, Male, willing to share data about self and activities, no change]*

In contrast, another participant thought to share a little data to *minimise any potential harm*:

> *"I don't see any harm in sharing some small data about myself, but further information which relates more to personality perhaps feels like it could result in bias."*

> *[P361, Control group, Female, willing to share data about self, no change]*

These two preceding examples demonstrate differing yet similar perspectives on possible harms to privacy arising from use of student data. The shared understanding in these two responses is that the participants do not perceive any possible harm. The use of data as described in Study 3 was not unexpected for the participant in the privacy risks group, while the participant in the control group held the view that there was no harm, but one still needed to be cautious. These views highlight some of the underlying assumptions held by participants. In stating that they saw no harm in sharing their data, participants might have lacked knowledge about how data use can cause harm or may not have cared about the issue. Additionally, instead of participants choosing not to share any data, they may instead have opted to share the data they thought was harmless. This view that data use can be harmless might have been driven by the context where the data was used, which was by the learning institution and focused on supporting students' learning. These two examples further highlight a noteworthy tension where both perceive little chance of harm but go on to express different data use preferences.

Purpose and Scope Limitation and Appropriate Support

Participants expected that *purpose limitation* (47.6% of codes, n = 10 (out of 21 codes)) would be put in place, for example, that the institution would only use academic data for academic purposes:

> *"As I indicated before, I do not think it is appropriate to use data about a student's private life and background to make a judgement on their academic performance. It is not fair to do so, as it could lead to discrimination and unfair bias. A student's academic performance and private life should be separate and it is not the place of the university to be able to access that data or use it to judge a person's abilities. Their abilities should be judged solely on their present engagement with the course and their previous academic record."*

> *[P269, Control group, Female, willing to share data about activities, no change]*

Akin to the perspective indicated in Cormack's (2016) data protection framework, participants expressed the view that personal data about students such as about their health could be useful to provide context about their learning, however, it needed to be handled appropriately, for

instance, being directed to relevant staff who could provide the necessary assistance, rather than teaching staff:

> *"I think that for the purposes of education and student development, the software should have access related to student's academic activities. If there are problems unrelated to academic studies, family problems, or other health issues that impact student's performance, the software may be useful in identifying the problem, however, these should be essentially addressed in person with counsellors and academic tutors...."*
>
> *[P325, Privacy risks and benefits group, Male, willing to share data about activities, no change]*

Another expectation was that the HEI would *provide students with assurance and appropriate support* (23.8% of codes, *n* = 5 (out of 21 codes))*,* including *opportunities to opt out of data use* (9.5% of codes, *n* = 2 (out of 21 codes))*:*

> *"I personally don't mind an institution collecting my data for analytics (as long as there are appropriate data safety regulations). Therefore, I am more inclined to share all the data for them to provide me with suggestions to benefit my learning - ideas which I may not have considered and the final decision is always up to me anyways on whether I use it or not."*
>
> *[P367, Privacy risks and benefits group, Male, willing to share data about self and activities, no change]*

Overall, students' responses demonstrated an expectation that the HEI would act in students' best interests.

Trade-offs in the Use of Student Data

Participants indicated that they would share just *enough data to protect privacy* (40% of codes, *n* = 4 (out of 10 codes)), and that they sought the *best balance between privacy and services for students (*20% of codes, *n* = 2 (out of 10 codes))*.* They thought they were *getting something back for their information* (20% of codes, *n* = 2 (out of 10 codes))*,* and that *the benefits outweighed the privacy risks* (10% of codes, *n* = 1 (out of 10 codes)). As one participant stated:

> *"I think it will provide the most amount of data and will therefore allow the tool to be as accurate as possible. There is no point creating a tool with little to no data as it's predictions will not be as good. For me the benefits of having such a tool outweigh the privacy issues when using the data specified on the previous page."*

[P21, Benefits group, Female, willing to share data about self and activities, no change]

It was evident that the trade-offs were an attempt to hold some data back to keep the *balance* at what participants felt was an acceptable level for them (20% of codes, *n* = 2 (out of 10 codes)):

"There is a breach of privacy I feel if there were to be any more information revealed."

[P211, Control group, Female, willing to share data about activities, no change]

"Any more would directly disregard my privacy."

[P336, Benefits group, Male, willing to share data about self, no change]

These two preceding examples demonstrate that participants thought differently about how sharing different data would compromise their privacy. The participant in the control group felt that their privacy would be compromised if they revealed data about themselves and consequently chose to share their activity data, yet the participant in the benefits group shared data about themselves and felt that additionally sharing their activity data would compromise their privacy. This is an important lesson that in allowing students to determine *which* data to share, HEIs should scaffold this support with information about privacy risks related to different data types to allow more informed decision-making.

Theme 2: Reservations about Institutional Use of Student Data

Despite participants' willingness to share some or all of their data (as seen in the quantitative data analysis in Chapter 6.3.1), and despite the support for institutional use of student data captured in the first theme, participants, as summarised in Table 6.9, additionally expressed *reservations about institutional use of student data* (45.3% of codes, *n* = 254). Participants raised *ethical and privacy considerations* (44.1% of codes, *n* = 112 (out of 254 codes)). They pointed to the *irrelevance of personal data for learning analytics and a preference for anonymity* (32.7% of codes, *n* = 83 (out of 254 codes)). There was *tension between understanding data use and discomfort with data use* (12.2% of codes, *n* = 31 (out of 254 codes)) where participants were seemingly of two minds about the use of student data. Finally, a small number of participants expressed a *preference for boundaries* or *separation* in data use across their personal lives and their lives as students (7.9% of codes, *n* = 20 (out of 254 codes)).

Ethical and Privacy Considerations

Reservations about institutional use of student data were due to *ethical and privacy considerations* (44.1% of codes, *n* = 112 (out of 254 codes)). Participants' responses captured their *concern about (potential) bias, discrimination, or prejudice* (16.1% of codes, *n* = 18 (out of 112 codes)). These concerns were shared by participants from all four groups who additionally expressed different data use preferences, thereby suggesting that willingness to share no data, some data, or all data for LA may not correspond with the presence or absence of concern about data use.

Additionally, participants raised concerns that institutional use of student data as described in the study could *negatively impact students* (14.28% of codes, *n* = 16 (out of 112 codes)), for instance that those from disadvantaged backgrounds would be treated unfairly, or students would be pressured to behave in a certain way*:*

> *"With more information, I could determine that the personal information used would be almost a breach of my privacy, and even giving away data about my use of the learning platform is somewhat private to me, as I would like to privately access learning materials without feeling pressure (for example if I downloaded some materials a little late in the course, or past a deadline)."*
>
> *[P425, Control group, Male, willing to share activity data, Change (activities to no data)]*

Finally, other concerns were raised by participants across the four groups as well as by individual students, including that the data use was *privacy invasive* (6.3% of codes, *n* = 7 (out of 112 codes)), and that the data could only give a *partial picture of the student* (6.3% of codes, *n* = 7 (out of 112 codes))*.*

Irrelevance of Personal Data for Learning and Preference for Anonymity

Participants shared the perspective that *personal details were not relevant for learning or should not be shared* (68.7% of codes, *n* = 57 (out of 83 codes)):

> *"I don't think it is necessary for the personal details to be shared as it is not about that, it's about what the person does regardless of, for arguments sake, their gender."*
>
> *[P115, Risks group, Female, willing to share data about activities, no change]*

A *preference for anonymity* (10.8% of codes, *n* = 9 (out of 83 codes)) was another reason why participants hesitated to share their data.

Tension between Understanding Need for Data Use and Discomfort with Data Use

With *the tension between understanding data use and discomfort with data use* (12.2% of codes, *n* = 31 (out of 254 codes)), participants were seen to express an understanding, for instance, that institutional use of student data was necessary, alongside seemingly contradictory views, such as expressing corresponding concerns about discrimination, or a sense that the data use was privacy invasive. A few examples are given in Table 6.10 to illustrate these views captured from participants' responses.

Table 6.10: Examples from participants' responses of tensions between understanding the need for institutional data use and discomfort with data use

Category	Example from participants' responses
No harm in sharing AND concerned about discrimination	"I don't see any harm in sharing some small data about myself, but further information which relates more to personality perhaps feels like it could result in bias." [P361, Control group, Female, willing to share data about self, no change]
Understanding AND does not want to be tracked	"I don't mind giving basic information about myself since that would be fairly easy to get anyway, but I do not like to have everything about me being tracked even it could have some minor benefits to helping me improve my performance." [P424, Privacy risks and benefits group, Female, willing to share data about self, no change]
Comfortable sharing data BUT seems too much	"I would be comfortable with sharing pieces of information from each category, however not all as this seems like too much to give away." [P352, Control group, Male, willing to share data about self and activities, no change]
Privacy invasive BUT potentially useful	"Because much of the personal data is irrelevant to studying and feels like an infringement on privacy in the name of education. However, certain (but not all) aspects of activities on the online learning platform could be genuinely useful, e.g., detecting whether a part of a module was inadvertently overlooked." [P373, Benefits group, Male, willing to share data about activities, no change]

Preference for a Boundary or Separation

Finally, there was a *preference for a boundary or separation* (7.9% of codes, *n* = 20 (out of 254 codes)) between students' academic and personal lives when it came to institutional use of student data. Participants wanted to keep *academic and private life separate,* or their *online activity separate from student life.* Three individual responses of interest are highlighted. The first response was from a participant who expressed an expectation that *some student data would not be used*:

> *"At first, I thought it might be a good idea to share some data, but I believe that the suggested options of the data shared/what will be done with it oversteps it's boundaries and could have negative effects on performance and mental health. I believe that if the only outcome of the data collection was to improve learning by providing support, then I'd be alright with sharing some of the suggested data."*
>
> *[P326, Control group, Female, not willing to share data, change (self and activities to no data)]*

The second response was where a participant indicated their expectation that *boundaries would be maintained* between different aspects of students' lives:

> *"My life outside of the learning platform does not seem applicable to predicting my grades."*
>
> *[P186, Benefits group, Male, willing to share data about activities, change (self and activities to activities)]*

The final response was from a participant who expressed an expectation about the type of data and the *locational boundaries:*

> *"I don't mind my data being used as long as the data is relevant, for example I don't want personal data from when I'm at home to be used but any data collected when I'm in college is fine."*
>
> *[P35, Benefits group, Male, willing to share data about self and activities, no change]*

Observation 1: A noteworthy observation from participants' responses was that there was a *change in opinion* (2.9% of codes, *n* = 16 (out of 561 codes)) due to the information about Study 3. The change in opinion was predominantly *negative* where participants indicated that their awareness about how data was used negatively influenced their willingness to share data:

"Looking at the type of personal data you want to use has put me off."

[P100, Control group, Female, willing to share data about activities, change (self and activities to activities)]

In contrast, another participant who was in the control group and thus was not made aware of the possible privacy risks and benefits given the use of their data for LA expressed a *positive* view that led to them changing their opinion:

"I've changed my mind a little after learning more. I think it would be helpful to my studies to receive tailored suggestions for what I might need help with, based on an analysis of my activities."

[P194, Control group, Male, willing to share data about self and activities, change (self to self and activities)]

Changes in opinion are to be expected in privacy research as it makes participants aware of issues they may not have previously considered (Krol, Spring, Parkin, & Sasse, 2016).

Observation 2: A total of 26 (5.8%) participants across all four groups (Control – 6 (23.1%); Privacy risks – 8 (30.8%); Benefits – 7 (26.9%); Privacy risks and benefits – 5 (19.2%)) indicated that they wanted *more information*. Participants predominantly raised questions that they wanted answered. One participant indicated that they would prefer to choose what personal data is used. Proposals to address these issues are discussed in Section 6.4.

The next section presents the results for the motivation for participants' data sharing preferences.

6.3.2.2   Motivation for Participants' Data Sharing Preferences

Across the four groups, a total of 422 open responses were analysed from participants' data sharing preferences. Seven hundred and thirteen (713) codes were identified, resulting in three main themes: (i) acceptance of data sharing (24.6% of codes, $n = 174$), (ii) concerns and questions about the impact of data sharing (31.7% of codes, $n = 224$), and (iii) (mis)match in students' expectations of data sharing (43.6% of codes, $n = 308$). A few participants indicated responses that were brief and therefore difficult to code or did not clearly answer the question asked. Therefore, these were coded as 'other' (0.9% of codes, $n = 7$). The themes are defined and a summary of the codes across the four groups shown in Table 6.11 and the codebook is provided in Appendix C.7. Note that the first and second themes and some of the codes are similar to those identified from participants' data use preferences (see Section 6.3.2.1) with the exception of the

third theme. As such, this suggests that participants' data use and data sharing preferences may be motivated by related factors.

Table 6.11: Summary and definition of themes for motivation for participants' data sharing preferences with corresponding number of codes per theme and experimental group

Theme	Definition of theme	%/no. codes from control group (n = 121)	%/no. codes from privacy risks group (n = 92)	%/no. codes from benefits group (n = 100)	%/no. codes from privacy risks and benefits group (n = 109)	Total
Theme 1: Acceptance of data sharing (24.6% of codes, n=174)	Statements where participants acknowledged the relevance of some of the data used for learning analytics and agreed to data sharing while indicating several conditions to be met.	5.6 (40)	3.9 (28)	6.5 (46)	8.4 (60)	**24.4 (174)**
Theme 2: Concerns and questions about the impact of data sharing (31.7% of codes, n = 224)	Statements related to concerns participants expressed including ethical and privacy considerations.	8.7 (62)	7.0 (50)	8.9 (64)	6.7 (48)	**31.4 (224)**
Theme 3: (Mis)match in students' expectations of data sharing (43.6% of codes, n = 308)	Statements where participants stated their expectations about how their data would (not) be used thereby highlighting (mis)matches between participants' expectations and the data that is actually used.	11.1 (79)	8.7 (62)	10.8 (77)	12.6 (90)	**43.2 (308)**
	Total	**25.4 (181)**	**19.6 (140)**	**26.2 (187)**	**27.7 (198)**	
Other						**0.9%**

The highest percentage of codes across the control group and the three experimental groups related to the third theme – (mis)match in students' expectations of data use. This third theme had a high proportion of codes from participants in the privacy risks and benefits group (12.6%). The lowest proportion of codes was identified from participants in the privacy risks group (8.7%). With respect to the second theme – concerns and questions about impact of data use – the highest proportion of codes was identified from participants in the benefits group (8.9%) and the lowest proportion of codes from participants in the privacy risks and benefits group (6.7%). For the first theme – acceptance of data use – the highest proportion of codes were identified from participants in the privacy risks and benefits group (8.4%), and the lowest proportion of codes from participants in the privacy risks group (3.9%). Further insights to explain these observations are discussed in this section.

Theme 1: Acceptance of Data Use

Participants who made statements indicating that they *accepted institutional use of student data* (24.6% of codes, *n*=174), were *supportive of data use* (50% of codes, *n* = 87 (out of 174 codes)). However, this did not seem to be support for the use of all data, despite participants indicating a preference to share both data about themselves and data about their activities on the learning platform. For example, one participant stated:

> *"As long as the data is held appropriately and only for the reasons outlined, I feel comfortable with all data except the external data. The reason I don't feel comfortable with the external data is because it doesn't feel relevant to what is being created with the data. Other than that, I can see the benefit to myself of providing the other data, so I am happy for this to be used."*

> *[P13, Benefits group, Female, willing to share data about self and activities, no change]*

The misgivings observed in the statement above might have contributed to participants *agreeing to data use under several conditions* (47.1% of codes, *n* = 82 (out of 174 codes)), for example, indicating that the data they shared was *relevant* or *useful for their learning*:

> *"I agree with the ones to do with my learning. The ones that do not affect my learning I have chosen no on."*

> *[P120, Control group, Male, willing to share data about activities, no change]*

Other relevant *conditions* (20.7% of codes, *n* = 17 (out of 82 codes)) influencing participants' data sharing preferences included *if data use helped them and other students* (29.4% of codes, *n* = 5 (out of 17 codes)), *if the data was anonymised* (17.6% of codes, *n* = 3 (out of 17 codes)), and *if the*

data was stored properly (11.8% of codes, *n* = 2 (out of 17 codes)). These particular codes highlighted the interest participants had in the benefits available for the use of their data, but additionally, the need to address concerns they might have, in this case with respect to being identified using their data, or having their data stolen.

The fact that the *university already has student data* or *some of the data is already public* (3.4% of codes, *n* = 6 (out of 174 codes)) was raised by participants:

> *"Characteristics such as your name are already public knowledge but other characteristics such as employment which is not related to your studies should not be considered… They're not information that is confidential or private, so I wouldn't mind if it was used to help me in my studies."*
>
> *[P428, Privacy risks and benefits group, Male, not willing to share data, no change]*

This view suggests that students may not be explicitly aware of the new knowledge about them that LA can create. The datafication model of privacy (Mai, 2016) focuses on new personal information created by processing and analysing data. While this new knowledge can be used to benefit students, there is the potential for it to harm them, for instance, if they are placed in particular categories which determine their access to educational support (Yeung, 2018; Reidenberg & Schaub, 2018; Solove, 2009).

Finally, *trust in their college/university* (1.7% of codes, *n* = 3 (out of 174 codes)) was an additional factor contributing to participants' data sharing preferences:

> *"I agree to all of these as I would trust it to use my information for what it says it is used for."*
>
> *[P5, Control, Female, willing to share data about self and activities, no change]*

It was noteworthy that only three participants referred to trust, which suggests that trust may in fact play a minor role in students' data sharing preferences. This contrasts with findings in related research (Slade & Prinsloo, 2014) where trust in the learning institution plays a greater role.

Theme 2: Concerns and Questions about the Impact of Data Use

Numerous *concerns and questions* (31.7% of codes, *n* = 224) were raised about how data use could impact students. *Concerns* (58.0 of codes, *n*=130 (out of 224 codes)) included the possible *harm to students* (32.30% of codes, *n*=42 (out of 130 codes)) such as discouraging them or lowering their confidence:

> *"I believe some of the data could result in poorer predictions for disadvantaged students (e.g., parents' education level, type of school attended, income). I think this could potentially result in students being discouraged and could be more likely to drop out of the program or become less motivated in their studies."*
>
> *[P142, Control group, Female, willing to share data about self and activities, no change]*

Bias (18.5% of codes, *n*=24 (out of 130 codes)) was yet another concern expressed by participants. They were concerned that the algorithms would be biased, or that they would experience bias from their lecturers:

> *"I am generally okay with studies related information being used and what may hinder my performance and why there might be problems at some points, so information about my previous achievements, number of courses, jobs I have during uni, financial status etc, what I do on study forums, which material I deem important to download from study material etc. I do not agree with the school keeping these details together with my name and address because I would not want to face any bias from lecturers etc. I also do not want them to have information of what I do and look for outside of when using their school sites and study sites from the university."*
>
> *[P74, Risks group, Female, willing to share data about self and activities, change (self and activities to activities)]*

> *"Some types of data are too personal e.g., parents' education level is strongly correlated with whether students go to university. Will there be unconscious bias in the AI/machine learning/analytics technology used for disadvantaged students?"*
>
> *[P297, Control group, Female, willing to share data about self and activities, change (self and activities to activities)]*

It is noted, however, that some of the concerns may point to participants' not being fully informed about how student data is used, for example, that in some algorithms used in LA, the students' performance over time has greater weight than their demographic data (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015).

> *"I think basing study recommendations on your engagement with university resources is reasonable but basing it on past performance before university or demographics is an unhealthy way to treat your students and is ripe for embedded bias in your algorithm."*
>
> *[P278, Benefits group, Male, willing to share data about activities, no change]*

Other concerns included *some of the data used causing discomfort* (13.1% of codes, *n*=17 (out of 130 codes)), and participants feeling that the data use was *invasive/intrusive* (12.3% of codes, *n*=16 (out of 130 codes)).

Privacy considerations were also noted as *privacy was noted to matter to participants in different ways* (24.6% of codes, *n* = 55 (out of 224 codes)). These considerations were raised by participants in the four groups including them *thinking it invades privacy* (40% of codes, *n* = 22 (out of 55 codes)):

> *"Some of these are completely unacceptable indicators of student progress and unnecessarily invade students' privacy."*
>
> *[P334, Risks group, Male, willing to share data about activities, no change]*

There was a *preference for data to remain private* (14.5% of codes, *n* = 8 (out of 55 codes)) and *participants valuing privacy* (18.2% of codes, *n* = 10 (out of 55 codes)).

With the code *crossing a boundary* (12.9% of codes, *n* = 29 (out of 224 codes)), participants wanted boundaries in how student data was used. Participants distinguished between types of data collected (for instance, personal and non-personal), and the environments where the data was collected from (for instance, the university and home):

> *"I feel like some of the more personally identifying information is a problem when it leaks outside of the university campus. It could be used in an erroneous way to potentially profit off my data without my consent. I also feel like certain things belong publicly or with institutions like universities, and some only to myself (e.g., tests, in my view, which are administered by the university also belong to the university as well as myself while something as personal as my family details, birth date and full name belong only to myself)."*
>
> *[P216, Risks group, Male, not willing to share data, no change]*

Student data as an incomplete snapshot (4% of codes, *n* = 9 (out of 224 codes)) captured participants' sense that the data only shows a partial picture of the student and their abilities. Participants indicated that the data *does not tell you everything about a student* (33.3% of codes,

n = 3 (out of 9 codes)), *does not accurately represent the person* (44.4% of codes, *n* = 4 (out of 9 codes))*,* and *does not indicate the student's ability* (22.2% of codes, *n* = 2 (out of 9 codes)):

> *"… Also, I do not think it is fair to base predictions upon downloads from the university platform - some people may be unorganised and download them multiple times whereas others may only download them once and log in very few times because they are more organised - but this would reflect badly upon them."*

> *[P282, Control group, Female, willing to share data about self, change (self to no data)]*

In line with this code, Slade and Prinsloo's (2013) ethical framework for LA discusses how student identity is temporary and changes over time with respect to the data that is collected and used for LA.

Theme 3: (Mis)match in Students' Expectations of Data Use

The third theme captured a *(mis)match in students' expectations of data use* (43.6% of codes, *n* = 308), as participants stated their expectations about how their data would (not) be used. It was clear that numerous students across the four groups felt that some of the *(personal) data selected was irrelevant* (63.6% of codes, *n* = 196 (out of 308 codes))*:*

> *"I have agreed to most of the things that would help with a learning analytics tool targeted at myself, whereas I have disagreed or put not applicable to things I would consider unnecessary information about myself or information that I would not be willing to have recorded and used by a learning analytics tool."*

> *[P92, Benefits group, Male, willing to share data about self and activities, no change]*

Others specifically highlighted the *personal nature of the data used* (26.3% of codes, n = 81 (out of 308 codes)):

> *"I wouldn't want to reveal too much information about my personal life if it is irrelevant to my studies."*

> *[P182, Risks group, Female, willing to share data about activities, no change]*

Participants also pointed out that some of the data was not the student's rather it was *other people's data* (3.2% of codes, *n* = 10 (out of 308 codes)). As such, participants felt that the data should not be used*:*

> *"I would be happy to provide information regarded to my education, such as previous grades or anything that can aid my academics and enhance my future learning, however I*

am less inclined to share more personal data, and data that is not initially my own, such as … some details about parents."

[P152, Privacy risks and benefits group, Female, willing to share data about self, change (self to self and activities)]

"Some things are not my data to give away."

[P128, Control group, Male, willing to share data about self and activities, change (self and activities to self)]

Yet another important consideration for participants was that *learning analytics may constrain student support or options if students were labelled* (2.6% of codes, *n* = 8 (out of 307 codes)). In general, participants indicated that if their background or parents' background was taken into *consideration then their study options might be unfairly restricted:*

"I think quite a few of them have the danger of restricting people's ability to achieve based on irrelevant factors such as employment, loan statistics, parents' education level which means that it further instils disadvantages and inherent biases."

[P402, Control group, Female, willing to share activities, no change]

One participant also indicated a *desire for a standard rather than a personalised experience*. Despite only being raised by one participant in the study, this response is highlighted as a unique perspective that is not shared by many other participants in the study:

"I value privacy significantly. Plus, I actually desire a standard experience - I don't want to be oriented and regimented into certain areas. I need a general overview of everything in my field and I want everything to have some kind of fun useful difficulty to it, rather than prematurely specialise."

[P443, Privacy risks and benefits group, Male, not willing to share data, no change]

The next section presents the results of what motivated participants' responses when they were made aware of the potential loss of LA benefits.

6.3.2.3   *Motivation for Participants To (Not) Choose to Add Excluded Features*

This section assesses the reasons participants gave for (not) adding excluded features to the learning dashboard. First, reasons for adding excluded features are discussed, followed by reasons for not adding excluded features. The reasons participants provided are summarised in Appendix C.8.

Reasons for Adding Features

There were 131 codes identified from participants' responses highlighting their reasons for adding features that were excluded from the learning dashboard. Those participants who opted to add more features did so because they thought the *learning resources and recommendations were useful* (67.2% of codes, *n* = 88 (out of 131 codes)). However, the feature to predict students' performance was not favourable to several participants in the study, as one stated:

> *"It's more for the personalised recommendation. I'm not too fond of the concept of predicting my grades but it might be a personal preference."*

> *[P138, Benefits group, Male, willing to share data about activities, no change]*

Predictive analytics has often been designed for use by staff to help them prepare how to support students who need it, for example with OU Analyse (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015). As such, the output of predictive analytics tools is often directed to student support teams. However, it is important to be transparent with students about the different ways in which student data can be used. Doing so can help identify why they may be concerned about it, and where possible, these concerns can be addressed.

Participants expressed additional *preferences* (4.6% of codes, *n* = 6 (out of 131 codes)), for instance, that *student data was protected*, or that the *amount of data used was limited*:

> *"I think it would be more beneficial to have personalised ways to improve your learning, but I think it's important to limit the amount of data used."*

> *[P316, Privacy risks and benefits group, Female, willing to share data about self and activities, change (self and activities to activities)]*

Finally, participants also raised *exceptions* (15.3% of codes, *n* = 20 (out of 131 codes)) as they added more features, including that they were concerned about the *impact on students,* either making them over-confident or demotivating them.

Only one participant questioned the 'all-or-nothing' approach surrounding data use and LA features:

> *"It feels like I'm being given an all or nothing choice - either give us all of your data or none of it. Why can't I still get personalised recommendations and predictions without having to divulge data like my income or marital status?"*

> *[P296, Benefits group, Female, willing to share data about activities, no change]*

This is an issue that needs investigation in future work given that students might not be willing to share data or only want to share some data, yet they should not be disadvantaged compared to other students. Furthermore, students who choose to share their data might question it being used to support students who opt out of data sharing.

Reasons for Not Adding Features

There were 179 codes identified from participants' responses indicated why they did not add features that were excluded from the learning dashboard. Participants thought the *features were not beneficial* (17.9% of codes, *n* = 32 (out of 179))*,* with one participant indicating they already accessed additional reading, suggesting that the offer made to students may not be what can make the trade-off truly beneficial in their view:

> *"Personalised recommended reading would be useful, but I normally go over all recommended reading anyway to choose what I believe applies to me anyway."*
>
> *[P267, Benefits group, Female, willing to share data about self and activities, change (self and activities to activities]*

This suggests that for some students, the potential benefits of LA may need to extend beyond what students think they can achieve by themselves, or support that they already have access to.

Participants also indicated that *the features provided were sufficient* (16.2% of codes, *n* = 29 (out of 179)). Others felt they had *provided enough data or preferred not to provide (additional) data* (8.4% of codes, *n* = 15 (out of 179))*:*

> *"The inclusion of personalised learning recommendations does seem very useful, but again, I'm not willing to give them the information for that to be possible. So, I don't want to add the excluded features to my dashboard."*
>
> *[P374, Benefits group, Female, willing to share data about activities, change (activities to no data)]*

Finally, they were concerned about the *impact on students* (16.8% of codes, *n* = 30 (out of 179)), for instance that students would be *demotivated by a predicted grade* (8.4 of codes, *n* = 15 (out of 179)). Some participants indicated that *students should have responsibility for their learning* (8.4% of codes, *n* = 15 (out of 179)) or thought that it *reduces the students' responsibility for their learning* (0.6% of codes, n = 1 (out of 179))*.*

> *"I don't see how a personalised recommendation can be made using the features and I simply think it would be better to provide students with all of the recommended texts so*

they can learn for themselves what works best for them rather than an algorithm saying

this is how you learn most optimally."

[P274, Privacy risks and benefits group, Male, willing to share data about self, no change]

It is noted that some of the objections raised by participants might be due to the novelty of LA features such that students are yet to use them for their own learning.

6.3.2.4 Summary of Qualitative Results

The qualitative results have provided in-depth insights into the motives for participants' data use and data sharing preferences. They have demonstrated contrasting responses from participants where they both accept and yet are reserved about institutional use of student data. Additionally, the results have demonstrated cases of (mis)match in students' expectations and practices in institutional use of student data. These contrasting perspectives were sometimes identified from the same participant, showing that their views are not polarised with individuals on opposing ends. Instead, participants seem to be uncertain or yet to determine what to think about institutional use of student data.

6.4 Study 3 Discussion

6.4.1 Implications of Findings

While there is a lot of empirical research that has been carried out on students and privacy in LA (as summarised in Section 2.4), Study 3 is the first to carry out a large-scale experimental study to explore how changing privacy risks and benefits might influence students' data use preferences. In designing Study 3, the expectation was that making students aware of the possible privacy risks and benefits of data use for LA would influence their data use preferences. Furthermore, it was anticipated that making students aware of the potential loss of benefits tied to their data use preferences would influence the same. Finally, Study 3 sought to identify the motives for participants' data use and data sharing preferences.

With respect to RQ 6, the findings from Study 3 suggest that an awareness of the potential privacy risks and benefits of data use for LA did not significantly influence participants' data use and data sharing preferences. Consequently, there may be other factors, unrelated to a privacy risk or benefit assessment, which played a greater role in participants' data use preferences. For example, in the context of health data sharing, Schairer and her colleagues (2019) found that participants in their study were motivated by negative past experiences, or a desire to help others; reasons which extended beyond the privacy risks and benefits contained in the privacy calculus theory (Laufer & Wolfe, 1977; Dinev & Hart, 2006).

A potential contributing aspect to the result is that the privacy risk and benefits interventions used in Study 3 did not evoke a strong response from participants. The interventions were designed so they could possibly be used in real life by a HEI. Thus, this constraint determined that a privacy risk intervention that would evoke a strong emotional response such as fear, was unsuitable. Additionally, allowing Study 3 participants to control whether and how their data is used for LA might have lowered their privacy concern (Culnan & Armstrong, 1999) and therefore their concern about privacy risk.

Another insight from the results of Study 3 is that students' data use preferences are pre-determined and not easily modified. This insight arises from the finding that the possible loss of benefits resulting from the use of student data did not cause students to change their stated data use preferences. The benefits provided, concerns about privacy and data protection, as well as ethical considerations played a role in participants' choosing to change or not change their post-test data use preferences.

It was observed that there was a mismatch between students' data use preferences and their data sharing preferences. Specifically, while students might have indicated a preference to not share data, or only share data about themselves *or* their activities on the learning platform, they later went on to choose specific data items they were willing to share. Conversely, those who indicated a preference to share data about themselves *and* their activities on the learning platform went on to not share all data items. One possible explanation is that this highlights the positive role that both transparency and control over data could play in the LA context, and how important it is to pair the two in practice. By empowering students to control which data is used for LA, the lesson learned is that while students are willing to have their data used for LA, they might not be willing to have *all* their data used. There might be specific data items that students want to hold back or share with different entities.

Participants were observed to engage in choosing specific data items that they were comfortable sharing for LA. For instance, some opted to share data about themselves or data about their activities, while others preferred to share their address and not their library statistics data. Students' engagement in choosing which data to share points to them having an interest in privacy and how their data is used by HEIs. Consequently, the engagement observed in Study 3 (and in other work such as (Ifenthaler & Schumacher, 2016)) does not align with the narrative that students do not care about privacy in LA (Vu, Adkins, & Henderson, 2019). Instead, their apparent apathy could be influenced by design decisions or pedagogical decisions, for example, the absence of opportunities to control the use of their data or being in collaborative learning environments where

data sharing (between students and others) is necessary and/or required (Scheffel, Drachsler, Kreijns, de Kraker, & Specht, 2017).

The qualitative results revealed further privacy and utility tensions as participants are both *accepting* of and *reserved* about the use of student data for LA. This finding is especially relevant for contextual integrity (Nissenbaum, 2010) as it gives greater insight into students' norms of appropriateness in the LA context. For example, some participants thought their personal data, health data, or data linked to their family members was irrelevant for use to support their learning. Yet another example is seen in students' preference for some separation between their student and personal lives. Other privacy and utility tensions were identified in the finding that data use in LA might not match students' expectations. Taken together, these insights lead to a recommendation that HEIs need to provide students with further information on the data used for LA and opportunities for them to control the data that is used and to make future updates if necessary, to take steps towards providing informed consent.

Participants' perspectives on boundaries and separation between different parts of their lives illustrate their expectations of the practices the HEI will engage in. These practices are often encapsulated in institutional policies (for example, the Open University LA policy (Slade & Boroowa, 2014)). However, students may not readily engage with these policies and more work is needed to propose ways the institutional practices detailed in these policies can align with students' expectations.

Participants' responses demonstrated that they made trade-offs in an attempt to arrive at what was an acceptable use of student data for them. This further suggests a need for HEIs to support these differences as a single approach to institutional use of student data might not suit all students. While some students might want to engage in choosing the details about what data is used, others might prefer not to engage, finding the resulting effort goes a step too far.

The negative reactions from some participants once they are given further details about possible privacy risks and benefits accompanying institutional use of student data are noteworthy and can guide HEIs as they seek to be more transparent about the data they want to use for LA and how they plan to use it.

Study 3 makes two theoretical contributions linked to the guiding theoretical frameworks. With respect to the privacy calculus theory (Culnan & Armstrong, 1999; Dinev & Hart, 2006) and its application in the LA context is the observation that additional factors apart from the potential privacy risks and benefits of data use for LA may be influential in determining students' data use preferences. Perceived usefulness and concern about data use for LA are two of the variables

which can be examined in future work for their contribution to predicting students' data use preferences.

A second theoretical contribution with respect to contextual integrity (Nissenbaum, 2010) is the observation that this particular understanding of privacy might not clearly explain the dual nature of participants' data use preferences in that they were both accepting *and* reserved about institutional use of student data for LA. Contextual integrity seems to support a clear determination of whether students are supportive or hesitant about certain institutional data use practices. What seems ill-fitting, however, is a determination where participants seem to be both, as is the case in the Study 3 results.

6.4.2   Limitations

Study 3 had several limitations, some of which are because of using a survey experiment and recruiting participants using a crowdsourcing platform. It was observed that participants had high willingness to share data at the start of Study 3. As discussed in Section 3.4, their willingness to share data could have been influenced by the fact that they were recruited from a crowdsourcing website where they participate in user studies and are used to regularly sharing data about themselves.

Another limitation of Study 3 lies in its ecological validity. The data was not collected while students were taking part in an actual lecture or studying as they ordinarily would be. Consequently, the results obtained might not be reflected similarly in an actual course. Future research should seek opportunities to carry out further studies with students in realistic learning contexts.

A difference in students' trust in the learning institutions was observed in Study 3. However, this may have been due to the study design as the students who were recruited studied at different institutions. The impact of the students' institution on students' data use preferences was not the focus of this research and should also be examined in future work.

6.4.3   Links to Other Studies in This Thesis

The findings of Study 3 have shown that the privacy risks and benefits interventions did not appear to influence participants' data use and data sharing preferences. This finding led to Study 4 being carried out to attempt to further unpack what contributed to this finding. Study 4 is reported in the next chapter.

7 Study 4 Methods and Results

This chapter discusses the methods and results from the fourth and final study in this thesis. The aim of Study 4 was to provide in-depth insights to unpack Study 3's results, specifically to better understand the factors that motivated them to change or not change their data use preferences. Section 7.1 is a summary of Study 3 and an introduction to the research question for Study 4. Section 7.2 is a discussion of the method used to address the research question. The results are presented and discussed in Section 7.3 after which the implications of Study 4's findings and the limitations are discussed in Section 7.4.

7.1 Introduction

Study 4 was carried out as a follow-up to Study 3, which is reported in Chapter 6. Study 3 aimed to examine whether and how an awareness of the possible privacy risks and/or benefits of data use for LA would influence participants' preferences for institutional use of student data. In Study 3, 447 participants were randomly assigned to four groups: i) a control group and three experimental groups – ii) privacy risks, iii) benefits, and iv) privacy risks and benefits. Participants in the privacy risks group were shown a description of potential privacy risks resulting from the use of student data for LA (the potential privacy risks were adapted from Solove (2009)). Participants in the benefits group saw a description of the possible benefits, namely, predictive LA and recommendations of learning materials (the potential benefits were adapted from Herodotou, Rienties, Boroowa, Zdrahal and Hlosta (2019), and Sclater, Peasgood and Mullan (2016)). Finally, participants in the privacy risks and benefits group were shown both the privacy risks and benefits descriptions.

The quantitative results from Study 3 (discussed in Section 6.3.1) indicated that the interventions (privacy risks and/or benefits) did not have an influence on participants' data use preferences and their willingness to share data for LA. Furthermore, Study 3's qualitative results demonstrated the dual nature of participants' data use preferences as they both *supported* and were *reserved* about institutional use of student data for LA. Similar results were observed from the qualitative data regarding students' data sharing preferences, while additionally, a (mis)match was observed between institutional data use practices and students' data use expectations.

Thus, Study 4 was carried out to provide in-depth insights to try to explain and unpack Study 3's results and to obtain further insights into students' data use and data sharing preferences. In this regard, Study 4 builds on the strengths of qualitative research to try to explain quantitative results as well as unexpected results in the context of mixed methods research (Bryman, 2016). Study 4 focused on answering the following research question:

RQ 9: What insights can be obtained about the motivation for students changing or not changing their data use preferences?

The next section describes the method used to address this research question.

7.2 Method

This section describes the method used to carry out Study 4. First, Section 7.2.1 discusses the study setting and describes the participants recruited to take part in Study 4. This is followed, in Section 7.2.2, by a discussion of the procedure and the study instruments that were used. Finally, a description of the data analysis procedure is provided in Section 7.2.3.

7.2.1 Setting and Participants

One hundred and seventy-five (175) participants who took part in Study 3 also expressed an interest to participate in the follow-up semi-structured interviews. These participants' data use preferences were used to recruit those participants who had expressed different data use preferences. Participants' data use preferences were categorised as Low (L; prefer not to share data), Medium (M; prefer to share data about themselves *or* their activity on the online learning platform), and High (H; prefer to share data about themselves *and* their activity on the online learning platform). Whether or not these participants changed their data use preferences was also used to create categories to try to better understand the change/no change dichotomy. Thus, possible categories from which to recruit participants were where participants were:

- Willing to share their data and changed their data use preference (M/H, change)
- Not willing to share their data and changed their data use preference (L, change)
- Willing to share their data and did not change their data use preference (M/H, no change)
- Not willing to share their data and did not change their data use preference (L, no change)

Given that some participants changed their data use preferences while others did not, emphasis was placed on recruiting participants in the following three groups: i) those who did not change their data use preferences, ii) those who changed their data use preferences to share less data, and iii) those who changed their data use preferences to share more data. In total, 61 participants (34.9%) fit this criterion and were invited to participate in the follow-up semi-structured interviews. Of those invited, in total 15 participants (24.6%) were included in the actual interview. There were no apparent systematic differences in the characteristics of those who took part in Study 4 versus those who did not.

Participants were contacted using the messaging service on the Prolific platform and invited to take part in the follow-up semi-structured interviews. The initial contact email sent to participants is shown in Appendix D.1. All the participants were compensated £5 which was paid as a bonus on

the Prolific platform (that is, as an additional payment to the compensation they received for taking part in Study 3).

Data collection was carried out in two phases and data analysis was done concurrently, as recommended by Schutt (2017, p. 267). In this way, it was possible to determine the focus of further exploration in a subsequent data collection phase.

The first phase of Study 4 recruited 10 participants. Six of these participants had not changed their data use preferences and four had changed their data use preferences to share less data. Of those who did not change their data use preferences, three were from the risks group and three from the benefits group. Of those who changed their data use preferences to share less data, there was one participant each from the four groups (privacy risks, benefits, privacy risks and benefits, and control).

The second phase of Study 4 recruited an additional 5 participants. Four of these participants had not changed their data use preferences, and 1 changed their data use preference to share more data. Of those who did not change their data use preferences, three were from the risks group and one from the benefits group. The participant who changed their data use preference to share more data was from the benefits group. Therefore, participants from all four study groups were recruited. While there is representation from participants who changed and did not change their data use preference, there was only one participant who chose to share more data and only one participant who opted to share no data. The resulting limitations of not recruiting more participants from these two categories are discussed in Section 7.4.2.

There were four male and 11 female participants recruited for Study 4. Six participants were from the risks group, seven from the benefits group and 1 participant each from the privacy risks and benefits and control groups. Table 7.1 shows the demographic details of the participants who took part in Study 4.

Change/no change	Group	ParticipantID	M/F	Pre-test Prefer to share:	Post-test Prefer to share:	Willingness to share data	Course
		P_001	M	No data	No data	Low	Biomedical Science
		P_008	F	Self & Activities	Self & Activities	High	Drug Development
		P_010	M	Self	Self	Medium	Law
		P_012	M	Activities	Activities	Medium	Business Computing
	Risks	P_013	F	Self & Activities	Self & Activities	High	Criminology
		P_003	M	Activities	Activities	Medium	PGCE (Prior Medical Genetics)
		P_004	F	Activities	Activities	Medium	Applied Psychology (Clinical)
		P_005	F	Self & Activities	Self & Activities	High	International Policy
		P_011	F	Self	Self	Medium	Biomedical Engineering
No change	Benefits	P_015	F	Activities	Activities	Medium	Psychology
	Risks	P_006	F	Self & Activities	Activities	High to Medium	Psychology
	Benefits	P_002	F	Self & Activities	Self	High to Medium	English
Change (Share less data)	Risks & Benefits	P_007	F	Self & Activities	Activities	High to Medium	Computing and IT
	Control	P_009	F	Self & Activities	Activities	High to Medium	Chemistry
Change (Share more data)	Benefits	P_014	F	Self	Self & Activities	Medium to High	Classics

7.2.2   Procedure and Study Instruments

During the interview, participants were asked to explain the data use preferences they expressed in Study 3, and why they were (un)willing to share different data items. Following on from this, participants were given a scenario describing how student data can be used in LA and asked to identify what they thought were positive and negative aspects in the scenario. The benefits of scenarios and their suitability for the thesis research are discussed in Section 3.4. Scenarios were used in Study 4 to aid participants to think deeply and beyond what they would ordinarily focus on without some prompting. Thus, participants could consider different perspectives thereby providing richer insights (Blandford, Furniss, & Makri, 2016).

A sample interview schedule is provided in Appendix D.2. Each participant had their own interview script as their data use and data sharing preferences were different. The interview focused on understanding and identifying:

- The motivation(s) for each participant's pre- and post-test data use preferences
- The motivation(s) for each participant's data sharing preferences
- The factors that contributed to participants (not) changing their data use preferences
- The potential factors participants identified that would contribute to them (not) sharing their data for LA

Before or during the interview, participants were sent their responses to the Study 3 questions alongside the participant information sheet and consent form. This was due to the limited time the interviews were scheduled to run for (15 minutes). In this way, participants had time to review the material and familiarise themselves with their responses prior to the interview.

Effort was made to maintain participants' privacy. As a result, they were not asked to provide their email or Skype addresses (although some participants provided their email addresses during Study 3). Instead, they were sent a link to an online room created for the interviews[20]. Additionally, participants were free to choose whether to turn on their video cameras during the interviews.

7.2.3   Data Analysis

The interviews were audio-recorded and transcribed for analysis. At the end of each interview, brief notes were taken to summarise relevant issues participants raised that were related to the research question and the focus of the interviews. The NVivo software (version 11) was used in the data analysis process to keep track of the codes and resulting themes across participants'

---

[20] The tool used to conduct the interviews was www.talky.io.

responses. The analysis process was iterative (Saldaña, 2016) using first and second cycle analysis methods as detailed in Section 3.5.1. Details of the steps followed in the first and second cycle analysis process are provided in this section to enhance the trustworthiness of the results as recommended by Nowell, Norris, White and Moules (2017).

In the first cycle of analysis, the data was analysed comparatively, looking at participants' responses in relation to the research question, as well as whether they changed their data use preferences in Study 3. The output from this first cycle of analysis focused on whether participants' responses addressed i) the motivation for their data use preferences, ii) the motivation for their data sharing preferences, iii) the motivation for a change in data use preferences, iv) the motivation for no change in data use preferences, and v) factors participants thought would contribute to them (not) sharing their data for LA. The codes for those who opted to share no data were compared with those who chose to share some data (either data about themselves *or* data about their activities on the online learning platform), or all data (that is both data about themselves *and* data about their activities on the online learning platform).

As the codes were examined in the second cycle of analysis, several categories were identified. This included, for example, *context-based expectations* where participants' expectations were based on the use of data specifically in the learning context. Some responses were not aligned to any of these categories and were coded as *other,* while some were specifically in response to the *scenario*. Responses from the latter two categories were excluded from the results.

Similar to the qualitative data collected for Study 3 (discussed in Chapter 6), Study 4 data was analysed collaboratively. Two transcripts were coded independently by the main researcher and two members of the supervision team. The codes were reviewed for areas of agreement and disagreement before the remaining transcripts were analysed by the main researcher. The process of collaborative qualitative data analysis is discussed in detail in Section 3.5.1.4. A further review of the codes identified in the second cycle analysis round resulted in the themes discussed in the next section.

7.3   Results

A total of 249 codes were identified from the 15 participants who took part in the follow-up semi-structured interviews. This led to the following five themes: i) Data use practices are not aligned to students' expectations (38.6% of codes, n = 96), ii) Data use practices are aligned to students' expectations (26.9% of codes, n =67) iii) Opportunities for institutional learning about students' data use expectations (28.5% of codes, n =71), iv) Students are resigned to institutional data use practices (3.6% of codes, n =9), and v) Knowledge disparities contribute to students'

misperceptions (2.4% of codes, *n* =6). The themes are defined in Table 7.2, and the themes and corresponding codes are shown in Appendix D.3.

Table 7.2: Themes summarising insights from the motivation for participants' data use and data sharing preferences

Theme	Definition of theme: Statements …
Theme 1: Data use practices and students' expectations are not aligned (38.6% of codes, n = 96)	that demonstrate a mismatch between LA data use practices and students' expectations.
Theme 2: Data use practices and students' expectations are aligned (26.9% of codes, n = 67)	that demonstrate a match between LA data use practices and students' expectations.
Theme 3: Opportunities for institutional learning about students' data use expectations (28.5% of codes, n = 71)	that highlight HEIs' potential lack of knowledge about students' data use preferences.
Theme 4: Students are resigned to institutional data use practices (3.6% of codes, n = 9)	where students seemed resigned to the institutional data use practices emphasising their lack of agency to bring about any changes.
Theme 5: Knowledge disparities contribute to students' misperceptions (2.4% of codes, n = 6)	that surface students' lack of knowledge about how their data is used which results in students' misperceptions about institutional data use practices.

A thematic map of the themes and corresponding codes identified in Study 4 is shown in Figure 7.1. In presenting these results, participants are identified alongside a sample of their responses by providing their participant ID and indicating whether they changed their data use preference and the extent of their willingness to share data for LA. For example, (P_001, not willing to share data, no change) refers to the first participant who was unwilling to share data for LA and did not change their data use preference at the end of the study.
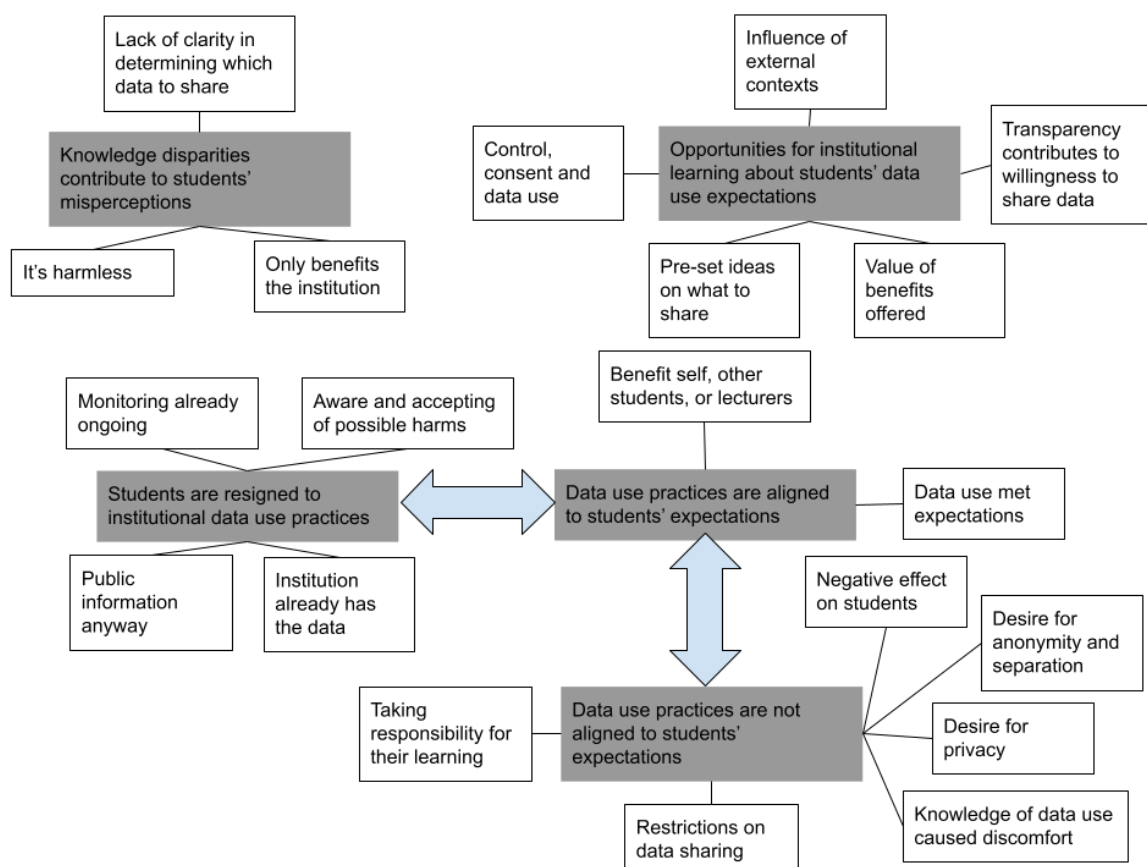
Figure 7.1: An illustration of Study 4's themes. The arrows represent tensions identified between the themes

7.3.1   Where Data Use Practices are not Aligned to Students' Expectations

The first theme identified showed that the data use practices described in the study were not aligned to students' data use expectations. Fourteen (14) of the participants, pointed to different ways they thought the use of student data as described in the study might have a **negative effect on students**. Only one student (P_013, self and activities, no change) had responses which did not correspond to this code.

Some of the negative effects that participants raised included *concerns about discrimination* (7 participants) which, for example, contributed to one participants' unwillingness to share data:

> *"I feel like those, those are the, the key things that could be used to discriminate against me if they want to."*
>
> *[P_001, not willing to share data, no change]*

Participants shared the perspective that the data collected did *not give the full picture* of the students' studying and learning practices (6 participants):

194

"… it can kind of reduce your grades to just numbers as opposed to, you might have not done very well in the past but now you're working harder and you know you're going to achieve better grades in the future, whereas the computer system might not see that. So I think you're kind of the best measure, at the end of the day of like how hard you can work and how well that you think you can do, but it's useful to see the data. But I think to take it like holistically, along with other factors like your own motivations and how hard you think you can work."

[P_014, willing to share data about self, change (share more data)]

Participants were also concerned about *data being used to withdraw opportunities* that they or other students would ordinarily have had access to (5 participants):

"I suppose if sharing my data would cause them to, I don't know like stop sharing some information. Like for example, they've been giving us lectures online, and if they think, you know, the student has not watched half of the lectures... and even right so let's say that was the scenario, and then they would say okay we're going to stop doing this for this particular student, or even for everybody because I don't know not everybody's using it, but I think, you know, in that case, I would be you know if I knew that that would happen, I would be uncomfortable sharing it. Because I would not want anybody to not get the information that they should be getting because somebody is not using it, you know. Because if there's one person who's going to use the information, then it's important information."

[P_006, willing to share data about self and activities, change (share less data)]

Other issues were highlighted by fewer than two participants. This included concerns about a *chilling effect* (2 participants) where they would not be comfortable to express themselves, for example, if forums were monitored, and they were made aware that the data was being analysed:

"And forum posts I feel like recording that information would feel people make people feel less comfortable to just express their opinion and like, ask questions."

[P_011, willing to share data about self, no change]

Other ethical considerations were identified as participants *did not want to act only on what was being measured,* which they thought would happen once it was clear what data the institution was collecting (2 participants). One participant was opposed to data being used to *limit available options* for students, and another participant thought that doing so *crossed a boundary.* One participant was concerned that *generalisations would disadvantage minority students*, while

another thought to act if they learned their data *was being used as a benchmark for other students.*

A desire for **anonymity and separation** was expressed by seven (7) participants. While one participant indicated a willingness to share data if it was anonymised, two (2) participants wanted to *separate their social life from their life as a student*, in particular referring to use of social media data, and one (1 participant) thought to *separate their physical from their online identity* such that their data was not be linked to their identity:

> *"There'd have to be a breakage between like my own identity and the online identity so the online would have to be completely separate to what I had, you know, like all of my like my name and date of birth that would all have to stay off it. And then, and like it would stay separate on the online."*

> *[P_001, not willing to share data, no change]*

Related to this was a **desire for privacy** expressed by seven (7) participants, with four (4) participants linking *privacy to their identity*, and one each *being aware of monitoring, uncomfortable with being monitored, and wanting to limit the data collected to maintain privacy.*

> *"So, I think for everyone's sort of confidentiality and privacy only the data that is necessary and required should be included."*

> *[P_015, willing to share data about activities, no change]*

Knowledge of data use caused discomfort for four (4) participants, as they considered the data uses described in the study and became aware of the data that was used and how it was used:

> *"And the more I thought about it, the more I thought it was kind of invasive to be having everything you're doing monitored by the uni, and I thought like I feel like you should be able to use the platform how you want to without that having an impact. And it just made me feel a bit uncomfortable."*

> *[P_002, willing to share data about self and activities, change (share less data)]*

Restrictions on data sharing were useful for four (4) participants who did not want their data shared with third parties. Finally, two (2) participants thought they should **take responsibility for their learning.** One of these students understood self-assessment tests to be for students' own learning and thus did not expect to share it with others or have others comment on their performance in such a case:

> *"It's self-assessment, and I feel like as long as I have an understanding of the situation and how far I'm along in the course, I don't believe that I need feedback on it, and therefore, I don't need a second opinion on this."*

> *[P_010, willing to share data about self, no change]*

On the other hand, the second of these students wanted to have the information available to them and for them to then decide how to proceed, rather than having already filtered options presented to them:

> *"So, if you do a survey and then, the results are that they'll show you this type of learning because you said that this as well too. You won't be able to see the other options whereas I think it would just be more helpful to see all the options and choose which is best for you."*

> *[P_011, willing to share data about self, no change]*

Both responses point to these students wanting to be in control of how (some) of their data is used and wanting all rather than filtered information presented to them.

7.3.2 Where Data Use Practices are Aligned to Students' Expectations

In contrast to the first theme, the second theme identified that the **data use practices described in the study were aligned with students' expectations** (14 participants). The observation that fourteen out of fifteen participants' responses were related to this theme highlights a duality similar to that observed in Study 3, that data use practices can be both aligned and misaligned for the same participant. One motivation was the *benefits for students themselves, other students, or their lecturers* (11 participants). Contributing to this was participants' *desire for an individualised experience* that was *tailored to them*, with one participant stating:

> *"I think that I would definitely want a more personalised like sort of feedback or support… that's more tailored to sort of my learning and my ability, would be sort of really encouraging and would be motivational for me because at least I know that, since it's tailored to me that it is something that I can do, and it would definitely benefit me in some way. As opposed to just the general sort of general support for everyone."*

> *[P_015, willing to share data about activities, no change]*

Participants also made several statements that indicated that the *data use met their expectations* (9 participants), for example:

"Some data has to be used. I think it's good that some data is used. I think it's beneficial for both student and university to have some data openly used. And, yeah, I think there's not anything that I would absolutely cut off completely. Mainly just because I think the majority of the data that is used is used for beneficial purposes and it's not used for any malicious intent."

[P_007, willing to share data about self and activities, change (share less data)]

Given the relationship between the first two themes, participants' responses were compared as shown in Table 7.3.

Table 7.3: Comparing codes for Theme 1 and 2

	Change/No Change	Willingness to share data	Data use practices are … to students' expectations							
			not aligned						aligned	
			Code 1	Code 2	Code 3	Code 4	Code 5	Code 6	Code 1	Code 2
P_001	No change	Low	X	X	X		X			
P_002	Change (Less)	High to Medium	X		X	X			X	
P_003	No change	Medium	X							X
P_004	No change	Medium	X	X	X		X		X	
P_005	No change	High	X						X	X
P_006	Change (Less)	High to Medium	X	X	X	X				X
P_007	Change (Less)	High to Medium	X			X			X	X
P_008	No change	High	X							X
P_009	Change (Less)	High to Medium	X			X			X	
P_010	No change	Medium	X		X		X		X	X
P_011	No change	Medium	X	X				X	X	X
P_012	No change	Medium	X	X				X	X	X
P_013	No change	High		X	X				X	X
P_014	Change (More data)	Medium to High	X	X			X		X	
P_015	No change	Medium	X		X				X	

Notes: Not aligned – code 1 = negative effect on student; code 2 = desire for anonymity and separation; code 3 = desire for privacy; code 4 = knowledge of data use caused discomfort; code 5 = restrictions on data sharing; code 6 = taking responsibility for their learning; Aligned – code 1 = benefit self, other students, and lecturers; code 2 = data use met expectations

As can be seen in Table 7.3, only one participant's (P_001) responses did not contain codes related to the second theme. As the participants' responses contained codes that demonstrated the data use practices were in some cases aligned and in others not aligned to students' expectations, this further highlights the dual nature of students' data use preferences identified in the results from Study 3 (discussed in Section 6.3.2).

7.3.3 Opportunities for Institutional Learning about Students' Data Use Expectations

The third theme pointed to the potential lack of knowledge HEIs have about how students expect their data to be used for LA, and therefore highlighted **opportunities for institutional learning** (14 participants). The results in this theme were identified as they emphasised that there is at present little opportunity for students to provide HEIs with positive or negative feedback, or to input into proposed and actual uses of student data at the institutional level. As a result, institutional data practices are seen in some cases to be at odds with students' data use expectations. These results identify possible misperceptions that HEIs might have due to the absence of this feedback provided in ways that can feed into institutional data use policies and practices. Participants' responses across this theme are illustrated in Table 7.4.

Table 7.4: Comparing codes for Theme 3

	Change/No Change	Willingness to share data	Opportunities for institutional learning about students' data use expectations				
			Code 1	Code 2	Code 3	Code 4	Code 5
P_001	No change	Low	X	X		X	
P_002	Change (Less)	High to Medium	X				
P_003	No change	Medium	X				
P_004	No change	Medium	X				
P_005	No change	High	X				
P_006	Change (Less)	High to Medium	X				
P_007	Change (Less)	High to Medium	X				
P_008	No change	High					X
P_009	Change (Less)	High to Medium	X		X		
P_010	No change	Medium	X	X			X
P_011	No change	Medium	X			X	
P_012	No change	Medium		X			
P_013	No change	High					
P_014	Change (More data)	Medium to High			X		
P_015	No change	Medium	X				

Notes: Code 1 = control, consent, and data use; Code 2 = Influence of external contexts; Code 3 = Transparency contributes to willingness to share data; Code 4 = Value of benefits offered; Code 5 – Pre-set ideas on what to share.

Issues related to c*ontrol, consent and data use* were highlighted by 11 participants. Insights included making it possible for students to update their data use preferences when they determined that there was a need to provide data that they would previously have withheld:

> *"… if you had a medical condition that meant that you had to go to the hospital once a week or once a month. And for that once a month, you weren't able to do any, any work or any coursework or anything, then that could become a reason why the medical information would be relevant."*
>
> [P_007, willing to share data about self and activities, change (share less data)]

Additionally, the *influence of data use in external contexts* on participants' data use preferences for LA became clear from three (3) participants' responses. One participant referred to events reported in the media and linked to social media and elections and how this then had influenced his data use preferences in the LA context:

> *"I think part of it is because of the historic abuse of data, especially online. Like Cambridge Analytica, the sort of worrying use of data by Facebook, and it just sort of turns me off using data for anyone because I sort of feel like if they can abuse that anyone can."*
>
> [P_001, not willing to share data, no change]

Additionally, one participant referred to how an awareness of the *permanence of digital footprints* caused them to change their forum posting behaviour:

> *"Records of my forum posts, obviously these posts on the learning platform so I'm not gonna put anything out on the internet that don't want to be there forever"*
>
> [P_010, willing to share data about self, no change]

It may be the case that this response was driven by the information no longer being under the student's control.

Transparency by the HEI was noted as *contributing to students' willingness to share data* (2 participants):

> *"So, as long as I'm aware what my data is being used for and where it's going. Then I'm kind of happy to share, share most [of] my data."*
>
> [P_014, willing to share data about self, change (share more data)]

Additionally, two participants assessed the *value of the benefits offered* and thought that they needed to be novel and thus 'worth' students' data, rather than benefits that they thought were readily accessible:

> *"I feel like some of them are like extras they're not necessary. So I feel like learning book recommendations, like it's useful but it's not necessary and enough that I would want to give up all of my data. I feel like I can get that off other people and other sources as well, so I feel like it'd be more useful if, like I'd be willing to give up my data for something that was unique to that area and was very undeniably useful. Like if it was suggesting [to] me future courses and stuff like that, then maybe I would look at giving some of my data."*
>
> *[P_001, not willing to share data, no change]*

Finally, two participants made clear that they had *pre-set ideas on the data they wanted to share*, and as demonstrated in Study 3 (reported in Section 6.3), *it might be difficult to change these ideas:*

> *"Well, when I went into the study, I was already aware of what I was willing to share and what I wasn't willing to share. I was very steadfast, and I knew my decision, unless there was something major in the study there was nothing that was really ever going to change why I thought I was willing to share what I wasn't willing to share. So, I was pretty secure in my choice at the beginning, and therefore didn't change."*
>
> *[P_010, willing to share data about self, no change]*

7.3.4   Knowledge Disparities Contribute to Students' Misperceptions

With the fourth theme, several **misperceptions that students held** (5 participants) with respect to the use of student data for LA were identified. These misconceptions can be attributed to a knowledge gap that HEIs can address by communicating clearly to students about the use of student data for LA. Participants' responses are illustrated in Table 7.5

Table 7.5: Comparing codes for Theme 4

	Change/No Change	Willingness to share data	Knowledge disparities contributed to students' misperceptions		
			Code 1	Code 2	Code 3
P_001	No change	Low			
P_002	Change (Less)	High to Medium			
P_003	No change	Medium	X		
P_004	No change	Medium			
P_005	No change	High			
P_006	Change (Less)	High to Medium			
P_007	Change (Less)	High to Medium	X		
P_008	No change	High			
P_009	Change (Less)	High to Medium			X
P_010	No change	Medium	X		
P_011	No change	Medium			
P_012	No change	Medium			
P_013	No change	High			
P_014	Change (More data)	Medium to High		X	
P_015	No change	Medium			

Note: Code 1 = It's harmless; Code 2 = Not only benefiting the institution; Code 3 = Lack of clarity in determining the data to share.

The data use was seen as *harmless* by three (3) participants:

> *"It was because I didn't see how they would use that information in any way that could be bad for me..."*
>
> *[P_003, willing to share data about activities, no change]*

Finally, there was a *lack of clarity in determining the data to share* for LA (1 participant) in choosing between sharing data about oneself, data about one's activity on the online learning platform or both. This presents an opportunity for HEIs to communicate to improve students' understanding of data use for LA.

7.3.5 Where Students are Resigned to Institutional Data Use Practices

With the fifth theme, five (5) participants were found to be resigned to HEIs using student data for LA. This perspective highlighted the students' lack of agency to bring about any changes. Participants' responses are illustrated in Table 7.6.

Table 7.6: Comparing codes for Theme 5

	Change/No Change	Willingness to share data	Students are resigned to institutional data practices			
			Code 1	Code 2	Code 3	Code 4
P_001	No change	Low				
P_002	Change (Less)	High to Medium				
P_003	No change	Medium				X
P_004	No change	Medium				
P_005	No change	High				
P_006	Change (Less)	High to Medium				
P_007	Change (Less)	High to Medium	X			
P_008	No change	High				
P_009	Change (Less)	High to Medium	X			
P_010	No change	Medium	X			
P_011	No change	Medium				
P_012	No change	Medium	X	X	X	
P_013	No change	High				
P_014	Change (More data)	Medium to High				
P_015	No change	Medium				

Note: Code 1 = Public information anyway; Code 2 = Institution already has the data; Code 3 = Monitoring already ongoing; Code 4 = Aware and accepting of possible

harms

The fact that the *information was already public* was a motivator for four (4) participants who held this view. For one participant, the *institution already had the data,* suggesting that there was no need to hold any data back:

> *"I wouldn't mind that being taken. Because that's information for the learning institution anyway."*
>
> *[P_012, willing to share data about activities, no change]*

There was also an awareness that *monitoring was already ongoing* (1 participant), and one student *accepted the possible harms:*

> *"Nine times out of 10, sharing loads of data isn't going to be bad for you and isn't going to be a problem. There may be instances where something goes wrong, there's a data breach, or you share the wrong information to the wrong person. And it could cause some level of stress in your life where you might have to go cancelling credit cards or something like that. But I don't see that as a bad thing to do, to share too much data. So I have less of a negative view on that."*
>
> *[P_003, willing to share data about activities, no change]*

The next section is a discussion of these study findings focusing on their implications and acknowledging the limitations of the study.

7.4   Discussion

7.4.1   Implications of Findings

Study 4 has identified various privacy and utility tensions with respect to students' data use preferences in the LA context. On one hand, the data use practices described in the study met participants' expectations, yet at the same time, participants made statements to suggest that the data use practices were not aligned to their expectations. For example, there were numerous statements to suggest that the data use described in the study met students' expectations, however, one student wanted to remain responsible for their learning while others wanted restrictions on who the data was shared with and wanted to keep their social life separate from their lives as students. Highlighting these privacy and utility tensions serves to demonstrate Study 4's contribution to understanding the nuances that are present as students respond to knowledge about the use of their data for LA.

Similar to the findings of Study 3 which first identified the dual nature of students' data use preferences (discussed in Section 6.3.2), Study 4's findings also demonstrate that institutional data use practices as examined in the study are in some cases aligned and in other cases not

aligned to students' data use expectations. An additional privacy and utility tension was identified between being transparent with students and informing them about the data that is used for LA and how, and a resulting desire for privacy. While HEIs that inform their students about the use of their data for LA are meeting their ethical obligations (Sclater, 2016), Study 4's results suggest that a balanced approach is necessary so as not to cause students' undue discomfort and concern. A recommendation based on this finding is for HEIs to be transparent with students about the data that is used for LA and the resulting privacy risks and benefits accompanying these uses. Furthermore, HEIs should provide information on steps they are taking to mitigate any of the possible privacy risks facing students from the use of their data. Institutional transparency initiatives should also concentrate on addressing students' misperceptions. Knowing that this is being done might address concerns about data use negatively impacting various student groups. Future work can explore transparency initiatives which also seek to develop students' self-efficacy to address ethics and privacy concerns.

Study 4 further emphasises that no one size fits all with respect to the use of student data for LA. Students appreciate the benefits from the use of their data, but also have contrasting expectations, such as remaining responsible for their learning, or expressing concern about possible negative effects such as causing students to only act on what they know is being measured. Students also highlighted that the data captured did not give a complete picture of the activities they carry out for their learning.

The ethical issues raised by students in Study 4, where they highlight potential negative effects to students, mirror issues raised in the ethics and privacy framework literature. For example, Slade and Prinsloo's (2013) framework on ethical issues and dilemmas argues that students' performance is dynamic with the information taken at one point in time and subject to change.

Study 4 has further identified the need for HEIs to allow students to provide feedback on data use initiatives. Additionally, there might be specific data items that students prefer not to share but then choose to share at a later time. Therefore, HEIs should allow students to update their data use preferences. This proposal differs from situations where students at present can choose to withdraw their data from use or add personal details that are relevant. Instead, it points to capturing questions which may arise about why certain data is needed, or how students determine what data to share or when to offer or withdraw their consent for data use. Questions that such an approach would address include: i) Do HEIs know if data use in contexts such as social media influence students' expectations in LA? and ii) Are these external influences positive or negative for LA? This approach would need to work in either an opt-out or an opt-in data

regime. Such a flexible approach allowing students to update their data use preferences can potentially be valuable to both students and HEIs.

Taken together, Study 4's findings demonstrate that there is a need for HEIs to develop transparency initiatives as well as control and feedback mechanisms for students regarding use of student data for LA. In this way, students will be clearly informed about the use of their data and be given the opportunity to control which data items are used. Their selections will act as feedback for the HEIs to identify what data students are willing to share. One challenge that has to be acknowledged with providing students with the opportunity to control the data that is used for LA is that it requires their effort and input, and some students may not be willing to expend the necessary effort. Allowing students to update their data use preferences as discussed is a learning opportunity for HEIs. This information can then feed into the data use policies through institutional review processes. This proposal is discussed in further detail in the final chapter of this thesis (Chapter 8).

7.4.2   Limitations

While recruiting 15 participants for Study 4 met the requirements for an appropriate number of participants for qualitative research (Guest, Bunce, & Johnson, 2006), there was only one participant who changed their data use preference to share more data and only one who did not want to share data. As such, the insights obtained are focused more on participants with a medium willingness to share data which they maintained during the study, or those who opted to share less data during the study. However, the insights obtained shed further light on students' data use and data sharing preference as discussed in the results.

The next chapter is a general discussion that brings together the findings and contributions of the four studies reported in this thesis research, identifies the implications, and makes recommendations for future work.

8 General Discussion and Conclusions

This final thesis chapter revisits the motivation for the thesis research and the research questions in Section 8.1. Key findings from the four thesis studies are summarised and the thesis contributions are detailed in Section 8.2. The thesis research findings are relevant for various stakeholders in LA including policy makers, researchers, and administrators in HEIs, thus, Section 8.3 contains relevant practical implications of the research findings and recommendations for efforts to address privacy and utility tensions in LA applications. Further research directions are also discussed. Finally, concluding remarks are made in Section 8.4.

8.1 Introduction

LA is being used in HEIs to bring about potentially worthwhile changes in educational practice. These changes include using data from past courses to inform the effective design of courses (Rienties & Toetenel, 2016), identifying students in need of further learning support (Jayaprakash, Moody, Lauría, Regan, & Baron, 2014), and recommending additional learning materials for students (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015). At the same time, there are reports of privacy and utility tensions in LA. Examples of these tensions include interventions based on LA to support student success (Herodotou, Naydenova, Boroowa, Gilmour, & Rienties, 2020; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014), while potentially limiting their responsibility for their own learning (Rubel & Jones, 2016), and raising students' awareness of LA (Roberts, Howell, Seaman, & Gibson, 2016; Adejo & Connolly, 2017; Brooker, Corrin, Fisher, & Mirriahi, 2017), while potentially causing them to be concerned about the negative impact of LA feedback, such as, on their motivation (Roberts, Howell, Seaman, & Gibson, 2016). It is important that the causes of these privacy and utility tensions are identified and where possible addressed to allow further innovation in LA (Gasevic, Dawson, & Jovanovic, 2016).

Against this background, the thesis research aim was to investigate the causes of privacy and utility tensions in LA from the perspective of two stakeholder groups: LA experts and students. It was important to characterise what contributes to privacy and utility tensions in LA to identify relevant insights which can be used to develop ethical LA applications. The perspectives of LA experts were sought because of their contribution to ethics and privacy frameworks (for example (Slade & Prinsloo, 2013; Pardo & Siemens, 2014; Heath, 2014; Hoel & Chen, 2018)), which are critical for the development and implementation of LA in higher education. Additionally, students' perspectives were deemed important to explore and understand because taking them into consideration might help maintain students' trust in HEIs and LA, and students' perspectives can also contribute to the design and development of ethical LA applications (Braunack-Mayer, Street, Tooher, Feng, & Scharling-Gamba, 2020). Furthermore, there have been calls to integrate more of

students' perspectives on privacy in LA research as their perspectives and preferences have yet to be fully considered and integrated (Roberts, Howell, Seaman, & Gibson, 2016). There is now more research focusing on students and privacy in LA as summarised in the literature review in Section 2.4.

There were three organising concepts chosen for the thesis research, namely privacy concepts, privacy concern, and data use preferences. Thus, the thesis research, in Study 1 sought to understand how LA experts conceptualised privacy in LA and what they thought the key privacy issues were. Due to mixed results reported in the empirical research literature as to whether students are concerned about privacy and the use of their data for LA, there was a need for further research leading to greater insight into the dimensions of students' privacy concerns or lack thereof. This formed the focus of Study 2. Finally, there was a lack of clarity about how students responded to the use of their data once they were made aware of the potential privacy risks and benefits of LA, and this formed the focus of Study 3 and Study 4. The research questions addressed in these four studies are shown in Table 8.1 along with the thesis chapters where the studies are reported and discussed.

Table 8.1: Thesis studies and research questions mapped to study, thesis chapter, organising concept, and stakeholder

Study	Chapter	Research Questions	Organising Concept	Stakeholder
Study 1	Chapter 4	**RQ 1**: How do learning analytics experts describe and talk about privacy in learning analytics? **RQ 2**: What do the experts consider to be key privacy issues in learning analytics?	Privacy concept	Learning analytics experts
Study 2	Chapter 5	**RQ 3:** To what extent are students concerned about the collection, use, and sharing of their data for learning analytics, and compared to e-commerce? **RQ 4:** To what extent are students' general privacy concerns and behaviour related to their concern about the collection, use, and sharing of student data for learning analytics? **RQ 5:** What issues contribute to students' concern or lack of concern about data collection, use, and sharing in learning analytics?	Privacy concept, Privacy concern	Students
Study 3	Chapter 6	**RQ 6:** To what extent does an awareness of the possible privacy risks and benefits of data use for learning analytics contribute to a difference in students' preferences for the use of their data? **RQ 7:** To what extent does not having access to learning analytics applications influence students' preferences for the use of their data? **RQ 8:** What do students indicate as the motivation for their preferences for use of their data?	Data use preferences	
Study 4	Chapter 7	**RQ 9:** What insights can be obtained about the motivation for students changing or not changing their data use preferences?		

The next section summarises and discusses the contribution of the thesis research based on these research questions.

8.2 Contributions of the Thesis

This thesis research has made both knowledge contributions and methodological contributions in the context of understanding privacy and utility tensions in LA. These contributions are elaborated upon in this section.

8.2.1 Contributions to Knowledge

The contributions this thesis research has made to knowledge on privacy and utility tensions in LA are summarised in this section corresponding to the three organising concepts introduced in Section 2.3 – privacy concepts, privacy concern, and data use preferences.

Privacy Concepts

Privacy concepts were investigated from the perspective of two stakeholder groups: LA experts (Study 1 in Chapter 4) and students (Study 2 in Chapter 5). The LA experts were found to conceptualise privacy in different ways from each other. Related to this, only 4 of the twelve LA experts shared a privacy concept (specifically, they thought of privacy as contextual integrity (Nissenbaum, 2010)). As discussed in Section 2.3.1.1, this means that their conceptualisation of privacy would be influenced by where the information was collected from, the type of information, who had access to it, and what it would be used for. The remaining experts held different privacy concepts from each other. Furthermore, the experts were found to hold more than one privacy concept – they were not exclusive. In addition, the experts described the privacy concepts using existing terms from well-known privacy literature. This included contextual integrity (Nissenbaum, 2004; 2011), privacy as a right (UN General Assembly, 1948), privacy as control over personal information (Westin, 1967), and privacy as identity (Floridi, 2005).

From the students' perspective, each individual student who took part in the follow-up interviews to Study 2 conceptualised privacy in their own unique way. Additionally, the students expressed different privacy concepts from those the LA experts expressed in Study 1. Furthermore, in discussing their privacy concepts the students expressed their expectations of privacy descriptively rather than using terms from privacy literature. In expressing their privacy concepts, one student referred to data not being shared with third parties and the need for students to provide informed consent, so students are aware about how their data is used. Another student referred to being free to act as they chose and viewed their data as their identity. A third student referred to the choice students might have whether to share data, while the fourth student

referred to the use of personal information and how this would impact her in the future. Thus, these privacy concepts might differ from LA implementations in practice.

Consequently, with respect to privacy concepts, the thesis research demonstrated that there are tensions as a result of the different ways LA experts and students conceptualise privacy, and therefore there are gaps between these concepts and how privacy is actualised in the field. It is noteworthy that there are differing views among the experts about what privacy in LA is. Furthermore, the lack of agreement between students' and the experts' privacy concepts suggests an additional factor contributing to the privacy and utility tensions in LA. This is because attempts to achieve privacy in LA may not match to students' expectations.

Privacy Concern

A key finding with respect to students' privacy concern in LA was that students were significantly more comfortable with the collection and use of their data in the university context than in the Amazon (e-commerce) context. This finding further supports the importance of context in privacy and LA from students' perspective. Students might want to be informed about who has access to their data, how it is or will be used, for how long, who will potentially benefit, and how, among other details. Students' privacy concerns were seen to lie in the university sharing student data with third parties. Finally, students in Study 2 were comfortable sharing their data with their tutors to receive better support from them. This lends further support to the consideration that a lack of institutional transparency towards students could contribute to increasing students' privacy concern.

From the follow-up interviews it was observed that being a student at the university, and therefore, having a relationship with the university might contribute to students' trust in the university using their data for LA. At the same time, there was mistrust whether the university would use student data as stated. One reason given for this mistrust was that students were unaware about how the data would be used but this might be because they did not engage with university data use policies (as seen, for example, in McDonald and Cranor (2008)). Students also expressed an interest in knowing how their data would be used and being able to control who had access to their data. Finally, the potential benefits of LA for students (that is, the features made available using LA) might be an important factor in participants' agreeing to provide their data in exchange for a service, in line with the privacy calculus theory (Dinev & Hart, 2006).

Thus, with respect to privacy concern, the findings from Study 2 suggest that contributing to privacy and utility tensions is the case of HEIs sharing data with third parties. Further contributing

factors include students' lack of engagement with the university (LA) data use policies and lack of opportunity to control the use of their data as with privacy self-management (Solove, 2013).

Data Use Preferences

An awareness of the potential privacy risks and benefits of LA did not influence participants' data use preferences in Study 3. Thus, students' data use preferences might be pre-determined and difficult to change. That said, it is important to emphasise the need for HEIs to (continue to) be fully transparent with students about the implications of LA. The qualitative data from Study 3 made clear that the potential benefits of LA motivated some students to share more data, while the potential privacy risks made others uncomfortable with data use for LA.

Participants were more willing to share learning and academic history data, such as information about their final school grade, followed by their digital exhaust data, for example, library loan statistics. They were less willing to share personal data, such as medical information. Students' willingness to engage with the data that is used does suggest an interest in controlling the data that is used for LA.

Finally, the potential loss of the LA benefits did not influence participants' data use preferences which lends further support to the finding indicated above that students' data use preferences are pre-determined and difficult to change.

In-depth insights regarding students' informational norms (that is, their expectations regarding the flow of personal data) were derived from the qualitative data. These are summarised below:

- A tension was observed as participants both supported and yet had reservations about institutional use of student data. This finding suggests that there are nuances in students' data use preferences which need further investigation and integration in practice to contribute to ethical LA.
- Participants accepted that some data needed to be shared for LA, but at the same time, they were concerned and had questions about what impact sharing data would have on them.
- In some cases, the data sharing practices met students' expectations, and at other times they did not. Thus, there were (mis)matches between students' data use expectations and the actual data use and data sharing practices for LA.
- The potential benefits of LA motivated some students to share more data. At the same time, they expressed conditions to be met, such as limiting what the data was used for, and concerns such as whether the data use would have a negative impact on students. Those participants who chose *not* to share more data did not perceive the potential benefits to

be worth more of their data and felt that they had reached the limit of the data they were willing to share. Thus, it would seem that not all potential benefits are perceived equally by students. Furthermore, there may be limits to the data students want used for LA due to their thinking that not all the data used is necessary.

- Trust in the HEI seemed to play a minor role as it was only suggested by three participants in Study 3, suggesting that other factors apart from trust in the HEI contribute more to students' comfort with the use of their data for LA. The apparent minor role of students' institutional trust in their data use preferences contrasts with the findings of Slade, Prinsloo, and Khalil (2019) who emphasise the role of students' trust in the HEI in their findings.

Study 4 provided further insights into students' data use preferences.

- Similar to the findings of Study 3, the data use practices described in the study were aligned and at other times misaligned with students' data use expectations.
- There are gaps in students' understanding about LA and these gaps seem to contribute to students' making incorrect assumptions about it.
- There are opportunities for HEIs to learn about students' privacy and data use preferences.

Thus, with respect to students' data use preferences, the findings from Study 3 and 4 suggest that what contributes to privacy and utility tensions in LA is a lack of effective transparency by HEIs towards students regarding institutional uses of student data, legitimate concerns about how institutional use of student data impacts students, and a lack of opportunities for HEIs to derive and act on students' privacy and data use preferences, for instance, integrating them into university (LA) data use policies.

8.2.2   Methodological Contributions

From a methodological perspective, the thesis research applied crowdsourcing and a survey experiment to understand privacy and utility tensions in LA. To the best of our knowledge, and as summarised in Section 2.4, this is the first time these research methods have been used in the context of privacy and LA research. Crowdsourcing allowed students from a variety of Further and HEIs to be recruited for the study. The survey experiment allowed an investigation of the impact of the potential privacy risks and benefits on students' data use preferences. The design of Study 3 did not investigate having the learning institution as an independent variable to determine whether and how this would influence privacy in LA. However, this can be considered in future research.

A further contribution of Study 3 was in the use of a modular prototype LA dashboard. Consequently, it was possible to mimic the influence of various data use preferences on the potential benefits and investigate whether this influenced students' data use preferences. Specifically, the prototype allowed an investigation of how students respond when it is made clear to them that the data which they provide for LA influences the features they can access.

Furthermore, this thesis research used a mixed-methods approach integrating follow-up interviews in Study 2 (reported in Chapter 5) and Study 3 (reported as Study 4 in Chapter 7) to provide in-depth insights to answer the 'why' question driving the quantitative data. As argued in Chapter 3, the qualitative data helped to complement the quantitative data and enhance the understanding of the research problem.

8.3   Practical Implications, Recommendations, and Directions for Future Research

With respect to privacy concepts, the thesis research has demonstrated that there are different concepts of privacy between and among members of the two stakeholder groups – LA experts and students. This work has documented these privacy concepts. These findings suggest the value of viewing privacy in LA not as a unitary concept, but rather as a cluster of concepts, acknowledging the complexity of privacy in this domain. This would be similar to the approach taken in the taxonomy of privacy harms (Solove, 2009). The benefit of taking such an approach is that it enables an understanding of the various ways privacy is thought about by various stakeholders in LA. With these perspectives available, deeper insights can be sought in future research, for instance, examining how these concepts relate to each other and which perspectives stakeholders might prefer. There is already recent work in this direction that has sought to identify the privacy concepts faculty members relate to the most (Jones, VanScoy, Bright, & Harding, 2021).

The findings also suggest that there is a strong relationship between ethics and privacy in LA. While the thesis research set out to investigate the privacy and utility tensions in LA, accompanying ethical considerations from both the LA experts and the students were identified. Thus, for the stakeholders who took part in the thesis research, ethics and privacy issues are intertwined. As they thought about the use of student data, they were also observed to consider, for example, its impact on stakeholders, especially students. While students provide the data, the output from LA applications can have a positive or negative influence on their learning experience. Thus, in thinking about privacy in LA, this work suggests that *there is support for a cluster concept of privacy, and there is a strong link between ethical and privacy issues from stakeholders' perspectives.*

Acknowledging the diverse views on privacy, HEIs and practitioners (including entities involved in the development of LA applications) can:

- Ensure that students are aware of the different data that is collected and the different ways it can be used at present and in the future (referred to as downstream data use (Cormack, 2016),
- Integrate ways to measure and improve student understanding of the use of their data,
- Give students an opportunity to opt-in to the use of their data for LA over the course of their stay at the institution (and once they have left), which is noted to align with legal requirements (GDPR) where user consent is needed,
- Make students aware of the possible benefits of LA which may improve informed consent and acceptance of LA,
- Make students aware of the possible privacy risks, how the institution is mitigating these risks, and what it would mean for students if any of these risks materialised while they are students, or in the future after they have completed their studies. The findings of Study 3 suggest that students are accepting of data use for LA and making them aware of possible privacy risks may not influence their strongly held privacy and data use preferences.

Another recommendation for HEIs and practitioners is that they seek to apply a cluster concept of privacy in LA to meet the needs and expectations of various stakeholders, especially those of students. Such an approach would be in line with contextual integrity (Nissenbaum, 2010) to the extent that data use meets the expectations of students and other stakeholders. Where stakeholders' expectations of data use cannot be met, then an explanation for any differences should be provided to ensure clarity for all involved. As an example, the thesis research has shown that some students might expect to have control over whether and how their data is used for LA. However, it may not be possible to provide this control for various reasons, such as meeting legal requirements or supporting robust statistical models. Given valid constraints, the expectations of student control over data used should be met to the extent that is possible and clear and accessible explanations about any limitations provided for the benefit of all stakeholders.

Catering to students' concerns about data sharing suggests that HEIs examine the data sharing that is ongoing and determine whether it is necessary or conflicts with students' expectations. There are various legitimate reasons why HEIs share data with third parties. The concern expressed by students may not mean that data sharing with third parties should cease, rather it points to the need for greater transparency on the purposes, outcomes, and implications for students.

Students in the studies reported in this thesis had different expectations about what data would be used for LA from the data that was used in practice by HEIs. Thus, there is an added need to allow students' data use preferences to be fed back to the HEI and, in particular, into the larger process of designing and developing LA applications. One way to do this might be through focusing on what can be perceived as a sub-process – the development of data use policies - as illustrated in Figure 8.1.
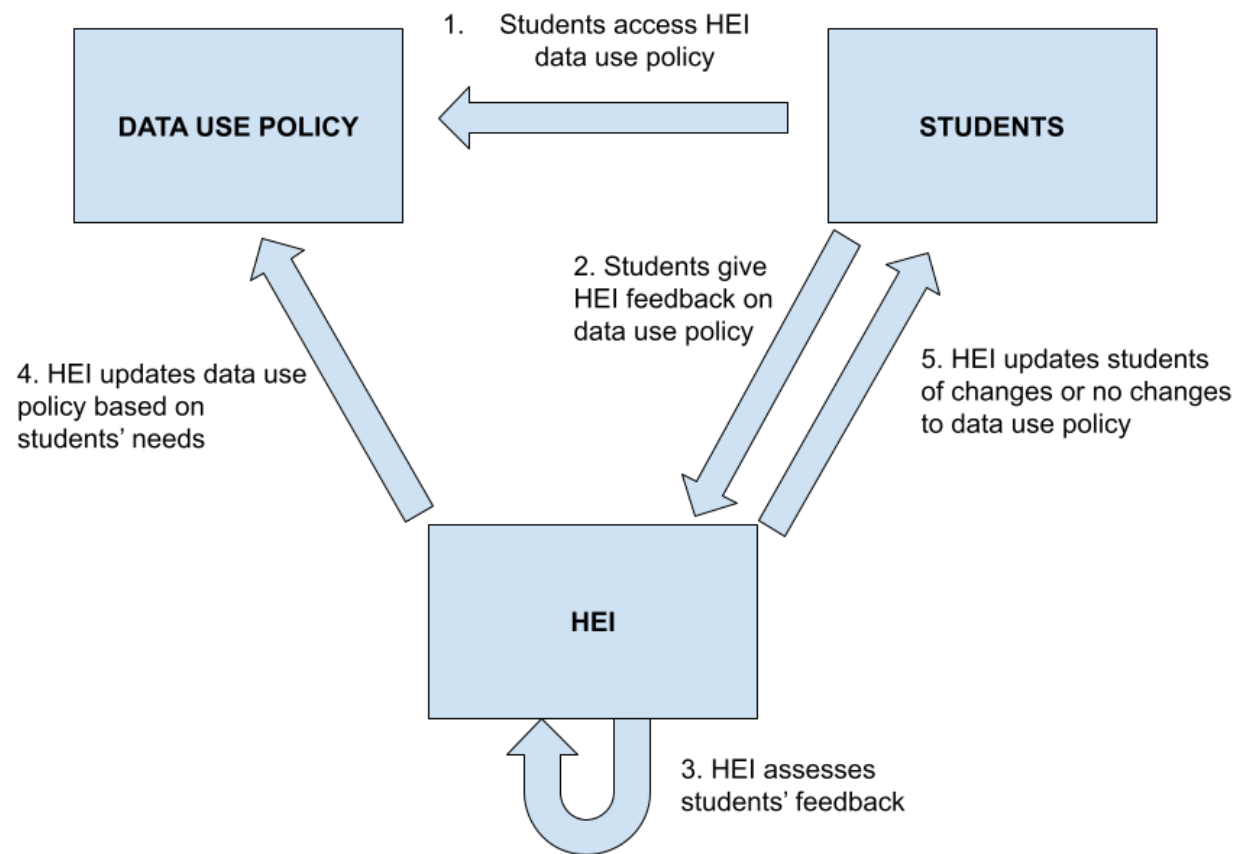
Figure 8.1: Proposed process of institutional learning from students' data use preferences

Figure 8.1 suggests that students access the data use policy from their learning institution [1] and are then provided a way to give feedback on the data use policy, for example in cases where the data use proposed fails to meet students' expectations [2]. Administrative staff from the HEI (possibly with technical support to alleviate workload concerns) review students' feedback [3] to determine those which can be acted on and those which cannot, for various reasons. The feedback from students which can be implemented is integrated into the data use policy [4] while those preferences which can and cannot be integrated, and the reasons for this, are communicated back to the students [5]. This proposal is relevant as it enhances students' data literacy (Knox, 2017) around how their data is used and provides a way for their data use preferences to be fed back to the HEI and for them to obtain feedback on the process. It raises questions about how HEIs can support students to engage with data use policies. Possible solutions include making it a part of their learning as seen in Knox (2017). Future work should investigate the feasibility of such a process, including determining its suitability from various stakeholders' perspectives, particularly those it would directly impact such as teaching and administrative staff at HEIs, as well as students.

8.4   Concluding Remarks

The thesis research has identified several factors contributing to privacy and utility tensions in LA. The two stakeholder groups (LA experts and students) think about privacy in very different ways from each other. More significant is that students' way of thinking about privacy does not match how privacy in LA is conceptualised, for example, in the ethics and privacy frameworks which guide the implementation of LA applications. Students' concern about their data being shared with third parties needs to be addressed by HEIs. Furthermore, it is recommended that HEI's communication with students through LA data use policies be simplified, and effort spent to ensure the communication takes place and students understand it. In addition, students need a way to communicate their data use preferences to HEIs. Where possible, students' data use preferences need to be integrated into the (LA) data use policies, and ultimately, the ethical frameworks which guide implementation of LA in higher education. In effect, the thesis research suggests that a 'bottom-up' approach is explored where students' needs feed into the development process for ethical LA applications. It is hoped that these steps will mark the way forward to minimising the identified privacy and utility tensions.

9 References

Acquisti, A., & Grossklags, J. (2005). Privacy and Rationality in Individual Decision Making. *IEEE Security & Privacy, 3*, 26–33.

Adejo, O., & Connolly, T. (2017). Learning analytics in a shared-network educational environment: Ethical issues and countermeasures. *International Journal of Advanced Computer Science and Applications, 8*, 22-29.

Aguilar, S., Lonn, S., & Teasley, S. D. (2014). Perceptions and Use of an Early Warning System During a Higher Education Transition Program. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 113–117). New York, NY, USA: ACM. doi:10.1145/2567574.2567625

Aldridge, A., & Levine, K. (2001). *Surveying the social world: Principles and practice in survey research.* Open University Press.

Amigud, A., Arnedo-Moreno, J., Daradoumis, T., & Guerrero-Roldan, A.-E. (2018). An integrative review of security and integrity strategies in an academic environment: Current understanding and emerging perspectives. *Computers & Security, 76*, 50 – 70. doi:https://doi.org/10.1016/j.cose.2018.02.021

Anton, A. I., Earp, J. B., & Young, J. D. (2010). How Internet Users' Privacy Concerns Have Evolved Since 2002. *IEEE Security Privacy, 8*, 21–27. doi:10.1109/MSP.2010.38

Arbaugh, J. B. (2000). Virtual classroom characteristics and student satisfaction with internet-based MBA courses. *Journal of management education, 24*, 32-54.

Arnold, K. E., & Sclater, N. (2017). Student Perceptions of their Privacy in Leaning Analytics Applications. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, (pp. 66–69).

Atkins, L., & Wallace, S. (2012). Interviewing in educational research. SAGE publications. doi:10.4135/9781473957602.n6

Baker, R. S., Gašević, D., & Karumbaiah, S. (2021). Four paradigms in learning analytics: Why paradigm convergence matters. *Computers and Education: Artificial Intelligence, 2*.

Barabas, J., & Jerit, J. (2010). Are survey experiments externally valid? *American Political Science Review, 104*, 226–242.

Barter, C., & Renold, E. (1999). The Use of Vignettes in Qualitative Research. *Social research update, 25*, 1-6.

Baruh, L., & Popescu, M. (2017). Big Data Analytics And The Limits Of Privacy Self-Management. *New Media and Society, 19*, 579–596. doi:10.1177/1461444815614001

Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication, 67*, 26–53.

Bautista, P., & Inventado, P. S. (2021). Protecting Student Privacy with Synthetic Data from Generative Adversarial Networks. *Artificial Intelligence in Education. AIED 2021. Lecture Notes in Computer Science. 12749.* Springer, Cham. doi:10.1007/978-3-030-78270-2_11

BBC. (2020, 8 20). A-levels and GCSEs: How did the exam algorithm work. *BBC News*. Retrieved from https://www.bbc.co.uk/news/explainers-53807730

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior research methods, 43*, 800–813.

Bielick, S. (2017). Surveys and Questionnaires. In *The BERA/SAGE Handbook of Educational Research: Two Volume Set* (pp. 640–658). SAGE Publications. doi:10.4135/9781473983953.n32

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research, 26*, 1802-1811. doi:10.1177/1049732316654870

Black, C., Setterfield, L., & Warren, R. (2018). Online Data Privacy from Attitudes to Action: an evidence review. *Carnegie UK Trust*.

Blandford, A. (2013). Semi-structured Qualitative Studies. In *The Encyclopedia of Human-Computer Interaction* (Second ed.). Interaction Design Foundation.

Blandford, A., Furniss, D., & Makri, S. (2016). *Qualitative HCI Research: Going Behind the Scenes.* Morgan & Claypool.

Bodily, R., & Verbet, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies, 10*, 405–418.

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education*(30), 572–598. doi:10.1007/s12528-018-9186-0

Boeren, E. (2018). The Methodological Underdog: A Review of Quantitative Research in the Key Adult Education Journals. *Adult Education Quarterly, 68*, 63–79. doi:10.1177/0741713617739347

Botes, A. (2000). A comparison between the ethics of justice and the ethics of care. *Journal of Advanced Nursing, 32*, 1071 – 1075.

boyd, d. m., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication, 13*, 210-230. doi:10.1111/j.1083-6101.2007.00393.x

Braunack-Mayer, A. J., Street, J. M., Tooher, R., Feng, X., & Scharling-Gamba, K. (2020). Student and staff perspectives on the use of big data in the tertiary education sector: A scoping review and reflection on the ethical issues. *Review of Educational Research, 90*, 788–823. doi:10.3102/0034654320960213

Brooker, A., Corrin, L., Fisher, J., & Mirriahi, N. (2017). Defining "data" in conversations with students about the ethical use of learning analytics. *Ascilite2017.*

Brooks, R., te Riele, & Maguire, M. (2014). Ethical theories, principles and guidelines. In *Ethics and education research* (pp. 18-43). SAGE Publications Ltd. doi:10.4135/9781473909762

Bryman, A. (2016). *Social Research Methods* (5th ed.). Oxford University Press.

Buchanan, T., Paine, C., Joinson, A. N., & Reips, U.-D. (2007). Development of Measures of Online Privacy Concern and Protection for Use on the Internet. *Journal of the American Society for Information Science and Technology, 58*, 157-165.

Bulger, M., McCormick, P., & Pitcan, M. (2017). The legacy of inbloom. *The legacy of inbloom*. Data &Society Research Institute.

Burnard, P., Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Analysing and presenting qualitative data. *British Dental Journal, 204*, 429–432.

Calvert, C. E. (2014). Developing a model and applications for probabilities of student success: A case study of predictive analytics. *Open Learning: The Journal of Open, Distance and e-Learning, 29*(2), 160-173. doi:10.1080/02680513.2014.931805

Canning, J. (2017). Conceptualising student voice in UK higher education: four theoretical lenses. *Teaching in Higher Education, 22*(5), 519 – 531. doi:10.1080/13562517.2016.1273207

Chaudhari, V. (2021). Learning in Uncertainty: Examining the Relationship Between Perceived EnvironmentalUncertainty and Self-Regulated Learning of Finance Professionals, and The

Role of Technology in Supporting It. *Learning in Uncertainty: Examining the Relationship Between Perceived EnvironmentalUncertainty and Self-Regulated Learning of Finance Professionals, and The Role of Technology in Supporting It*.

Cho, H., Rivera-Sánchez, M., & Lim, S. S. (2009). A Multinational Study on Online Privacy: Global Concerns and Local Responses. *New Media & Society, 11*, 395–416. doi:10.1177/1461444808101618

Christ, T. W. (2013). The worldview matrix as a strategy when designing mixed methods research. *International Journal of Multiple Research Approaches, 7*, 110–118. doi:10.5172/mra.2013.7.1.110

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). routledge.

Coles-Kemp, L., & Kani-Zabihi, E. (2010). On-line Privacy and Consent: A Dialogue, Not a Monologue. *Proceedings of the 2010 New Security Paradigms Workshop* (pp. 95–106). ACM. doi:10.1145/1900546.1900560

Colman, A. M. (2015). *A dictionary of psychology.* Oxford University Press. doi:10.1093/acref/9780199657681.001.0001

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Buckingham Shum, S., . . . Fisher, J. (2016). Student retention and learning analytics: a snapshot of Australian practices and a framework for advancement. *Student retention and learning analytics: a snapshot of Australian practices and a framework for advancement*. Australian Government Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/project-student-retention-and-learning-analytics-snapshot-current-australian-practices-and-framework

Cormack, A. N. (2016). A Data Protection Framework for Learning Analytics. *Journal of Learning Analytics, 3*, 91-106.

Cormack, A. N. (2016). Downstream Consent: A Better Legal Framework for Big Data. *Journal of Information Rights, Policy, and Practice, 1*.

Cornish, F., Gillespie, A., & Zittoun, T. (2013). Collaborative Analysis of Qualitative Data. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 79 – 93). SAGE Publications Ltd.

Corrin, L., Kennedy, G., French, S., Shum, S. B., Kitto, K., Pardo, A., . . . Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper. *The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper*. Retrieved from

https://melbourne-cshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics

Cram, F., & Mertens, D. M. (2015). Transformative and indigenous frameworks for multimethod and mixed methods research. In *The Oxford handbook of multimethod and mixed methods research inquiry* (pp. 91–109).

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research.* SAGE Publications.

Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science, 10*, 104–115.

Curry, L., & Nunez-Smith, M. (2015). Assessing quality in mixed methods studies. In *Mixed Methods in Health Sciences Research: A Practical Primer* (pp. 169–200). SAGE Publications, Inc.

Daniel. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British journal of educational technology, 46*, 904–920.

Daniel. (2017). Overview of Big Data and Analytics in Higher Education. In B. K. Daniel (Ed.), *Big Data and Learning Analytics in Higher Education.* Springer International Publishing Switzerland. doi:10.1007/978-3-319-06520-5_1

Daniel, B. K., & Butson, R. (2013). Technology Enhanced Analytics (TEA) in Higher Education.

Daniel, E. M., & White, A. (2005). The Future of Inter-organisational System Linkages: Findings of an International Delphi Study. *European Journal of Information Systems, 14*, 188-203.

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. *Qualitative research, 14*, 603–616.

Delbecq, A. L., Ven, A. H., & Gustafson, D. H. (1975). *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes.* Scott, Foresman and Company.

DeSantis, L., & Ugarriza, D. N. (2000). The Concept of Theme as Used in Qualitative Nursing Research. *Western Journal of Nursing Research, 22*, 351–372. doi:10.1177/019394590002200308

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*, 61-80.

Doring, A., Hodge, A., & Heo, M. (2014). Online learners and their self-disclosure preferences. *Journal of Information Technology Education: Research, 13*, 163–175.

Drachsler, H., & Greller, W. (2012). The Pulse of Learning Analytics Understandings and Expectations from the Stakeholders. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (pp. 120–129). New York, NY, USA: ACM. doi:10.1145/2330601.2330634

Drachsler, H., & Greller, W. (2016). Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 89-98). ACM.

Drachsler, H., Hoel, T., Cooper, A., Kismihók, G., Berg, A., Scheffel, M., . . . Ferguson, R. (2016). Ethical and Privacy Issues in the Design of Learning Analytics Applications. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 492–493). New York, NY, USA: Association for Computing Machinery. doi:10.1145/2883851.2883933

Drachsler, H., Hoel, T., Scheffel, M., Kismihók, G., Berg, A., Ferguson, R., . . . Manderveld, J. (2015). Ethical and privacy issues in the application of learning analytics. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, (pp. 390–391).

Dunlap, J., Palmer, I., & Wesley, A. (2021). Keeping Student Trust: Student Perceptions of Data Use within Higher Education. *Keeping Student Trust: Student Perceptions of Data Use within Higher Education.* New America.

Edinburgh, T. U. (2018). Policy and procedures for developign and managing Learning Analytics activities. Retrieved from https://www.ed.ac.uk/files/atoms/files/learninganalyticspolicy.pdf

Egelman, S., Chi, E. H., & Dow, S. (2014). Crowdsourcing in hci research. In *Ways of Knowing in HCI* (pp. 267-289). Springer.

Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*.

Falcao, T. P., Ferreira, R., Rodrigues, R. L., Diniz, J., & Gašević, D. (2019). Students' perceptions about learning analytics in a Brazilian higher education institution. *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), 2161*, pp. 204–206.

Fallon, G., & Brown, R. B. (2002). Focusing on focus groups: lessons from a research project involving a Bangladeshi community. *Qualitative Research, 2*, 195-208. doi:10.1177/146879410200200204

Feathers, T. (2021). Major universities are using race as a "high impact predictor" of student success. *Major universities are using race as a "high impact predictor" of student success*. The Markup. Retrieved from https://themarkup.org/news/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success

Ferguson, R. (2012). Learning Analytics: Drivers, Developments and Challenges. *International Journal of Technology Enhanced Learning, 4*, 304–317.

Ferguson, R. (2019). Ethical Challenges for Learning Analytics. *Journal of Learning Analytics, 6*, 25-30.

Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S. (2014). Setting learning analytics in context: overcoming the barriers to large-scale adoption. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, (pp. 251–253).

Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. *Journal of learning analytics, 3*, 5–15.

Finch, J. (1987). The Vignette Technique in Survey Research. *Sociology, 21*, 105-114.

Finlay, L., & Gough, B. (2008). *Reflexivity: A practical guide for researchers in health and social sciences.* John Wiley & Sons.

FitzGerald, E., Kucirkova, N., Jones, A., Cross, S., Ferguson, R., Herodotou, C., . . . Scanlon, E. (2018). Dimensions of personalisation in technology-enhanced learning: A framework and implications for design. *British Journal of Educational Technology, 49*, 165–181. doi:10.1111/bjet.12534

Floridi, L. (2005). The Ontological Interpretation of Informational Privacy. *Ethics and Information Technology, 7*, 185-200. doi:10.1007/s10676-006-0001-7

Follmer, D. J., Sperling, R. A., & Suen, H. K. (2017). The role of MTurk in education research: Advantages, issues, and future directions. *Educational Researcher, 46*, 329–334.

Foster, C., & Francis, P. (2020). A systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes. *Assessment & Evaluation in Higher Education, 45*, 822–841. doi:10.1080/02602938.2019.1696945

Gable, G. G. (1994). Integrating case study and survey research methods: an example in information systems. *European journal of information systems, 3*, 112–126.

Gasevic, D., Dawson, S., & Jovanovic, J. (2016). Ethics and Privacy as Enablers of Learning Analytics. *Journal of Learning Analytics, 3*, 1–4.

Gaskell, G., & Bauer, M. W. (2000). Towards public accountability: Beyond sampling, reliability and validity. In *Qualitative researching with text, image and sound: A practical handbook* (pp. 336–350). Sage Thousand Oaks, CA.

Gergle, D., & Tan, D. S. (2014). Experimental research in HCI. In *Ways of Knowing in HCI* (pp. 191–227). Springer.

Given, L. M. (2008). *The Sage encyclopedia of qualitative research methods.* Sage publications.

Glasow, P. A. (2005). Fundamentals of survey research methodology. *Fundamentals of survey research methodology*. Retrieved from http://www.mitre.org/sites/default/files/pdf/05_0638.pdf

Glogowska, M., Young, P., & Lockyer, L. (2011). Propriety, process and purpose: considerations of the use of the telephone interview method in an educational research study. *Higher Education, 62*, 17–26. doi:10.1007/s10734-010-9362-2

Gray, G., Schalk, A., Rooney, P., & Lang, C. (2021). A Stakeholder Informed Professional Development Framework to Support Engagement with Learning Analytics. *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 237–247). New York, NY, USA: Association for Computing Machinery. doi:10.1145/3448139.3448162

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society, 15*, 42.

Griffiths, D. (2017). An ethical waiver for learning analytics? In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & Pérez-Sanagustín (Eds.), *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017, Proceedings* (pp. 557–560). Springer International Publishing. doi:10.1007/978-3-319-66610-5_62

Griffiths, D., Drachsler, H., Kickmeier-Rust, M., Steiner, C., Hoel, T., & Greller, W. (2016). Is Privacy a Show-stopper for Learning Analytics? A Review of Current Issues and their Solutions. *Is Privacy a Show-stopper for Learning Analytics? A Review of Current Issues and their Solutions*. Retrieved from http://www.laceproject.eu/learning-analytics-review/privacy-show-stopper/

Groves, R. M., Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). John Wiley & Sons.

Groves, R. M., Presser, S., & Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. *The Public Opinion Quarterly, 68*, 2–31. Retrieved from http://www.jstor.org/stable/3521534

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods, 18*, 59–82. doi:10.1177/1525822X05279903

Gupta, U. G., & Clarke, R. E. (1996). Theory and Applications of the Delphi Technique: A Bibliography (1975–1994). *Technological Forecasting & Social Change, 53*(2), 185–211.

Gursoy, M. E., Inan, A., Nergiz, M. E., & Saygin, Y. (2017). Privacy-Preserving Learning Analytics: Challenges and Techniques. *IEEE Transactions on Learning Technologies, 10*, 68-81. doi:10.1109/TLT.2016.2607747

Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. *Review of Educational Research*. doi:10.3102/00346543211020116

Halcomb, E., & Hickman, L. (2015). Mixed Methods Research. *Nursing Standard, 29*(32), 41–47.

Hasson, F., & Keeney, S. (2011). Enhancing Rigor in the Delphi Technique Research. *Technological Forecasting & Social Change, 78*(9), 1695–1704.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research Guidelines for the Delphi Survey Technique. *Journal of Advanced Nursing, 32*(4), 1008–1015.

Hawkins, J. E. (2018). The practical utility and suitability of email interviews in qualitative research. *The Qualitative Report, 23*(2), 493–501.

Heath, J. (2014). Contemporary Privacy Theory Contributions to Learning Analytics. *Journal of Learning Analytics, 1*, 140-149.

Herodotou, C., Naydenova, G., Boroowa, A., Gilmour, A., & Rienties, B. (2020). How Can Predictive Learning Analytics and Motivational Interventions Increase Student Retention and Enhance Administrative Support in Distance Education? *Journal of Learning Analytics, 7*, 72–83.

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. *Educational Technology Research and Development, 67*, 1273-1306.

Herschel, R., & Miori, V. M. (2017). Ethics & Big Data. *Technology in Society, 49*, 31 – 36.

Hesse-Biber, S. (2015). Introduction: Navigating a Turbulent Research Landscape: Working the Boundaries, Tensions, Diversity, and Contradictions of Multimethod and Mixed Methods Inquiry. In *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry.*

Hewson, C. (2009). Gathering data on the internet: qualitative approaches and possibilities for mixed methods research. In *Oxford Handbook of Internet Psychology* (pp. 405-428).

Ho, A. (2017). Advancing Educational Research and Student Privacy in the "Big Data" Era. *Workshop on Big Data in Education: Balancing the Benefits of Educational Research and Student Privacy.* National Academy of Education.

Hoel, T., & Chen, W. (2016). Privacy-driven design of learning analytics applications–exploring the design space of solutions for data sharing and interoperability. *Journal of Learning Analytics, 3*, 139–158.

Hoel, T., & Chen, W. (2018). Advancing the Delicate Issue of Ethics and Privacy for Learning Analytics. *Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge.* Sydney: Society for Learning Analytics Research.

Hoel, T., Griffiths, D., & Chen, W. (2017). The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 243–252). New York, NY, USA: Association for Computing Machinery. doi:10.1145/3027385.3027414

Hoel, T., Mason, J., & Chen, W. (2015). Data Sharing for Learning Analytics - Questioning the Risks and Benefits. *Proceedings of the 23rd International Conference on Computers in Education.*

Hommel, M., Egetenmeier, A., & Maier, U. (2019). Supporting Stakeholders with Learning Analytics to Increase Study Success. In *Utilizing learning analytics to support study success* (pp. 37–60). Springer.

Hoofnagle, C. J., King, J., Li, S., & Turow, J. (2010). How Different are Young Adults from Older Adults When it Comes to Information Privacy Attitudes and Policies? doi:10.2139/ssrn.1589864

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies.* Sage publications.

Iachello, G., & Hong, J. (2007). End-User Privacy in Human–Computer Interaction. *Foundations and Trends® in Human–Computer Interaction, 1*, 1–137. doi:10.1561/1100000004

Ifenthaler, D., & Schumacher, C. (2015). Divulging Personal Information within Learning Analytics Systems. *International Association for Development of the Information Society*.

Ifenthaler, D., & Schumacher, C. (2016). Student Perceptions of Privacy Principles for Learning Analytics. *Educational Technology Research and Development, 64*, 923-938.

Ifenthaler, D., & Schumacher, C. (2016). Student Perceptions of Privacy Principles for Learning Analytics. *Educational Technology Research and Development, 64*, 923–938.

Ifenthaler, D., & Schumacher, C. (2019). Releasing personal information within learning analytics systems. In *Learning Technologies for Transforming Large-Scale Teaching, Learning, and Assessment* (pp. 3-18). Springer.

Ifenthaler, D., & Yau, J. Y.-K. (2019). Higher Education Stakeholders' Views on Guiding the Implementation of Learning Analytics for Study Success. *Personalised learning. Diverse Goals. One heart. ASCILITE 2019 Singapore*, (pp. 453-457).

Ifenthaler, D., Mah, D.-K., & Yau, J. Y.-K. (2019). Utilising learning analytics for study success: Reflections on current empirical findings. In *Utilizing learning analytics to support study success* (pp. 27–36). Springer.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*, 6–47.

Jiang, Z., Heng, C. S., & Choi, B. C. (2013). Research note - Privacy concerns and privacy-protective behavior in synchronous online social interactions. *Information Systems Research, 24*, 579–595.

Johnson. (2017). Ethics and justice in learning analytics. *New Directions for Higher Education, 2017*, 77–87.

Johnson, D. R., & Borden, L. A. (2012). Participants at Your Fingertips: Using Amazon's Mechanical Turk to Increase Student–Faculty Collaborative Research. *Teaching of Psychology, 39*, 245–251. doi:10.1177/0098628312456615

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33*, 14–26. doi:10.3102/0013189X033007014

Joksimović, S., Kovanović, V., & Dawson, S. (2019). The journey of learning analytics. *HERDSA Review of Higher Education, 6*, 27–63.

Jones. (2019). Learning Analytics And Higher Education: A Proposed Model for Establishing Informed Consent Mechanisms to Promote Student Privacy and Autonomy. *International Journal of Educational Technology in Higher Education, 16*. doi:https://doi.org/10.1186/s41239-019-0155-0

Jones, K. M., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*.

Jones, K. M., Perry, M. R., Goben, A., Asher, A., Briney, K. A., Robertshaw, M. B., & Salo, D. (2019). In Their Own Words: Student Perspectives on Privacy and Library Participation in Learning Analytics Initiatives. *2019 Association of College and Research Libraries Conference.* American Library Association.

Jones, K. M., Rubel, A., & LeClere, E. (2020). A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics. *Journal of the Association for Information Science and Technology, 71*, 1227–1241.

Jones, K., VanScoy, A., Bright, K., & Harding, A. (2021). Do They Even Care? Measuring Instructor Value of Student Privacy in the Context of Learning Analytics. *Proceedings of the 54th Hawaii International Conference on System Sciences*, (pp. 1529 – 1537).

Kappers, W. M., & Cutler, S. L. (2014). Poll Everywhere! Even in the Classroom: An Investigation into the Impact of Using Poll Everywhere in a Large Lecture Classroom. *The ASEE Computers in Education (CoED) Journal, 6*, 21.

Kay, D., Korn, N., & Oppenheim, C. (2012). Legal, Risk and Ethical Aspects of Analytics in Higher Education. *CETIS Analytics Series, 1*, 1–30.

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the Oracle: Ten Lessons From Using The Delphi Technique in Nursing Research. *Journal of Advanced Nursing, 53*(2), 205–212.

Kelly, B., Margolis, M., McCormack, L., LeBaron, P. A., & Chowdhury, D. (2017). What affects people's willingness to participate in qualitative research? An experimental comparison of five incentives. *Field Methods, 29*, 333–350.

Kelly, B., McCormack, M., Reeves, J., Brooks, D. C., & O'Brien, J. (2021). 2021 EDUCAUSE Horizon Report, Information Security Edition. *2021 EDUCAUSE Horizon Report, Information Security Edition*. EDUCAUSE.

Kennett, D. J., O'Hagan, F. T., & Cezer, D. (2008). Learned Resourcefulness and the Long-Term Benefits of a Chronic Pain Management Program. *Journal of Mixed Methods Research, 2*, 317–339. doi:10.1177/1558689808319732

Kenworthy, B. (2001). Supporting the student in new teaching and learning environments. In *The 9th Cambridge International Conference on Open and Distance Learning; Supporting the student in open and distance learning; Collated Conference Papers* (pp. 76 – 82). Routledge.

Kitchin, R. (2013). Big Data and Human Geography: Opportunities, Challenges and Risks. *Dialogues in Human Geography, 3*, 262-267. doi:10.1177/2043820613513388

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology, 50*, 2855–2870.

Knight, D. B., Brozina, C., & Novoselich, B. (2016). An Investigation of First-Year Engineering Student and Instructor Perspectives of Learning Analytics Approaches. *Journal of Learning Analytics, 3*, 215-238. doi:10.18608/jla.2016.33.11

Knight, S., Rienties, B., Littleton, K., Mitsui, M., Tempelaar, D., & Shah, C. (2017). The relationship of (perceived) epistemic cognition to interaction with resources on the internet. *Computers in Human Behavior, 73*, 507–518.

Knox, J. (2017). Data Power in Education: Exploring Critical Awareness with the "Learning Analytics Report Card". *Television & New Media, 18*, 734–752. doi:10.1177/1527476417690029

Kobsa, A. (2007). Privacy-enhanced Personalization. *Communications of the ACM, 50*, 24–33. doi:10.1145/1278201.1278202

Kokolakis, S. (2017). Privacy Attitudes and Privacy Behaviour: A Review of Current Research on the Privacy Paradox Phenomenon. *Computers & Security, 64*, 122-134. doi:https://doi.org/10.1016/j.cose.2015.07.002

Korir, M. (2021). *A systematic and privacy-oriented analysis of ethics and privacy frameworks for learning analytics.* Tech. rep.

Krol, K., Spring, J. M., Parkin, S., & Sasse, M. A. (2016). Towards Robust Experimental Design for User Studies in Security and Privacy. *The LASER Workshop: Learning from Authoritative Security Experiment Results (LASER 2016)* (pp. 21–31). USENIX Association. Retrieved from https://www.usenix.org/conference/laser2016/program/presentation/krol

Krosnick, J. A. (2018). Questionnaire Design. In D. L. Vannette, & J. A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research* (pp. 439-455). Springer International Publishing. doi:10.1007/978-3-319-54395-6_53

Kruse, A., & Pongsajapan, R. (2012). Student-Centered Learning Analytics. *CNDLS Thought Papers*, 1–9.

Kukulska-Hulme, A., Bossu, C., Coughlan, T., Ferguson, R., FitzGerald, E., Gaved, M., . . . Zhang, S. (2021). *Innovating Pedagogy 2021: Open University Innovation Report 9.* Milton, Keynes, UK: The Open University.

Kumaraguru, P., & Cranor, L. F. (2005). *Privacy Indexes: A Survey of Westin's Studies.* Carnegie Mellon University, School of Computer Science, Institute for Software Research International.

Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., & Wolff, A. (2015). OU Analyse: analysing at-risk students at The Open University. *Learning Analytics Review*, 1-16.

Laakso, M.-J., Kurvinen, E., Enges-Pyykönen, P., & Kaila, E. (2018). Designing and creating a framework for learning analytics in Finland. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, (pp. 0695–0700).

Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can Lottery Incentives Boost Web Survey Response Rates? Findings from Four Experiments. *Research in Higher Education, 52*, 537–553. doi:10.1007/s11162-010-9203-2

Laufer, R. S., & Wolfe, M. (1977). Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. *Journal of Social Issues, 33*, 22-42. doi:10.1111/j.1540-4560.1977.tb01880.x

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction.* Wiley Publishing.

Lee, H., Wong, S. F., Oh, J., & Chang, Y. (2019). Information privacy concerns and demographic characteristics: Data from a Korean media panel survey. *Government Information Quarterly, 36*, 294 – 303. doi:https://doi.org/10.1016/j.giq.2019.01.002

Li. (2012). Theories in Online Information Privacy Research: A Critical Review and an Integrated Framework. *Decision Support Systems, 54*, 471 – 481. doi:https://doi.org/10.1016/j.dss.2012.06.010

Li, H., Rathindra, S., & Xu, H. (2010). Understanding situational online information disclosure as a privacy calculus. *The Journal of Computer Information Systems, 51*, 62 – 71.

Li, W., Brooks, C., & Schaub, F. (2019). The impact of student opt-out on educational predictive models. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, (pp. 411–420).

Li, W., Sun, K., Schaub, F., & Brooks, C. (2021). Disparities in Students' Propensity to Consent to Learning Analytics. *International Journal of Artificial Intelligence in Education*. doi:10.1007/s40593-021-00254-2

Lichtman, M. (2013). *Qualitative Research in Education: A User's Guide* (3rd ed.). Sage Publications.

Lietz, P. (2010). Research into Questionnaire Design. *International Journal of Market Research, 52*, 249–272.

Linneberg, & Korsgaard, S. (2019). Coding Qualitative Data: a Synthesis Guiding the Novice. *Qualitative Research Journal, 19*, 259–270. doi:10.1108/QRJ-12-2018-0012

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review, 46*, 30–40.

Mahroeian, H., Daniel, B. K., & Butson, R. (2017). The perceptions of the meaning and value of analytics in New Zealand higher education institutions. *International Journal of Educational Technology in Higher Education, 14*, 1–17.

Mai, J.-E. (2016). Three Models of Privacy. *Nordicom Review*, 171 -- 175.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research, 15*, 336-355.

Mandal, J., Acharya, & Parija, S. C. (2011). Ethics in human research. *Tropical parasitology, 1*(1), 2–3. doi:10.4103/2229-5070.72105

Mason, J. (2002). *Qualitative researching* (Second ed.). Sage Publications Ltd.

May, M., Fessakis, G., Dimitracopoulou, A., & George, S. (2012). A Study on User's Perception in E-Learning Security and Privacy Issues. *2012 IEEE 12th International Conference on Advanced Learning Technologies*, (pp. 88–89).

McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *IS: A Journal of Law and Policy for the Information Society, 4*(3).

McIntosh, N. E. (1979). Barriers to implementing research in higher education. *Studies in Higher Education, 4*, 77–86. doi:10.1080/03075077912331377121

McKim, C. A. (2017). The value of mixed methods research: A mixed methods study. *Journal of Mixed Methods Research, 11*, 202–222.

Mittelmeier, J., Edwards, R. L., Davis, S. K., Nguyen, Q., Murphy, V. L., Brummer, L., & Rienties, B. (2018). 'A double-edged sword. This is powerful but it could be used destructively': Perspectives of Early Career Education Researchers on Learning Analytics. *Frontline Learning Research, 6*, 20-38.

Mittelmeier, J., Rienties, B., Tempelaar, D., & Whitelock, D. (2018). Overcoming Cross-Cultural Group Work Tensions: Mixed Student Perspectives on the Role of Social Relationships. *Higher Education, 75*, 149–166. doi:10.1007/s10734-017-0131-3

Mittelmeier, J., Rienties, B., Tempelaar, D., Hillaire, G., & Whitelock, D. (2018). The influence of internationalised versus local content on online intercultural collaboration in groups: A randomised control trial study in a statistics course. *Computers & Education, 118*, 82–95. doi:10.1016/j.compedu.2017.11.003

Morgan, D. L., & Spanish, M. T. (1984). Focus groups: A new tool for qualitative research. *Qualitative sociology, 7*, 253–270.

Müller, H., Sedley, A., & Ferrall-Nunge, E. (2014). Survey research in HCI. In *Ways of Knowing in HCI* (pp. 229–266). Springer.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science, 2*, 109–138.

Murmann, P., Reinhardt, D., & Fischer-Hübner, S. (2019). To Be, or Not to Be Notified. *IFIP International Conference on ICT Systems Security and Privacy Protection*, (pp. 209–222). doi:/10.1007/978-3-030-22312-0_15

Murry, J. W., & Hammons, J. O. (1995). Delphi: A Versatile Methodology for Conducting Qualitative Research. *The Review of Higher Education, 18*(4), 423.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization, 17*, 2–26. doi:https://doi.org/10.1016/j.infoandorg.2006.11.001

Naeini, P. E., Bhagavatula, S., Habib, H., Degeling, M., Bauer, L., Cranor, L. F., & Sadeh, N. (2017). Privacy Expectations and Preferences in an IoT World. *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)* (pp. 399-412). USENIX Association. Retrieved from https://www.usenix.org/conference/soups2017/technical-sessions/presentation/naeini

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). The Belmont report: Ethical principles and guidelines for the protection of human subjects of research: Superintendent of Documents. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research: Superintendent of Documents.*

Nelson, K. J., & Creagh, T. A. (2013). A Good Practice Guide: Safeguarding Student Learning Engagement.

Nevaranta, M., Lempinen, K., & Kaila, E. (2020). Students' Perceptions about Data Safety and Ethics in Learning Analytics. *Proceedings of the Conference on Technology Ethics*, (pp. 23–37).

Newland, B., & Trueman, P. (2017). *Learning Analytics in UK HE 2017 A HeLF Survey Report. Heads of eLearning Group.* Tech. rep. Retrieved from https://drive.google.com/file/d/0Bx5p8mXFK743dWFvdGFranZSUk0/view

Nielsen, K., Randall, R., & Christensen, K. B. (2017). Do Different Training Conditions Facilitate Team Implementation? A Quasi-Experimental Mixed Methods Study. *Journal of Mixed Methods Research, 11*, 223-247. doi:10.1177/1558689815589050

Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review, 79*, 119-158.

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life.* Stanford University Press.

Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus, 140*, 32–48.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs, 41*, 100-126. doi:10.1111/j.1745-6606.2006.00070.x

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*. doi:10.1177/1609406917733847

Olson, J. S., Grudin, J., & Horvitz, E. (2005). A Study of Preferences for Sharing and Privacy. *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1985–1988). ACM. doi:10.1145/1056808.1057073

Open University Human Research Ethics Committee. (2021). Conducting Human Research durign the Covid-19 Pandemic. *Conducting Human Research durign the Covid-19 Pandemic*. Retrieved from https://www.open.ac.uk/research/governance/ethics/human/review-process/conducting-human-research-during-covid-19

Oppenheim, A. N. (2000). *Questionnaire Design, Interviewing and Attitude Measurement.* Bloomsbury Publishing.

Oulasvirta, A., Suomalainen, T., Hamari, J., Lampinen, A., & Karvonen, K. (2014). Transparency of Intentions Decreases Privacy Concerns in Ubiquitous Surveillance. *Cyberpsychology, Behavior, and Social Networking, 17*, 633–638.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*.

Paine, C., Reips, U.-D., Stieger, S., Joinson, A., & Buchanan, T. (2007). Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies, 65*(6), 526–536.

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments., *17*, pp. 22-27. doi:10.1016/j.jbef.2017.12.004.

Pardo, A., & Siemens, G. (2014). Ethical and Privacy Principles for Learning Analytics. *British Journal of Educational Technology, 45*, 438-450.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*, 128–138. doi:https://doi.org/10.1111/bjet.12592

Paré, G., Cameron, A.-F., Poba-Nzaou, P., & Templier, M. (2013). A Systematic Assessment of Rigor in Information Systems Ranking-type Delphi Studies. *Information & Management, 50*, 207–217.

Parrish, A. H., & Richman, L. J. (2020). Dual perspectives on learning analytics in higher education. *Journal of Applied Research in Higher Education, 12*, 4–14. doi:10.1108/JARHE-12-2018-0254

Pedersen, M. J., & Nielsen, C. V. (2016). Improving Survey Response Rates in Online Panels: Effects of Low-Cost Incentives and Cost-Free Text Appeal Interventions. *Social Science Computer Review, 34*, 229-243. doi:10.1177/0894439314563916

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153-163. doi:10.1016/j.jesp.2017.01.006

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2016). Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. *Proceedings of the Sixth international conference on Learning Analytics & Knowledge*, (pp. 64–73).

Plano Clark, V. L., & Badiee, M. (2010). Research questions in mixed methods research. In *SAGE handbook of mixed methods in social & behavioral research* (pp. 275–304). SAGE Publications, Inc. doi:10.4135/9781506335193

Plesch, C., Kaendler, C., Rummel, N., Wiedmann, M., & Spada, H. (2013). Identifying Areas of Tension in the Field of Technology-enhanced Learning: Results of an International Delphi Study. *Computers and Education, 65*, 92-105. doi:10.1016/j.compedu.2013.01.018

Ponciano, L., Barbosa, P., Brasileiro, F., Brito, A., & Andrade, N. (2017). Designing for Pragmatists and Fundamentalists: Privacy Concerns and Attitudes on the Internet of Things. *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, (pp. 1-10).

Preibusch, S. (2013). Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human-Computer Studies, 71*, 1133–1143.

Prettyman, S. S., Furman, S., Theofanos, M., & Stanton, B. (2015). Privacy and security in the brave new world: The use of multiple mental models. *International conference on human aspects of information security, privacy, and trust*, (pp. 260–270).

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, (pp. 240–244).

Prinsloo, P., & Slade, S. (2015). Student Privacy Self-management: Implications for Learning Analytics. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, (pp. 83–92).

Prinsloo, P., & Slade, S. (2017). Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. In *Big data and learning analytics in higher education* (pp. 109–124). Springer International Publishing. doi:10.1007/978-3-319-06520-5_8

Prinsloo, P., & Slade, S. (2018). Student Consent in Learning Analytics: The Devil in the Details? In *Learning analytics in higher education: Current innovations, future potential, and pratical applications* (pp. 118-139). Routledge.

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies, 3*(9). doi:10.46827/ejes.v0i0.1017

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide.* John Wiley & Sons.

Reeve, D. (2017). Preparing for the General Data Protection Regulation (GDPR). *Preparing for the General Data Protection Regulation (GDPR)*. JISC. Retrieved from https://www.jisc.ac.uk/guides/preparing-for-the-general-data-protection-regulation-gdpr

Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: Twenty-first century student sorting and tracking. *Ethics and Information Technology, 21*, 167–179.

Reidenberg, J. R., & Schaub, F. (2018). Achieving Big Data Privacy in Education. *Theory and Research in Education*, 1-17.

Richards, N. M. (2012). The dangers of surveillance. *Harvard Law Review, 126*, 1934 – 1965.

Rienties, B., & Héliot, Y. (2018). Enhancing (In)Formal Learning Ties in Interdisciplinary Management Courses: A Quasi-Experimental Social Network Study. *Studies in Higher Education, 43*, 437-451. doi:10.1080/03075079.2016.1174986

Rienties, B., & Toetenel, L. (2016). The Impact of Learning Design on Student Behaviour, Satisfaction and Performance: A Cross-institutional Comparison Across 151 Modules. *Computers in Human Behavior, 60*, 333–341.

Rienties, B., Herodotou, C., Olney, T., Schencks, M., & Boroowa, A. (2018). Making sense of learning analytics dashboards: A technology acceptance perspective of 95 teachers. *International Review of Research in Open and Distributed Learning, 19*.

Roberts, L. D., Howell, J. A., & Seaman, K. (2017). Give me a Customizable Dashboard: Personalized Learning Analytics Dashboards in Higher Education. *Technology, Knowledge and Learning, 22*, 317–333.

Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student Attitudes toward Learning Analytics in Higher Education: ``The Fitbit Version of the Learning World". *Frontiers in Psychology, 7*.

Roessler, B. (2018). Three Dimensions of Privacy. In B. van der Sloot, & A. de Groot (Eds.). Amsterdam University Press.

Rubel, A., & Jones, K. M. (2016). Student Privacy in Learning Analytics: An Information Ethics Perspective. *The Information Society, 32*, 143-159.

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.). Sage.

Sale, J. E., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and quantity, 36*, 43–53.

Sax, M. (2018). Privacy from an Ethical Perspective. In B. van der Sloot, & A. de Groot (Eds.). Amsterdam University Press.

Schairer, C. E., Cheung, C., Kseniya Rubanovich, C., Cho, M., Cranor, L. F., & Bloss, C. S. (2019, 4). Disposition toward privacy and information disclosure in the context of emerging health technologies. *Journal of the American Medical Informatics Association, 26*, 610-619. doi:10.1093/jamia/ocz010

Scheffel, M., Drachsler, H., Kreijns, K., de Kraker, J., & Specht, M. (2017). Widget, Widget As You Lead, I Am Performing Well Indeed!: Using Results from an Exploratory Offline Study to Inform an Empirical Online Study About a Learning Analytics Widget in a Collaborative Learning Environment. *Proceedings of the Seventh International Learning Analytics and Knowledge Conference* (pp. 289–298). ACM.

Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development, 64*, 939–955.

Schumacher, C., & Ifenthaler, D. (2018). Features Students Really Expect from Learning Analytics. *Computers in Human Behavior, 78*, 397–407.

Schutt, R. K. (2017). *Investigating the social world: The process and practice of research.* SAGE publications.

Sclater, N. (2016). Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics, 3*, 16-42.

Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education: A review of UK and international practice.

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics, 6*, 11–19. doi:10.18608/jla.2019.63.3

Selwyn, N., Henderson, M., & Chao, S.-H. (2018). 'You need a system': exploring the role of data in the administration of university students and courses. *Journal of Further and Higher Education, 42*, 46–56.

Shannon-Baker, P. (2016). Making Paradigms Meaningful in Mixed Methods Research. *Journal of Mixed Methods Research, 10*(4).

Shapiro, B., & Baker, C. R. (2001). Information technology and the social construction of information privacy. *Journal of Accounting and Public Policy, 20*, 295-322. doi:10.1016/S0278-4254(01)00037-0

Sharples, M. (2020). A Short Introduction to Educational Technology. *A Short Introduction to Educational Technology*.

Sheehan, K. B. (1999). An Investigation of Gender Differences in On-line Privacy Concerns and Resultant Behaviors. *Journal of Interactive Marketing, 13*, 24-38. doi:https://doi.org/10.1002/(SICI)1520-6653(199923)13:4<24::AID-DIR3>3.0.CO;2-O

Sheehan, K. B. (2002). Toward a Typology of Internet Users and Online Privacy Concerns. *The Information Society, 18*, 21–32. doi:10.1080/01972240252818207

Shuster, E. (1997). Fifty Years Later: The Significance of the Nuremberg Code. *New England Journal of Medicine, 337*, 1436–1440. doi:10.1056/nejm199711133372006

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist, 57*, 1380-1400.

Silvola, A., Näykki, P., Kaveri, A., & Muukkonen, H. (2021). Expectations for supporting student engagement with learning analytics: An academic path perspective. *Computers & Education, 168*, 104192. doi:https://doi.org/10.1016/j.compedu.2021.104192

Singer, N. (2014). InBloom Student Data Repository to Close. *Business Innovation Technology Society - BitsNYTimes*. Retrieved from https://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close/

Slade, S., & Boroowa, A. (2014). Policy on Ethical Use of Student Data for Learning Analytics. Retrieved from http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/fil es/ecms/web-content/ethical-use-of-student-data-policy.pdf

Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist, 57*, 1510-1529.

Slade, S., & Prinsloo, P. (2014). Student Perspectives on the Use of their Data: Between Intrusion, Surveillance and Care. *Challenges for Research into Open & Distance Learning: Doing Things Better ? Doing Better Things*, (pp. 291-300).

Slade, S., & Tait, A. (2019). Global Guidelines: Ethics in Learning Analytics. *Global Guidelines: Ethics in Learning Analytics*. International Council for Open and Distance Education.

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning Analytics at the Intersections of Student Trust, Disclosure and Benefit. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 235–244). ACM. doi:10.1145/3303772.3303796

Smith, H. J., Dinev, T., & Xu, H. (2011). Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly, 35*, 989–1016.

Snelson, C., Rice, K., & Wyzard, C. (2012, 1). Research Priorities for YouTube and Video-sharing Technologies: A Delphi Study. *British Journal of Educational Technology, 43*, 119–129.

Solove. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review, 154*.

Solove. (2009). *Understanding Privacy.* Harvard University Press.

Solove. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review, 126*.

Solove, D. J. (2002). Conceptualizing privacy. *California Law Review, 90*, 1087.

Sønderlund, A. L., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology, 50*, 2594-2618.

Spall, S. (1998). Peer debriefing in qualitative research: Emerging operational models. *Qualitative inquiry, 4*, 280–292.

Steiner, C. M., Kickmeier-Rust, M. D., & Albert, D. (2016). LEA in Private: A Privacy and Data Protection Framework for a Learning Analytics Toolbox. *Journal of Learning Analytics, 3*, 66–90.

Strunk, K. K., & Mwavita, M. (2020). *Design and analysis in educational research: Anova designs in SPSS.* Routledge.

Sullivan, A., & Calderwood, L. (2017). Surveys: Longitudinal, Cross-sectional and Trend Studies. In *The BERA/SAGE Handbook of Educational Research: Two Volume Set* (pp. 395–415). SAGE Publications. doi:10.4135/9781473983953.n20.

Sun, K., Mhaidli, A. H., Watel, S., Brooks, C. A., & Schaub, F. (2019). It's My Data! Tensions Among Stakeholders of a Learning Analytics Dashboard. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 594:1–594:14). ACM. doi:10.1145/3290605.3300824

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education, 50*, 1183-1202. doi:https://doi.org/10.1016/j.compedu.2006.11.007

Swenson, J. (2014). Establishing an ethical literacy for learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, (pp. 246–250).

Taylor, H. (2003). Most People Are ``Privacy Pragmatists'' Who, While Concerned About Privacy, Will Sometimes Trade It Off for Other Benefits. *Most People Are ``Privacy Pragmatists'' Who, While Concerned About Privacy, Will Sometimes Trade It Off for Other Benefits*. Harris Interactive.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157–167. doi:https://doi.org/10.1016/j.chb.2014.05.038

The British Psychological Society. (2021). BPS Code of Human Research Ethics. *BPS Code of Human Research Ethics*. The British Psychological Society.

Tolley, E. E., Ulin, P. R., Mack, N., Robinson, E. T., & Succop, S. M. (2016). *Qualitative methods in public health: a field guide for applied research* (Second ed.). John Wiley & Sons.

Tsai, Y.-S., Perrotta, C., & Gašević, D. (2019). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2019.1676396

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, (pp. 230–239).

Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C.-C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education, 106*, A1–A9.

UN General Assembly. (1948). Universal Declaration of Human Rights.

van der Sloot, B. (2018). Privacy from a Legal Perspective. In B. van der Sloot, & A. de Groot (Eds.). Amsterdam University Press.

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The Current Landscape of Learning Analytics in Higher Education. *Computers in Human Behavior, 89*, 98 – 110. doi:https://doi.org/10.1016/j.chb.2018.07.027

Voida, A., Mynatt, E. D., Erickson, T., & Kellogg, W. A. (2004). Interviewing over Instant Messaging. (pp. 1344–1347). Association for Computing Machinery. doi:10.1145/985921.986060

Vu, P., Adkins, M., & Henderson, S. (2019). Aware, But Don't Really Care: Students' Perspective on Privacy and Data Collection in Online Courses. *Journal of Open Flexible and Distance Learning, 23*, 42-51.

Waheed, H., Hassan, S.-U., Aljohani, N. R., & Wasif, M. (2018). A bibliometric perspective of learning analytics research landscape. *Behaviour & Information Technology, 37*, 941-957. doi:10.1080/0144929X.2018.1467967

Warren, S. D., & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review, 4*, 193-220.

Welsh, S., & McKinney, S. (2015). Clearing the Fog: A Learning Analytics Code of Practice. *Conference Proceedings: Australasian Society for Computers in Learning and Tertiary Education (Ascilite2015).*

West, D., Huijser, H., & Heath, D. (2016). Putting an Ethical Lens on Learning Analytics. *Educational Technology Research and Development, 64*, 903–922.

West, D., Luzeckyj, A., Toohey, D., Vanderlelie, J., & Searle, B. (2020). Do academics and university administrators really know better? The ethics of positioning student perspectives in

learning analytics. *Australasian Journal of Educational Technology*. doi:10.14742/ajet.4653

Westin. (1967). *Privacy and Freedom.* New York: Atheneum.

Westin, A. F., & Maurici, D. (1998). *E-commerce and Privacy: What Net Users Want.* Privacy and American Business Hackensack, NJ.

White, T. I. (2017). *Right and wrong: A practical introduction to ethics.* John Wiley & Sons.

Whitelock-Wainwright, A., Gašević, D., Tejeiro, R., Tsai, Y.-S., & Bennett, K. (2019). The student expectations of learning analytics questionnaire. *Journal of Computer Assisted Learning, 35*, 633–666.

Whitley, E. A. (2009). Informational privacy, consent and the "control" of personal data. *Information Security Technical Report, 14*, 154–159. doi:10.1016/j.istr.2009.10.001

Wiles, R. (2013). *What are Qualitative Research Ethics?* Bloomsbury Academic. doi:10.5040/9781849666558

Wilkinson, S. (1998). Focus group methodology: a review. *International Journal of Social Research Methodology, 1*, 181–203. doi:10.1080/13645579.1998.10846874

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*. doi:10.33151/ajp.8.3.93

Williamson. (2002). The Delphi Method. In *Research Methods for Students, Academics and Professionals : Information Management and Systems.* ProQuest Ebook Central: Elsevier Science & Technology.

Williamson. (2017). *Big Data in Education: The Digital Future of Learning, Policy and Practice.* Sage Publishing.

Williamson, C., Rooyen, A. V., Shuttleworth, C., Binnekade, C., & Scott, D. (2020). Wuity as a Philosophical Lens for Qualitative Data Analysis. *International Journal of Qualitative Methods, 19*. doi:10.1177/1609406920926885

Williamson, K., & Kizilcec, R. F. (2021). Learning Analytics Dashboard Research Has Neglected Diversity, Equity and Inclusion. *Proceedings of the Eighth ACM Conference on Learning @ Scale* (pp. 287–290). Association for Computing Machinery. doi:10.1145/3430895.3460160

Willis, J. E. (2014). Learning analytics and ethics: A framework beyond utilitarianism. *Educause Review*.

Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Annual Meeting Proceedings*. doi:10.5465/AMBPP.2018.15903abstract

Wintrup, J. (2017). Higher Education's Panopticon? Learning Analytics, Ethics and Student Engagement. *Higher Education Policy, 30*, 87-103. doi:10.1057/s41307-016-0030-8

Wollny, S., Schneider, J., Mitri, D. D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? - A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence, 4*. doi:10.3389/frai.2021.654924

Wong, B. T.-M., Li, K. C., & Choi, S. P.-M. (2018). Trends in Learning Analytics Practices: A Review of Higher Education Institutions. *Interactive Technology and Smart Education, 15*, 132-154. doi:10.1108/ITSE-12-2017-0065

Woodruff, A., Pihur, V., Consolvo, S., Brandimarte, L., & Acquisti, A. (2014). Would a Privacy Fundamentalist Sell Their DNA for $1000...If Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences. *10th Symposium On Usable Privacy and Security (SOUPS 2014)* (pp. 1–18). Menlo: USENIX Association.

Woodyatt, C. R., Finneran, C. A., & Stephenson, R. (2016). In-Person Versus Online Focus Group Discussions: A Comparative Analysis of Data Quality. *Qualitative Health Research, 26*, 741–749. doi:10.1177/1049732316631510

Worcester, R. M., & Burns, T. R. (1975). Statistical Examination of Relative Precision of Verbal Scales. *Journal of the Market Research Society, 17*, 181–197.

World Medical Association. (2013). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *JAMA, 310*, 2191–2194. doi:10.1001/jama.2013.281053

Xu, H., & Teo, H.-H. (2004). Alleviating Consumers' Privacy Concerns in Location-Based Services: A Psychological Control Perspective. *International Conference on Information Systems (ICIS).*

Xu, H., Dinev, T., Smith, J., & Hart, P. (2011). Information privacy concerns: Linking individual perceptions with institutional privacy assurances. *Journal of the Association for Information Systems, 12*, 798 – 824.

Xu, H., Luo, X. R., Carroll, J. M., & Rosson, M. B. (2011). The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decision Support Systems, 51*, 42–52.

Yan, Z., Robertson, T., Yan, R., Park, S. Y., Bordoff, S., Chen, Q., & Sprissler, E. (2018). Finding the weakest links in the weakest link: How well do undergraduate students make cybersecurity judgment? *Computers in Human Behavior, 84*, 375–382. doi:https://doi.org/10.1016/j.chb.2018.02.019

Yeung, K. (2018). Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism. *International Data Privacy Law, 8*(3).

York, C. S., & Ertmer, P. A. (2011). Towards an Understanding of Instructional Design Heuristics: An Exploratory Delphi Study. *Educational Technology Research and Development, 59*, 841–863.

Zafeiropoulou, A. M., Millard, D. E., Webber, C., & O'Hara, K. (2013). Unpicking the Privacy Paradox: Can Structuration Theory Help to Explain Location-based Privacy Decisions? *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 463–472). New York, NY, USA: ACM. doi:10.1145/2464464.2464503

Appendix A   Study 1

Appendix A.1   Ethical Approval

Human Research Ethics Committee (HREC)

From	Dr Duncan Banks
	The Open University Human Research Ethics Committee
Email	duncan.banks@open.ac.uk
Extension	(6) 59198
To	Maina Korir, WELS CREET
Project title:	Experts' Views of Ethics and Privacy in Learning Analytics.

Memorandum

HREC ref   HREC/2018/2826/Korir

Date application submitted: 23/03/18
Date of HREC response: 26/03/18

This memorandum is to confirm that the research protocol for the above-named research project, as submitted for ethics review, has been given a favourable opinion by Chair's action by the Open University Human Research Ethics Committee.

Please note the following:

1. You are responsible for notifying the HREC immediately of any information received by you, or of which you become aware which would cast doubt on, or alter, any information contained in the original application, or a later amendment which would raise questions about the safety and/or continued conduct of the research.

2. It is essential that any proposed amendments to the research are sent to the HREC for review, so they can be recorded and where required, a favourable opinion given prior to any changes being implemented (except only in cases of emergency when the welfare of the participant or researcher is or may be affected).

3. Please include your HREC reference number in any documents or correspondence. It is essential that it is included in any publicity related to your research, e.g. when seeking participants or advertising your research so it is clear that it has been reviewed by HREC and adheres to OU ethics review processes.

4. You are authorised to present this memorandum to outside bodies such as NHS Research Ethics Committees in support of any application for future research clearance. Also, where there is an external ethics review, a copy of the application and outcome should be sent to the HREC.

5. OU research ethics review procedures are fully compliant with the majority of grant awarding bodies and where they exist, their frameworks for research ethics.

6. At the end of your project, you are required to assess your research for ethics related issues and/or major changes. Where these have occurred you will need to provide the Committee with a HREC final report to reflect how these were dealt with using the final report template on the research ethics website - http://www.open.ac.uk/research/ethics/human-research/full-review-process-and-proforma#final_report

Best regards

Dr Duncan Banks, The Open University Human Research Ethics Committee

www.open.ac.uk/research/ethics/   November 2017

Appendix A.2   Recruitment Email

VIEWS ON PRIVACY IN LEARNING ANALYTICS

Dear [expert's name],

My name is Maina Korir and I am a doctoral researcher at the Open University, UK. I am a member of the Institute of Educational Technology. My research focuses on privacy in learning analytics and is supervised by Dr. Sharon Slade, Dr. Wayne Holmes, and Professor Bart Rienties.

I would like to investigate privacy issues that arise in learning analytics and how they are being addressed. Many of these issues are documented in ethical and privacy frameworks published since 2012. However, the field of learning analytics is constantly changing and improving, and I would like to obtain a current *and* future-oriented perspective.

You have been identified as having specialist knowledge of ethics and privacy in general or ethics and privacy in learning analytics. I would like to invite you to take part in a Delphi study. The aim of this study is to seek your input on the following themes:

1. Conceptualising privacy
2. Identifying the main privacy issues
3. Identifying the privacy issues that are yet to be tackled
4. Identifying the questions that are yet to be asked about privacy in learning analytics.

Delphi Study Protocol

I plan to conduct the Delphi study in three rounds from April to July 2018. The first round is exploratory in nature and will run in April 2018. In this round, I will seek your views and opinions on privacy in learning analytics. I will analyse your responses and use them, along with findings from ongoing research, to prepare a questionnaire for the second round. I will conduct the second round in May and June 2018. In the third round, I will analyse the responses from the second round and send these back to you along with other participants' responses for your comment and input. I will carry out the third round in June and July 2018.

Only I will have access to your individual responses. I will use pseudonyms during data processing and analysis to remove the link between you and your responses. At the end of the study, I will share the names of the participating experts to enhance the trustworthiness of the project and results. Your participation in this study is entirely voluntary and you can withdraw your participation until the 30th of April, 2018.

If you would like to take part in this study, please send me an email at maina.korir@open.ac.uk to express your interest.

I would appreciate it if you could forward this invitation to others who might be willing to participate. I am looking for input from people who conduct research or write on ethics and privacy in general or ethics and privacy in learning analytics. Thank you very much for your help.

This research has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee - HREC reference number: 2826 (http://www.open.ac.uk/research/ethics/).

Appendix A.3   Interview Script

Introduce myself and my research – 3 minutes

Good morning/afternoon. Thank you for making time for today's interview. I am a first year PhD student at the Open University in the UK. In this study, I want to engage with experts to explore student privacy in learning analytics and to pursue insights into how the field might develop in the future. My work is focused on students in higher education.

Confirm consent for audio recording – 2 minutes

Before we begin, I would like to confirm your consent for me to record our interview today. Are you comfortable with me audio recording our interview?

Participant consents – begin audio recording.

Thank you very much. [turn on audio recorder]

Participant does not consent –

I have noted that you do not consent to my recording our interview today. I will now proceed to the questions and I will take notes of your responses during the interview.

This recording is on <day> at <time> and I have received consent for audio recording.

Settling in questions – 5 minutes

1.   To settle us in as we begin, could you tell me more about your work? What is your role in the organisation?

2.   [If relevant] How did you start working in learning analytics? How did your research focus on ethical and/or privacy issues?

Study questions – 20 minutes

Let's now move on to specific questions about student privacy and learning analytics.

3.   There are a number of views in the literature about what student privacy is. For example, some authors refer to limiting access to student data, others refer to the subject retaining control over their own data. How do you define or conceptualise student privacy in learning analytics?

4.   Has your view of student privacy changed over time? What contributed to this change?

5.   In your view, what are the three main student privacy issues in learning analytics?

6.   Does the ranking matter, that is, is one more important than the other?

Prompt 1: Do you think data protection, student control over learning analytics, or transparency over data collection and sharing are the most important issues?

7.   Why do you think these are the top issues?

8.   Do you think that student privacy in learning analytics has changed over the last few years?

Prompt: If yes, please describe how you think it has changed.

If no, do you have any thoughts on why there have been no changes? – got it right, more pressing issues to address first?

9.   Are there things you feel should change in student privacy in learning analytics? Prompt: For example, a shift to more student control, less institutional control….

10.   There have been calls for more research focusing on students in learning analytics. What is the biggest hurdle to students' use of learning analytics in the coming years?

11.   As an expert in the field, what do you think I am missing? Is there a question you think I should have asked but haven't?

Thank you and wrap up – 2 minutes

12.   Do you have any comments to make about the study or the questions?

Thanks again for your time. If you have any questions or comments, contact information is available on the participant information sheet.

Appendix B   Study 2

Appendix B.1   Screenshots Shown to Participants

Nearest students

Demographic 50%   VLE 50%
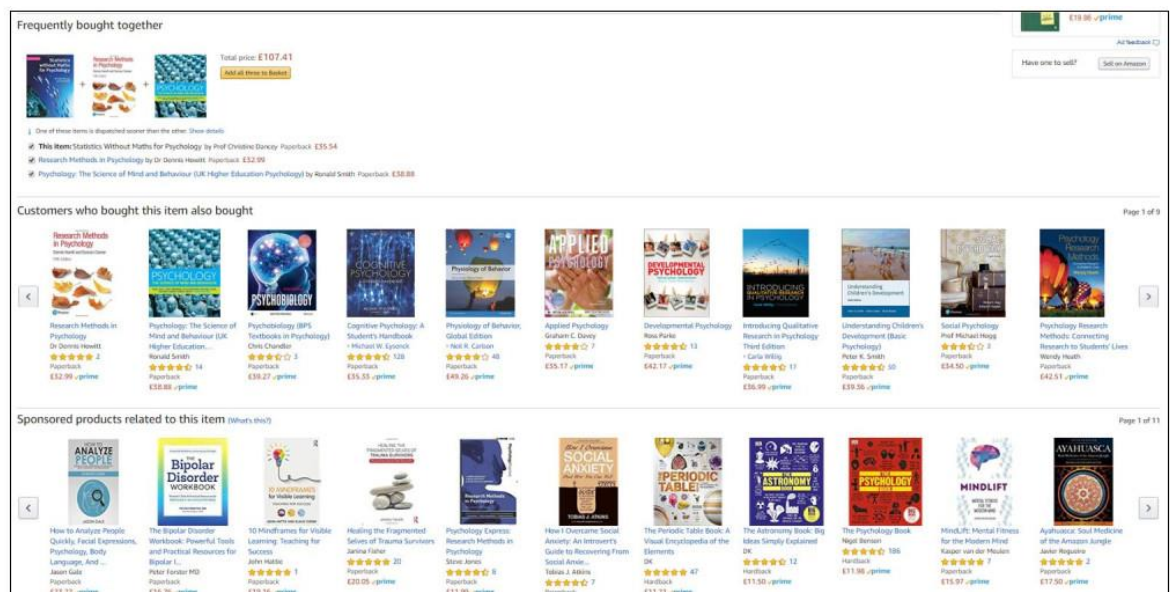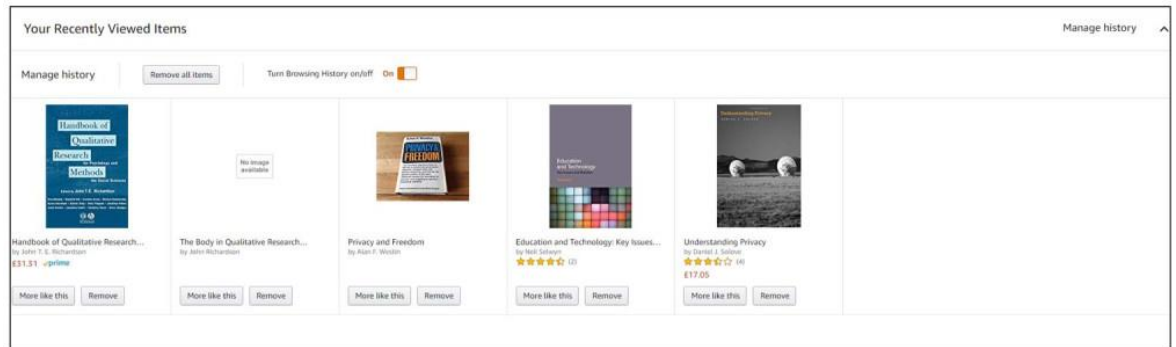
Scores

Assignment	Prediction	REAL	Justification
TMA 01	Submit	88	Resource VLE activity in week 4 >=0 Resource VLE activity in week 3 >=0 quiz VLE activity in week 3 >=0
TMA 02	Submit	74	quiz VLE activity in week 7 >=0 oucollaborate VLE activity in week 7 >=0 quiz VLE activity in week 8 >=0
TMA 03	Submit	Not submit	Forum VLE activity in week 10 >=0 Homepage VLE activity in week 12 >=0 summary VLE activity in week 12 >=0
TMA 04	NA	NA	NA
TMA 05	NA	NA	NA

Activity recommender

Visit Block 2 Part 2 (online version).	Visit Activity 28.	Visit Introduction to spreadsheets.	Visit Activity 18.	Visit The Penalty Kick (story for Activity 37).	Visit TMA questions and guidance.

Appendix B.2   Participant Information Sheet

Participant Information Sheet

Study title: Ethical Practices in Organizations in Collecting and Analysing Customer Data

What is the purpose of the study?
We are inviting you to take part in a study evaluating organisation's data practices in collecting and analysing customer data.

Why have I been approached?
For the purposes of the study we need to recruit a number of adult participants.

Do I have to take part?
No. Participation is entirely voluntary. If you change your mind about taking part in the study you can withdraw at any point during the session and at any time in the four weeks following (upto 15th November, 2018). If you decide to withdraw all your data will be destroyed and will not be used in the study. There are no consequences to deciding that you no longer wish to participate in the study.

What happens during the study?
The study is divided into two parts: The first part will be conducted in the lab session. The second part will take place 1 or 2 weeks after the lab session and will involve interviews which you can opt to attend to discuss your responses to the questions in the study.

The study will involve a number of scenarios on organisational practices of collection and use of customer data. Each scenario will be described, and you will have the opportunity to answer several questions on these scenarios. The session is interactive, and you will also be able to discuss your peers' responses to the same questions. In collaboration with Dr Yingfei Heliot we will collate demographic data for analysis about students' views on organisational data practices. The demographic data will include student username, course of study, nationality/ethnicity, age, and gender.

What are the possible disadvantages and risks of taking part?
We do not anticipate any risks associated with participation in this study.

What are the possible benefits of taking part?
You will gain an insight into how a research project is conducted and what it is like to be a participant in such a study. You will analyse the scenarios according to your individual perspective and compare this with your peers. You will receive feedback on the responses and gain insights into ethical issues surrounding the collection and use of customer data in organisations.

Will my taking part in this study be kept confidential?
Yes. No personally identifying information will be shared.

What will happen to the results of the research study?

This research forms part of Maina Korir's doctoral research at the Open University supervised by Dr. Sharon Slade, Dr. Wayne Holmes and Professor Bart Rienties. The research is being undertaken in collaboration with Dr Yingfei Heliot at the University of Surrey.

Data will be shared between the OU team and Dr. Heliot for research, educational, and dissemination purposes only. All data will be available to Dr. Heliot in her capacity as course leader. De-anonymised data will not be shared other than within the OU supervisory and external-examiner team except where we are legally bound to do so. Pseudonyms will be used in reporting.

The data will be anonymised after the lab session. The anonymised data will be kept until 2023, after which it will be reviewed with a view to deletion when appropriate. Data stored will be kept in a password protected file in accordance with the General Data Protection Regulation requirements.

Who is organising and funding the research?
The research is organised by Maina Korir, who is a research student at the Open University's Institute of Educational Technology. It is funded by a Leverhulme Doctoral Scholarship.

Who has reviewed the study?
The University of Surrey and Open University Ethics Committees have both reviewed and approved this study.

Contact for Further Information
Maina Korir
Institute of Educational Technology,
Walton Hall, Milton Keynes, MK7 6AA
Email: maina.korir@open.ac.uk

Consent
This study and these materials have been designed in accordance with the Ethical Guidelines for Educational Research (2018) by the British Educational Research Association, in addition to the University of Surrey and The Open University's ethics and data protection guidance and procedures and the associated regulatory frameworks.

This research project research project has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee - HREC reference number:

Appendix B.3   Privacy Profile Sent to Participants

Dear [Student Name],

Thank you for your participation in our organizational behaviour lab session on Monday 15th of October 2018. We looked at privacy and personality and used two scenarios of organisational data practices as discussion pieces. The lab session was a follow up to the personality lecture where you received a personality profile. Research has identified that personality plays a role in privacy concerns. For example, Korzaan and Boswell (2008) found that people who score high on agreeableness have high concern for information privacy.

Privacy profile

We used four scales to measure your privacy attitudes. The first scale (Fig. 1(a)) was developed by Alan Westin (Westin, 1998), whereby we looked at your overall privacy concerns (e.g., Most businesses handle the personal information they collect about consumers in a proper and confidential way). A high score on this scale suggests that privacy has high value and you hold the view that organisations do not need or are not entitled to customers' personal information for their programs. The second, third and fourth scales were developed by Buchanan et al. (2007). General caution (Fig. 1(b)) measures general caution in protecting your privacy (e.g., Do you hide your bank card PIN number when using cash machines/making purchases?) A high score on this scale suggests that you are generally cautious when online. Technical protection (Fig. 1(c)) looks at using technology to protect privacy and requires some knowledge of technical tools (e.g., Do you check your computer for spy ware?). A high score on this scale suggests some technical awareness on how to protect your privacy when online. Finally, privacy concern (Fig. 1(d)) looks at your attitudes towards privacy on the Internet (e.g., Are you concerned about people online not being who they say they are?). A high score on this scale suggests high concern about privacy online. As highlighted by the frequency graphs below, there is some substantial variation in the scores, indicating that you and your peer students have different perspectives on these privacy issues. As indicated before, there is no right or wrong perspective on this, but it might be useful to know that like your peers customers in organisations will have different perspectives on privacy, and therefore this will influence how organisations need to deal with variation in perspectives.

Your score: 2.7000000000000002
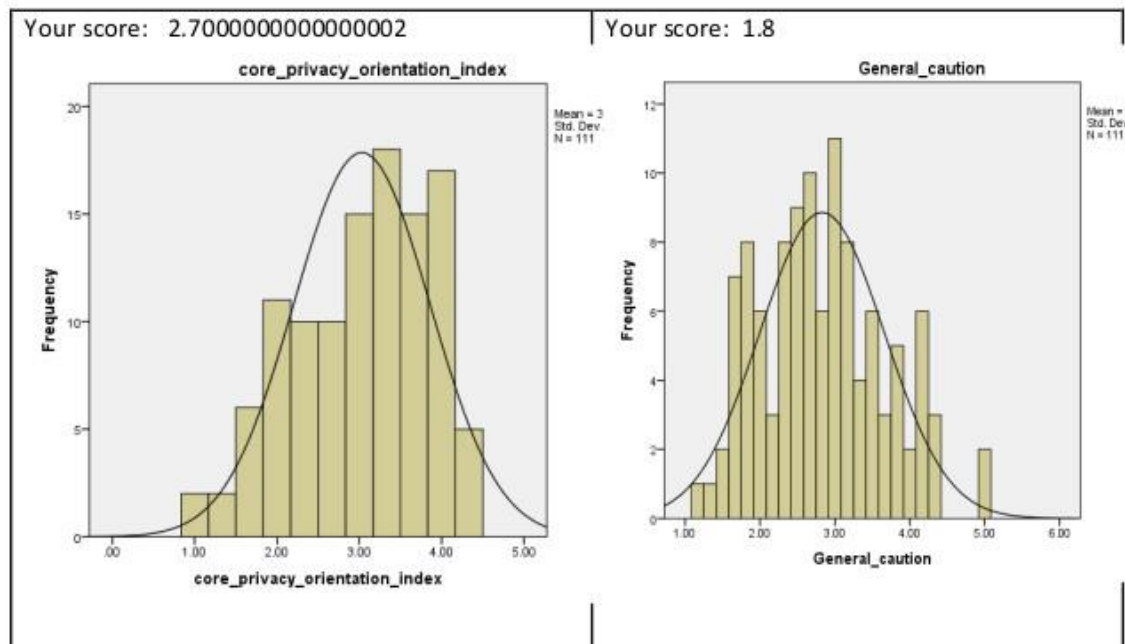
Your score: 1.8

Your score: 2

Your score: 2.7999999999999998
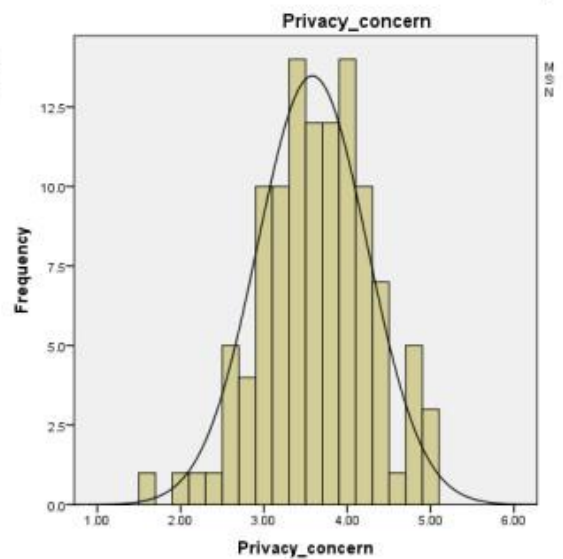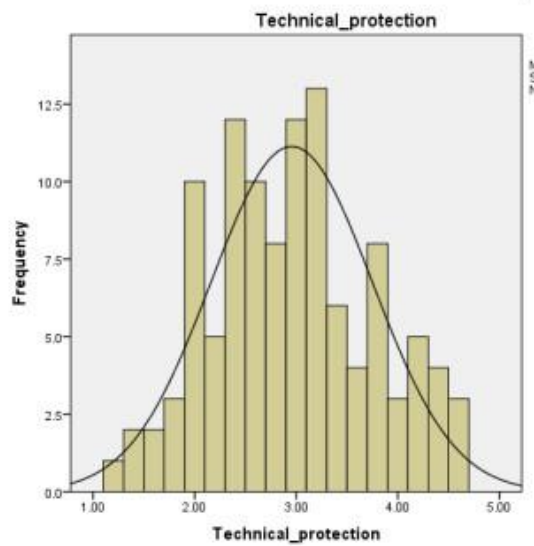
As indicated during the lab sessions, we are keen to hear more about your perspectives. You can share your perspectives by replying to this email. If you opt to send an email, we will deal with it in complete confidence and will not share it with the lecturer or the University of Surrey. Alternatively, we will arrange follow-up interviews, so if you have not already indicated that you may want to participate, please let me know.

Best wishes,

Maina Korir

More information on general protection of your privacy
To find out more about protecting your privacy online take a look at the following resource:
Surveillance self-defence: Tips, tools and how-tos for safer online communications

REFERENCES
Buchanan, Tom, Paine, Carina, Joinson, Adam N., and Reips, Ulf-Dietrich (2007). "Development of Measures of Online Privacy Concern and Protection for Use on the Internet". In *Journal of the American Society for Information Science and Technology*, 58.2, pp. 157-165.

Korzaan, Melinda L. and Boswell, Katherine T. (2008). "The Influence of Personality Traits and Information Privacy Concerns on Behavioral Intentions". In *Journal of Computer Information Systems*, 48.4, pp. 15-24.

Westin, Alan, and Harris Louis & Associates (1998). E-Commerce & Privacy: What Net Users Want. Technical report. Conducted for Privacy & American Business and PricewaterhouseCoopers.

Appendix B.4   Study Instruments

Participants were asked the following questions from the privacy index questionnaire:

For each of the following statements, how strongly do you agree or disagree? [1= Strongly disagree, 2 = Somewhat disagree, 3 = Neither agree nor disagree, 4 = Somewhat agree, 5 = Strongly agree]:

- Statement 1: Consumers have lost all control over how personal information is collected and used by companies.
- Statement 2: Most businesses handle the personal information they collect about consumers in a proper and confidential way.
- Statement 3: Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Questions on general caution	Response options	No. of students with missing data
Do you shred/burn your personal documents when you are disposing of them?	Never, Rarely, Sometimes, Very often, Always, Not applicable	1
Do you hide your bank card PIN number when using cash machines / making purchases?	Never, Rarely, Sometimes, Very often, Always, Not applicable	
Do you only register for websites that have a privacy policy?	Never, Rarely, Sometimes, Very often, Always, Not applicable	2
Do you read a website's privacy policy before you register your information?	Never, Rarely, Sometimes, Very often, Always, Not applicable	1
Do you look for a privacy certification on a website before you register your information?	Never, Rarely, Sometimes, Very often, Always, Not applicable	1
Do you read license agreements fully before you agree to them?	Never, Rarely, Sometimes, Very often, Always, Not applicable	1

Questions on technical protection	Response options
Do you watch for ways to control what people send you online (such as check boxes that allow you to opt-in or opt-out of certain offers)?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Do you remove cookies?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Do you use a pop-up window blocker?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Do you check your computer for spyware?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Do you clear your browser history regularly?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Do you block messages/emails from someone you do not want to hear from?	This question was excluded as it was not relevant for the study context

Questions on privacy concern	Response options
In general, how concerned are you about your privacy while you are using the Internet?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Are you concerned about online organisations not being who they claim they are?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Are you concerned that you are asked for too much personal information when you register or make online purchases?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Are you concerned about online identity theft?	Never, Rarely, Sometimes, Very often, Always, Not applicable
Are you concerned about people online not being who they say they are?	Never, Rarely, Sometimes, Very often, Always, Not applicable

Are you concerned that information about you could be found on an old computer?	Never, Rarely, Sometimes, Very often, Always, Not applicable	
Are you concerned who might access your medical records electronically?	Never, Rarely, Sometimes, Very often, Always, Not applicable	
Are you concerned about people you do not know obtaining personal information about you from your online activities?	Never, Rarely, Sometimes, Very often, Always, Not applicable	
Are you concerned that if you use your credit card to buy something on the Internet your credit card number will be obtained/intercepted by someone else?	Never, Rarely, Sometimes, Very often, Always, Not applicable	
Are you concerned that if you use your credit card to buy something on the internet your card will be mischarged?	Never, Rarely, Sometimes, Very often, Always, Not applicable	

Questions on the Amazon vignette	Response options	No. of students with missing data
Have you signed up for an Amazon account?	Y/N	
I feel comfortable that Amazon can offer me a better service (e.g., offers based on my buying or search patterns) by collecting my personal data?	Totally disagree, Disagree, Neutral, Agree, Totally agree	32
I feel comfortable that Amazon shares my personal and online activity data, in a personally identifiable way, with third parties?	Totally disagree, Disagree, Neutral, Agree, Totally agree	33
I feel comfortable that Amazon shares my personal and online	Totally disagree, Disagree, Neutral, Agree, Totally agree	36

activity data, in an anonymised format, with third parties?		
I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online?	Totally disagree, Disagree, Neutral, Agree, Totally agree	35
I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online and assures me that my data will not be shared with third parties?	Totally disagree, Disagree, Neutral, Agree, Totally agree	38
I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online on condition that my data will be shared with third parties?	Totally disagree, Disagree, Neutral, Agree, Totally agree	48
What is the overwhelming feeling you have about the Amazon scenario?	Open responses	

Questions on the university vignette	Response options	No. of students with missing data
I would feel comfortable that my personal and online activity data is shared with my tutor to help him/her to improve support to me	Totally disagree, Disagree, Neutral, Agree, Totally agree	
I feel comfortable that the University can offer me a better service (e.g., alerts on potential problems or recommendations of	Totally disagree, Disagree, Neutral, Agree, Totally agree	34

learning resources) by collecting my personal data?		
I feel comfortable that the University shares my personal and online activity data, in a personally identifiable way, with third parties?	Totally disagree, Disagree, Neutral, Agree, Totally agree	60
I feel comfortable that the University shares my personal and online activity data, in an anonymised format, with third parties?	Totally disagree, Disagree, Neutral, Agree, Totally agree	34
I feel comfortable that the University offers me specific benefits in exchange for tracking me online?	Totally disagree, Disagree, Neutral, Agree, Totally agree	34
I feel comfortable that the University offers me specific benefits in exchange for tracking me online and assures me that my data will not be shared with third parties?	Totally disagree, Disagree, Neutral, Agree, Totally agree	36
I feel comfortable that the University offers me specific benefits in exchange for tracking me online on condition that my data will be shared with third parties?	Totally disagree, Disagree, Neutral, Agree, Totally agree	33
What is the overwhelming feeling you have about the University student learning dashboard scenario?	Open responses	

Appendix B.5 Follow-Up Interview Schedule

INTERVIEW SCHEDULE

Introduce myself and my research

Good morning/afternoon. Thank you for making time for today's interview. My name is Maina and I am a 2nd year PhD student at the Open University.

This interview is a follow up to the Organizational Behaviour (OB) lab where we focused on personality and privacy using two scenarios – personalised recommendations on things you can purchase from Amazon, and using student data for a student learning dashboard and to improve learning at the University of Surrey (this was a hypothetical scenario). I would like to discuss your responses to the OB lab questions, and understand more about your perspective on how organizations use customer data.

Settling in

1. To get us settled in, could you tell me briefly about your experience as a graduate student at the University of Surrey?

   Probe: Build on what they mention of interest/relevance to settle in

Questions on responses in the lab session

2. In the lab session we looked at two scenarios – Amazon and a student facing learning dashboard. I will recap the questions and remind you of your response. I will then invite you to tell me more about your response.

   a. The first question I would like to focus on asked "I feel comfortable that Amazon can offer me a better service (e.g. offers based on my buying or search patterns) by collecting my personal data?" and "I feel comfortable that the University can offer me a better service (e.g. alerts on potential problems or recommendations of learning resources) by collecting my personal data?" You stated [insert student's answer]. Could you tell me more about your responses to these questions?

   b. Next, we asked "I feel comfortable that Amazon shares my personal and online activity data, in a personally identifiable way, with third parties?" and "I feel comfortable that the University shares my personal and online activity data, in a personally identifiable way, with third parties?" You stated [insert student's answer]. Could you tell me more about your responses to these questions?

   c. Next, we asked "I feel comfortable that Amazon shares my personal and online activity data, in an anonymised format, with third parties?" and "I feel comfortable that the University shares my personal and online activity data, in an anonymised format, with third parties?" You stated [insert student's answer]. Could you tell me more about your responses to these questions?

   d. Next, we asked "I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online?" and "I feel comfortable that the University offers me specific benefits in exchange for tracking me online?" You stated [insert student's response]. Could you tell me more about your responses to these questions?

   e. We asked "I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online and assures me that my data will not be shared with third parties?" and "I feel comfortable that the University offers me specific benefits in exchange for tracking me online and assures me that my data will not be shared with third parties?" You stated [insert student's answer]. Could you tell me more about your responses to these questions?

   f. We asked "I feel comfortable that Amazon offers me specific benefits in exchange for tracking me online on condition that my data will be shared with third parties?" and "I feel comfortable that the University offers me specific benefits in exchange for tracking me online on condition that my data will be shared with third parties?" You stated [insert student's answer]. Could you tell me more about your responses to these questions?

3. There are a number of ways that people define or think about privacy. Could you tell me what you think privacy is in the specific context where your data is used for the learning dashboard and to improve your learning?

Controlling use of data

4. Is there data that you would not want to be used in preparing the learning analytics dashboard and to improve your learning?

   a. **Probe**: Could you tell me more about why you would want to exclude some data from use?

   b. **Alternative**: Could you tell me more about why you would not want to exclude some data from use?

Benefits of the learning dashboard

5. Do you think there are benefits to you personally if you use the student learning dashboard?

   a) **Probe**: Could you tell me more about these benefits and why they are relevant to you?

   b) **Alternative**: Could you tell me more about why you think you get no benefits from using the learning dashboard?

Thank you and wrap up

6. Do you have anything to add that we have not talked about?

Appendix B.6   Mean, Standard Deviation: Online Privacy Questionnaire

General Caution	Mean	Std. dev.
Do you shred/burn your personal documents when you are disposing of them?	3.23	1.21
Do you hide your bank card PIN number when using cash machines / making purchases?	4.05	1.22
Do you only register for websites that have a privacy policy?	3.18	1.26
Do you read a website's privacy policy before you register your information?	2.07	1.16
Do you look for a privacy certification on a website before you register your information?	2.39	1.34
Do you read license agreements fully before you agree to them?	1.99	1.13

Technical protection		
Do you watch for ways to control what people send you online (such as check boxes that allow you to opt-in or opt-out of certain offers)?	3.39	1.13
Do you remove cookies?	2.59	1.08
Do you use a pop-up window blocker?	3.19	1.34
Do you check your computer for spyware?	2.65	1.28
Do you clear your browser history regularly?	2.96	1.23

Privacy concern		
In general, how concerned are you about your privacy while you are using the Internet?	3.43	.93
Are you concerned about online organisations not being who they claim they are?	3.34	1.04

Are you concerned that you are asked for too much personal information when you register or make online purchases?	3.63	.94
Are you concerned about online identity theft?	3.70	1.02
Are you concerned about people online not being who they say they are?	3.63	1.03
Are you concerned that information about you could be found on an old computer?	3.32	1.18
Are you concerned who might access your medical records electronically?	3.05	1.20
Are you concerned about people you do not know obtaining personal information about you from your online activities?	3.81	.93
Are you concerned that if you use your credit card to buy something on the Internet your credit card number will be obtained/intercepted by someone else?	3.96	.99
Are you concerned that if you use your credit card to buy something on the internet your card will be mischarged?	3.92	1.05

Appendix C   Study 3

Appendix C.1   Power Analysis
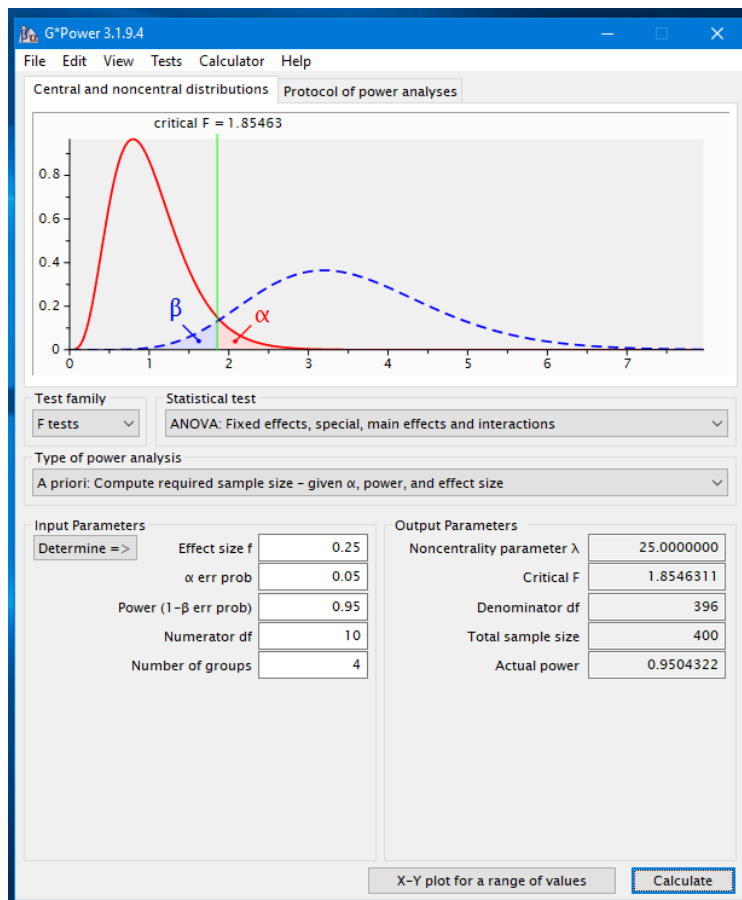
Appendix C.2   Study Instruments

Several questionnaires and scales are used. These are summarised as follows:

- Utility is measured by [1] utility of dashboard features and [2] perceived usefulness.
- Concern is measured by [3] concern about data use, [6] concern about privacy risks, and [8] privacy concern.
- Benefit perception is measured by [7] benefit perception.
- Acceptance of data use is measured by [4] sharing of data, and [5] data use preference.

1. Utility of learning analytics dashboard features (benefits)

Using a 5-point Likert scale (1 – not at all useful, 5 – very useful), participants were asked: "How useful are the following features for your studies?":

1. The feature to see your predicted performance
2. The feature to see your actual performance
3. The feature to receive personalised reading and activity recommendations

This question was adapted from work by Ponciano, Barbosa, Brasileiro, Brito and Andrade (2017).

2. **Perceived usefulness**

This question serves to measure students' satisfaction with the learning analytics dashboard and its features by considering their perception of its usefulness.

Using a 5-point Likert scale (1 - strongly disagree, 5 – strongly agree), participants were asked to indicate the extent to which they agreed or disagreed with the following statements:

1. Using learning analytics and the learning analytics dashboard would enhance my effectiveness in my studies
2. Using learning analytics and the learning analytics dashboard would improve my performance in my studies
3. I would find learning analytics and the learning analytics dashboard useful in my studies
4. Using learning analytics and the learning analytics dashboard in my studies would enhance my productivity

These questions were adapted from Arbaugh (2000).

3. **Concern about data use**

These questions (which were developed for Study 3) aimed to determine the extent to which participants were comfortable with the use of student data for learning analytics, to predict student performance, and to recommend learning resources. Using a 5-point Likert scale (1 - strongly disagree, 5 – strongly agree), participants were asked:

"Please indicate to what extent do you agree or disagree with the following statements? I would be concerned about my learning institution using student data ..."

1. ... for learning analytics
2. ... to predict students' performance
3. ... to recommend study resources

4. **Sharing of data**

The sharing of data questionnaire was adapted from work by Ifenthaler and Schumacher (2016). The questionnaire asked participants to identify specific information they were willing to share with learning analytics applications. The questionnaire has 28 items answered on a Thurstone scale (1 – Agree, 0 - Do not agree). The 28 items represent the types of data that can be shared with learning analytics applications. The questionnaire had a Cronbach's alpha of 0.74. Participants were also asked to share the reasons for the responses they gave to whether to share the specified data.

5. **Data use preferences**

This question was designed for Study 3. Participants were asked to express whether two categories of data (data about themselves and data about their activities on the learning platform) could be used for LA. Students' preferences were then used to derive a willingness score which indicated high, medium or low willingness to share data for learning analytics. Students who prefer that no data is used were assigned as low willingness to share. Those who prefer to share either data about themselves or data about their activities will be assigned a medium willingness to share score. Finally, those who prefer to share both data about themselves and their activities on the learning platform will be assigned a high willingness to share score. Participants were asked to opt-in rather than opt-out to encourage their active participation in the tasks for Study 3.

6. **Concern about privacy risks**

Participants were shown a description of possible privacy risks arising from the use of student data for predictive and prescriptive learning analytics. These privacy risks were identified using Solove's taxonomy of privacy harms (Solove, 2009). Using a 5-point Likert scale (1 – not at all concerned, 5 - extremely concerned), participants were asked: "Please indicate whether you are concerned with the possible risks associated with sharing data for learning analytics":

1. We will monitor what you and other students are doing on the online learning platform

2. Data that you and other students have provided to separate information systems at your learning institution (for example, during registration) will be combined to form a digital profile

3. This digital profile can be linked to the individual student

4. This information will be used to make decisions about you and other students, such as predicting your performance and giving you study recommendations.

7. Benefit perception

Using a 5-point Likert scale (1- strongly disagree, 5 – strongly agree), participants were asked to respond to the following question: "The following use of my data would be beneficial for me":

1. Personalised support to help me complete the course
2. Personalised recommendations of learning materials to improve my understanding

These questions were adapted from Naeini et al. (2017).

8. Privacy concern

Participants' privacy concerns were measured using the Internet Users Information Privacy Concerns (IUIPC) questionnaire by Malhotra, Kim and Agarwal (2004). Participants were asked: "Please indicate to what extent do you agree or disagree with the following statements":

1. Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared
2. Consumer control of personal information lies at the heart of consumer privacy
3. I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction
4. Companies seeking information online should disclose the way the data are collected, processed, and used
5. A good consumer online privacy policy should have a clear and conspicuous disclosure
6. It is very important to me that I am aware and knowledgeable about how my personal information will be used
7. It usually bothers me when online companies ask me for personal information
8. When online companies ask me for personal information, I sometimes think twice before providing it
9. It bothers me to give personal information to so many online companies
10. I am concerned that online companies are collecting too much personal information about me

Appendix C.3 Translating the Sharing Data Questionnaire into English

ORIGINAL	1: German to English	2: English to German	3: German to English	4: German to English	5: German to English
Bitte geben Sie an, ob Sie der Nutzung der folgenden Daten in einem Learning Analytics Tool zustimmen würden (ja/nein)	**Please indicate whether you would agree to the use of the following data in a Learning Analytics Tool (yes/no).**	Bitte kennzeichnen Sie, welche der folgenden Daten wir in Zustimmung mit Ihnen in einem Lernanalysewerkzeug verwenden koennen.	Please let us know whether you would agree to the following data being used by a Learning Analytics Tool (yes/no)	Please advise if you are happy for us to use the following data in a learning analytics tool. (Yes/No)	-
8 Testergebnisse zu individuellem Vorwissen	8 Test results of my previous knowledge	8 Testergebnisse meiner Vorkenntnisse	8 Test results concerning my prior knowledge	8 Test results regarding individual previous (subject) knowledge	8. **Test results of individual prior knowledge**
9 Aufzeichnung meiner Online-Nutzerpfade	9 Record of my online user pathways / the web pages I	9 Aufzeichnung zugegriffener Webseiten	9 Record of my online navigation	9 Keeping a record of my online pathways	9 **Record of my online user journey**

	have accessed				
10 Aufzeichnung meiner Onlinezeiten	10 Record of my online time	10 Aufzeichnung meiner Onlinezeit	10 Record of the times I spent online	10 Keeping a record of my time spent online	10 **Record of my time online**
11 Aufzeichnung meiner Downloads in der Lernumgebung	11 Record of my downloads in the learning environment	11 Aufzeichnung meiner heruntergeladenen Daten innerhalb des 'Learning Environment'	11 Record of downloads from the learning environment	11 Keeping a record of my downloads within the learning environment	11 **Record of my downloads within the learning environment**
12 Aufzeichnung meiner Forenbeiträge	12 Record of my posts in the forum	12 Aufzeichnung meiner Forumsbeitraege	12 Record of my Forum contribution	12 Keeping a record of my forum activity	12 **Record of my forum posts**
13 Inhaltsanalyse meiner Beiträge	13 Content analysis of my posts	13 Inhaltliche Untersuchung meiner Forumsbeitraege	13 **Content analysis of my contributions**	13 Content analysis of my contributions	
14 Testergebnisse zu Kompetenzen	14 Test results regarding competences	14 Testresultate in Bezug auf Faehigkeiten	14 Competency test results	14 Test results regarding competency	14 **Competency test results**

15 Abschlussnote der Schule	15 Final grade	15 Abschlussnote	15 Final grades achieved in school	15 Final school grades	15 **Final school grade**
16 Schulische Laufbahn	16 Schooling	16 Schulausbildung	16 School education pathway	16 **Time spent at school and the type of schools attended**	16 School career
17 Fragebogenergebnisse zur Motivation	17 Survey results regarding motivation	17 Umfrageergebnisse in Bezug auf Motivation	17 Responses to the motivation questionnaire	17 Results from the questionnaire for motivational purposes (or about motivation?)	17 **Survey results about motivation**
18 Fragebogenergebnisse zu Interessen	18 Survey results regarding interests	18 Umfrageergebnisse in Bezug auf Interessen/Hobbys	18 Responses to the interests questionnaire	18 Results from the questionnaire about personal interests	18 **Survey results about interests**
19 Fragebogenergebnisse zu Lernstrategien	19 Survey results regarding learning styles	19 Umfrageergebnisse in Bezug auf Lernstile	19 Responses to the learning strategies questionnaire	19 Results from of the questionnaire about learning strategies	19 **Survey results about learning strategies**

20 Fragebogenergebnisse zu Intelligenz	20 Survey results regarding intelligence	20 Umfrageergebnisse in Bezug auf Intelligenz	20 Responses to the intelligence questionnaire	20 Results from the questionnaire about intelligence	**20 Survey results about intelligence**
23 Vorherige akademische Leistungen	23 Previous academic performance	23 Vorherige Schulleistung	23 Prior academic achievements	23 Previous academic achievements	**23 Prior academic achievements**
24 Informationen zur Erwerbstätigkeit während des Studiums	**24 Information about employment during your studies**	24 Information ueber Anstellungen waehrend des Studiums	24 Information on your work commitments parallel to your studies	24 Employment information during (university) studies	24 Information about employment during studying
25 Informationen zur Kursbelegung an der Universität	25 Information about courses taken at the university	25 Information ueber studierte Universitaetskurse	25 Information on all (your) module registrations with the university	25 Information on courses studied at university	**25 Information about selected/chosen university modules/courses**
26 Ergebnisse von Selbsttests	26 Results of self-	26 Selbstbewertungstest	**26 Results of**	26 Results of self-	26 Results of self-

(Verständnistest zu einem zu lesenden Text)	assessment tests (text comprehension), for example, quizzes	ergebnisse (Textverstaendnis), zum Beispiel Quizfragen	**self-assessment test to evaluate your understanding of set reading**	testing (tests for understanding of written texts)	assessment tests (Reading comprehension test about a text)
27 Beteiligung an Diskussionen in Foren der Universität	27 Engagement with discussions in forums at the university	27 Teilnahme an Universitaetsforumsdiskussionen	27 Contributions to University Forums	27 Taking part in study forums/university forums	27 **Participation in discussion forums at the university**
28 Ausleihstatistik der Bibliothek	28 **Library loan statistics**	28 Bibliotheksausleihstatistiken	28 Library lending statistics	28 Library borrowing statistics	28 Library loan statistic. Or: loan statistic of library

The final list of items translated to English (as used in Study 3):

1. Name

2. Address

3. Email address

4. Date of birth

5. Information about your family (marital status, names of children, etc)

6. Medical information

7. Income

8. Test results of individual prior knowledge

9. Record of my online user journey

10. Record of my time online

11. Record of my downloads within the learning environment

12. Record of my forum posts

14. Competency test results

15. Final school grade

16. Please answer "no" for us to test the reliability of the responses

17. Time spent in school and the types of schools attended

18. Survey results about motivation

19. Survey results about interests

20. Survey results about learning strategies

21. Survey results about intelligence

22. Analysis of external data (e.g., search terms, personal profiles in social media, etc)

23. Parents' education level

24. Prior academic achievements

25. Information about employment during your studies

26. Information about selected/chosen university modules/courses

27. Results of self-assessment tests to evaluate your understanding of set reading

28. Participation in discussion forums at the university

29. Library loan statistics

Appendix C.4   Pilot Test Participants

Stage of pilot test	Pilot tester ID	Gender	Course
1	1	F	Language studies
	2	F	Professional learning
	3	M	Computing
	4	F	Educational technology
	5	F	Psychology
2	6	F	Educational technology
	7	F	Public leadership and social enterprise
	8	F	Education
	9	F	Organisational learning
	10	F	Language studies
	11	F	Educational technology
	12	M	Educational technology
	13	F	Health and wellbeing
	14	M	Educational technology
	15	F	Educational technology
	16	F	Mathematics
	17	F	Psychology
	18	F	Politics and international studies
	19	F	Learning analytics
	20	F	Computing
	21	F	Language studies

Appendix C.5   Factor Analysis of the Sharing Data Scale

Scree Plot
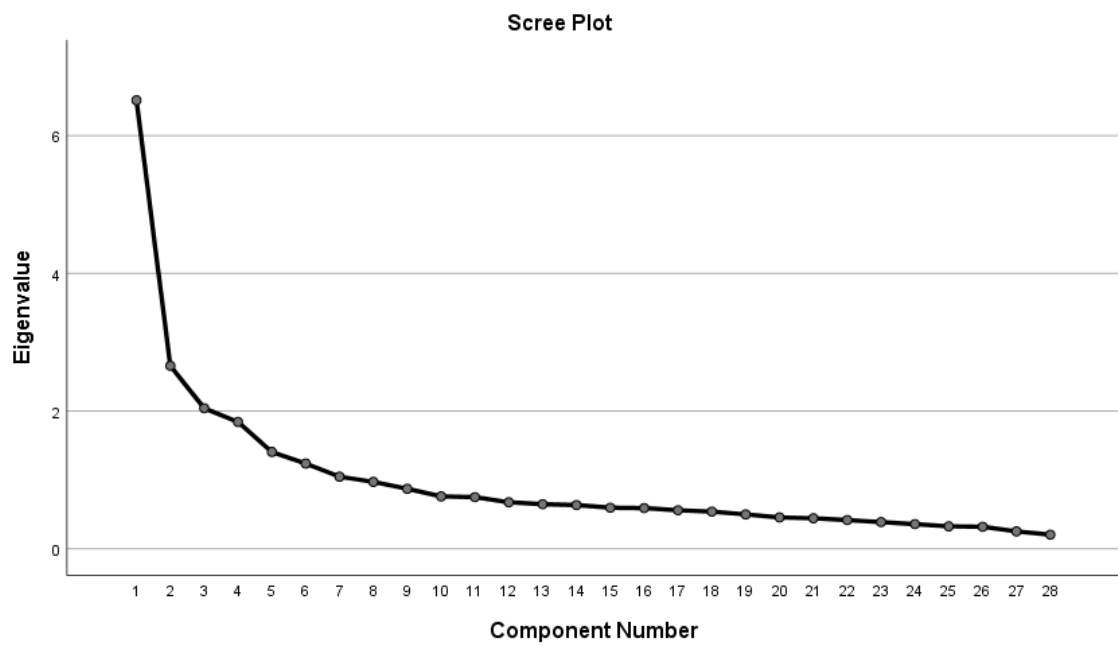
Results of the factor analysis identifying items loaded onto corresponding factors

	1	2	3
Learning strategies	0.693		
Motivation	0.678		
Final school grade	0.662		
Interests	0.656		
Prior academic achievements	0.647		
Self-assessment results on set reading	0.639		
Information about chosen university modules/courses	0.625		
Competency test results	0.560		
Intelligence	0.544		
Tests of individual prior knowledge	0.500		
Time spent in school and types of schools	0.409		
Information about employment during studies	0.334		
Information about family		0.732	
Medical information		0.669	
Address		0.661	
Income		0.636	
Date of birth		0.558	
Parents' education level		0.545	
Email		0.459	
Name		0.375	
Forum posts			0.734
Content analysis of posts			0.731
Time online			0.639
Downloads			0.556
Online user journey			0.551
University discussion forums			0.533
Library loan statistics			0.372
External data e.g., social media			0.307

Appendix C.6   Themes and Codes from Participants' Data Use Preferences

Theme/Categories	Definition. Statements where participants:
Reservations about institutional use of student data	**were hesitant about the use of data for learning analytics due to ethical and privacy considerations. Participants also focused on whether and how student data is (not) used and expressed a desire for boundaries in the use of student data.**
Ethical and privacy considerations	raised ethical and privacy issues, in many cases as concerns due to the use of student data, including how it could possibly have a negative impact on students.
Expectations about whether and how student data is (not) used	provided reasons for hesitating to support institutional use of student data
Tension between understanding data use and discomfort with data use	demonstrated a conflict where they claimed to understand the need for data to be used, but at the same time, expressed discomfort or uncertainty. They seemed to be of two minds about the use of data for learning analytics.
Desire for boundaries or separation	indicated in their responses a desire for a boundary or separation between aspects such as their school and private life.
Support for institutional use of student data	**expressed support for institutional use of student data, while stating conditions they thought need to be considered alongside this use of data. The statements also demonstrated that some participants made a trade-off in the use of their data in exchange for a service from the institution.**
Reasons to support institutional data use	expressed why they supported institutional use of student data, focusing on the data that is used, the experience for students and others, and the learning analytics tools.
Expectations students have of the learning institution	stated the expectations they had of what the learning institutional would do with, or how it would handle, student data.
Trade-offs in data use for benefits	stated the trade-offs they were making in supporting institutional use of student data for learning analytics.
Further information participants needed	**indicated what further information they would need in order to determine their data use preference.**

Change in opinion	**indicated that they had had a change of opinion having gone through the study. Their change in opinion was both negative and positive, however the negative changes were dominant.**
Other	**Statements unrelated to the motivation for participants' data use preferences**

Appendix C.7   Themes and Codes from Participants' Data Sharing Preferences

Theme/Code	Definition
(Mis)match in students' expectations of data use	**participants state their expectations about how their data would (not) be used thereby highlighting (mis)matches between participants' expectations and how the data is actually used**
Irrelevance of (personal) data selected for use	participants state that the data used is not relevant for the stated purposes or context
Personal nature of data used	participants indicate that the data used is personal and goes beyond what they expect
Separation of academic and personal records	participants distinguish between use of academic data and use of personal data and indicate a preference for use of academic data in an academic context
Use of other people's data	participants indicate they do not expect data about other people such as their parents to be used.
LA may constrain student support or options (labelling)	participants indicate that the use of data for LA will lead to them receiving limited or unhelpful options
Student's preference for independent work	participant expressed a preference to retain the students' agency in the work they do
Concerns and questions about effect of data use	**Statements related to concerns participants expressed including ethical and privacy considerations.**

Concerns	participants indicate concern about possible negative impact of data use on students.
Privacy considerations	participants express their privacy concerns and preference for privacy
Crossing a boundary	participants express a preference for a boundary or a separation, indicating that data use may be crossing a boundary they have in their expectations about how data is used.
Student data as an incomplete snapshot	participants point to the data collected as showing only a partial picture of the students' learning-related activities.
Acceptance of data use	**Statements where participants acknowledged the relevance of some of the data used for learning analytics and agreed to data use while indicating a number of conditions to be met.**
Supportive of data use	participants indicated support for use of some but not all data.
Agree to data use under several conditions	participants indicated the conditions they thought should be met
University already has some data, or the data is already public	participants seemed resigned to data use, acknowledging that the university already collects some of the data items or they are considered already public knowledge.
Trust college or university	participants stated they trusted the university to use student data appropriately.
Other	**Statements unrelated to the motivation for participants' data sharing preferences**

Appendix C.8   Reasons for (Not) Adding Excluded Features

Code name	Number of codes
Reasons for not adding features	**179**
Not beneficial	32
Features are sufficient	29
Prefers not to share (additional) data	15
Predictions can demotivate students	15
Better for students to have responsibility for their learning	13
Not necessary	10
Knows what they need to study	7
Feels has shared enough data	7
Predicted outcomes not valuable	6
Value my privacy	5
Feels privacy invasive	5
Get recommendations from staff or fellow students	5
The data is personal	4
Additional features cause privacy concerns	4
Would confuse student	4
Features not relevant	2
Does not want activities monitored	2
Some factors contributing to predicted grade excluded	1
Benefit without compromising personal information	1
Reduce student's responsibility for their learning	1
Does not want data analysed	1

Where is my information going	1
Needs more information about recipient to decide whether to share	1
Doesn't agree with dashboard	1
Lose control over privacy	1
Personal data may cause bias in staff	1
Not worth the invasion of privacy	1
Do not trust university to secure data	1
Other	2
Reasons for Adding Features	**131**
Further learning resources and recommendations useful	88
Exceptions while adding features	*20*
Does not want predicted grades	8
Demotivate students	4
Make students over-confident	2
Worried about giving up personal data	2
May be biased	1
Seems like a lot of data	1
Why the data already shared cannot be used	1
Concerned about sharing data	1
Preferences while adding features	*6*
Protect student data	1
Uses reliable technology	1
Do not use personally identifiable information	1
Restrict to education features	1

Limit amount of data used	1
Data collected carefully	1
Curious	4
Comfortable providing more data	2
Suggests more features	1
Best choice	1
Questioning the 'all or nothing'	1
Opportunity to make changes later	1
Motivate student to work hard	1
Other	4

Appendix D   Study 4

Appendix D.1   Participant Recruitment Email

Thank you for taking part in a study I conducted between September and October 2020. The study was titled Students' Preferences for the Use of Data to Improve Teaching and Learning in Higher Education.

As you indicated your interest in taking part in the follow-up interviews, I would like to invite you to discuss your responses to the survey questions.

I plan to run the interviews on [date of interviews] between [time of interviews]. The interviews will be held remotely (e.g., over Skype or Google Hangouts). You will receive £5 for your time. This will be paid as a bonus on the Prolific platform.

Please use this link (http://whenisgood.net/kkwc8nz) to pick the times that are suitable for you. If there's a schedule conflict, I'll request you to pick another time slot.

Please enter your name in the NAME field and your ProlificID and email address in the Comments box before you send your response. I will then be in touch to confirm your time slot.

Thank you.

Appendix D.2   Sample Interview Schedule

Interview schedule

Introduce researcher and research study

Good morning. Thank you for making time for today's interview. My name is Maina.

1. What name would you like me to call you during the interview?

Confirm consent for audio recording

Before we begin, please confirm whether you consent to me recording our interview today?

If participant consents

Begin audio recording

I have started the audio recording. We will now move on to the study questions.

If participant does not consent

I have noted that you do not consent to my recording of our interview today. We will now move on to the study questions and I will note down your responses during the interview.

Background

This interview is a follow-up to the study you took part in in September. The study focused on the use of student data to improve teaching and learning in higher education. I would like to focus specifically on the data use preferences you expressed in the study, compare them to how other participants responded, and discuss *whether there is anything that would cause you to change your response*.

Study-related questions

At the start of the study you indicated *that you prefer only data about your activities on the learning platform* was used.

1. Why did you say that?

At the end of the study you indicated *the same preference*.

2. Why did you say that?

On sharing specific data items, I have reminded you of the items you were willing and not willing to share.

Yes	No

1. Test results of individual prior knowledge Record of my downloads within the learning environment
2. Record of my forum posts
3. Content analysis of my posts
4. Competency test results
5. Final school grade
6. Survey results about motivation
7. Survey results about interests
8. Survey results about learning strategies
9. Survey results about intelligence
10. Prior academic achievements
11. Information about selected/chosen university modules/courses
12. Results of self-assessment tests to evaluate your understanding of set reading
3. Name
4. Address
5. Email address
6. Date of birth
7. Information about your family (marital status, names of children, etc)
8. Medical information
9. Income
10. Record of my online user journey
11. Record of my time online
12. Time spent in school and the types of schools attended
13. Analysis of external data (e.g., search terms, personal profiles in social media, etc)
14. Parents' education level
15. Information about employment during your studies
16. Participation in discussion forums at the university
17. Library loan statistics

18. Why did you indicate that?

For participants who did not change their data use preference

I noted that you did not change your data use preference

19. Why is that?

Scenario

When Robinson started college in 2017, he knew the career he wanted… he wanted to major in nursing. "I always knew I had a passion for helping people," he said. Biology had been his best subject in high school. During his first year, Robinson earned a B average. But the university was closely tracking his academic performance and knew from 10 years of student records that Robinson wasn't likely to make the cut for the nursing program. In meetings with his academic

adviser during the second semester of his first year, Robinson said he learned that though his GPA was solid, the school's computer algorithm saw trouble. … Robinson's file was showing yellow, a sign that his plan to go into nursing was risky. His adviser told Robinson he would need at least a 3.5 GPA – a high B+ average – to be admitted into the nursing program. Robinson's grades were a little short of that. An adviser at the university sometimes steers students like Robinson into another healthcare major that accepts students with lower grades. "They stay on track and we still get them to graduation," ... Robinson was pointed toward a related major, respiratory therapy, that he likes.

20. Is there anything in the scenario that's made you think 'That's good for Robinson'?
21. Is there anything in the scenario that's made you think 'That's not good for Robinson'?
22. If you were Robinson and this was your learning institution, how would you respond to your data being used in this way?

23. What do you think about the following benefits described in the study?

Our use of your data entails the following benefits:

*We can offer you **personalised support** to help you complete the course, including nudging to submit assignments or follow-up from the student support team. We can also provide you with **personalised recommendations** of learning materials that can improve your understanding of the course material.*

24. Is there anything that I've not touched on that would cause you to change your data use preferences
    a. to share more data?
    b. to share less data?
    c. to share no data?

Wrap up

25. Is there anything I have not asked you that you want to tell me?
26. Please confirm your prolific ID so I can compensate you for taking part in this interview.

Thank you very much for your time and for your feedback.

Appendix D.3   Themes and Codes for Study 4

Theme	Definition of theme: Statements …
Theme 1: Data use practices and students' expectations are not aligned (38.6% of codes, n = 96)	**that demonstrate a mismatch between LA data use practices and students' expectations.**
Negative effect on student	that refer to how data use can impact students in a negative way.
Desire for anonymity and separation	that show students' preference for their identity to be separate from other data.
Desire for privacy	where students express a preference for privacy.
Knowledge of data use caused discomfort	that show students' discomfort as they become aware of the data that was used.
Restrictions on data sharing	that show students wanting to limit who their data is shared with.
Taking responsibility for their learning	that show students' preference to be responsible for their learning.
Data use practices and students' expectations are aligned (26.9% of codes, n = 67)	**that demonstrate a match between LA data use practices and students' expectations.**
Benefit self, other students, and lecturers	that show students motivated by how data use would benefit them and others.
Data use met expectations	that show situations where data use as described in the study met students' expectations.
Opportunities for institutional learning about students' data use expectations (28.5% of codes, n = 71)	**that highlight higher educational institutions potential lack of knowledge about students' data use preferences.**
Control, consent, and data use	that focus on students' providing consent to the data that is used.

Influence of external contexts	that show how data use in other contexts can influence students' data use preferences in the learning context.
Transparency contributes to willingness to share data	that show how higher education institutions being transparent about how they use student data can lead to students being willing to share their data.
Value of benefits offered	where students assess the value of the benefits on offer.
Pre-set ideas on what to share	that show students had already determined the data they would share.
Students are resigned to institutional data use practices (3.61% of codes, n = 9)	**where students seemed resigned to the institutional data use practices emphasising their lack of agency to bring about any changes.**
Public information anyway	that show students' view that the data is already known to others.
Institution already has the data	that show students' view that the higher education institution already has their data.
Monitoring already ongoing	that show students' view that they are already being monitored.
Aware and accepting of possible harms	that show student acknowledging and discounting possible harms to data use.
Knowledge disparities contribute to students' misperceptions (2.4% of codes, n = 6)	**that surface students' lack of knowledge about how their data is used which results in students' misperceptions about institutional data use practices.**
It's harmless	that show the perception that data use is harmless for the student.
Only benefiting the institution	that show the perception that only the institution stands to benefit from data use.
Lack of clarity in determining the data to share	that show that students are unclear on what data they should share.