

Student Perceptions of Privacy in Learning Analytics: A Quantitative Study of Norwegian Students

Stian Botnevik

May 14, 2021



Master's thesis
Information Science
Department of Information Science and Media Studies

Abstract

Learning Analytics is concerned with the collection, analysis, and reporting of student data for the goals of optimizing learning experiences, and improving learning outcomes. The research field is young and unexplored in the Norwegian setting. This research explores student perceptions of privacy in relation to learning analytics.

This research was carried out within the frames of quantitative research through a questionnaire that gathered empirical data. The questionnaire was developed with inspiration from secondary analysis as well as a literature review. The main feature of the questionnaire was that it utilized privacy principles as indicators to measure privacy perceptions. The questionnaire was implemented at the University of Bergen, obtaining 394 student responses. The responses were analyzed using descriptive statistics.

The results indicate that students perceive privacy in general, and in relation to learning analytics, as highly important. The results also indicate that most students have a information privacy centered privacy understanding. Data security and consent are the most important privacy principles for students, based on findings from this research.

Student expectations of learning analytics have also been explored. The findings indicate that the majority of students accept the use of learning analytics at their higher education institution. According to the results of the questionnaire, students desire the following benefits of learning analytics: improved learning outcomes, improved courses, and improved grades.

Acknowledgements

I would like to thank all of the 403 students who answered my questionnaire – special thanks to you. I also want to thank Datatilsynet, represented by Dag Mostuen Grytli, for the inspiring meetings and access to data of their latest privacy questionnaire. Last but not least, I would like to thank my supervisors Prof. Barbara Wasson and Dr. Mohammad Khalil. Thank you for your assistance in the distribution of the questionnaire, which was immensely helpful. I also want to thank you for our constructive video calls and your helpful feedback on my drafts.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research questions	2
1.3	Structure of the thesis	2
2	Theory	3
2.1	What is privacy?	3
2.2	Solove's privacy theory	3
2.3	Personvern	6
2.4	Privacy principles	6
2.5	Summary	7
3	Literature review	8
3.1	Learning analytics	8
3.1.1	Review method	8
3.1.2	Defining literature for the research field	9
3.1.3	Insights from networks	14
3.1.4	Summary	16
3.2	Privacy in learning analytics	19
3.2.1	Review method	19
3.2.2	Privacy principles	19
3.2.3	Privacy concepts	24
3.2.4	Summary	26
3.3	Student perceptions of privacy principles in learning analytics	27
3.3.1	Review method	27
3.3.2	Research methodology	30
3.3.3	Explored privacy principles	32
3.3.4	Important principles for students	33
3.3.5	Other insights	35
3.4	Summary	36
4	Methodology	37
4.1	Desk research	37
4.2	Addressing the research questions	37
4.3	Quantitative research	38
4.4	Survey research	40
4.5	Questionnaire	40
4.5.1	Questionnaire development	42
4.5.2	Expert review	42
4.5.3	Reliability and validity	42
4.6	Secondary analysis	43
4.6.1	Limitations	44
4.7	Method of analysis	44
4.7.1	Descriptive statistics	44
4.8	Summary	45
5	Questionnaire development	46
5.1	Design	46
5.1.1	Anonymity	46
5.1.2	Cover letter	46
5.1.3	Introductory information	47
5.1.4	Language	47
5.1.5	Length	47
5.1.6	Clear presentation	47
5.1.7	Question formats	48
5.1.8	Likert scale	49
5.2	Formulation of questions	51

5.3	Justification of questions	52
5.3.1	Question 1 Student confirmation	52
5.3.2	Question 2 Privacy understanding	52
5.3.3	Question 3 Privacy importance	52
5.3.4	Question 4 Privacy statements	53
5.3.5	Question 5 Learning analytics desirability	53
5.3.6	Question 6 Privacy principles	53
5.3.7	Question 7 Terms and Conditions	55
5.3.8	Question 8 Pressure to consent	55
5.3.9	Question 9 Learning analytics services and benefits	55
5.4	Implementation	56
5.4.1	Distribution	56
5.5	Limitations	57
5.5.1	General limitations	57
5.5.2	Specific limitations	57
5.6	Summary	58
6	Results	59
6.1	Secondary analysis results	59
6.1.1	Respondents	59
6.1.2	Privacy importance	59
6.1.3	Feeling of lack of control and powerlessness	59
6.2	Questionnaire results	59
6.2.1	Respondents	61
6.2.2	Privacy understanding and importance	61
6.2.3	Feeling of control and powerlessness	63
6.2.4	Privacy self-management and expert-management	63
6.2.5	Learning analytics desirability	65
6.2.6	Importance of privacy principles	67
6.2.7	Terms & Conditions and the feeling of pressure to consent	69
7	Discussion	70
7.1	Privacy principles in learning analytics	70
7.2	Privacy perception	71
7.3	Privacy priorities	76
7.4	Learning analytics desirability	77
7.5	Summary	78
8	Conclusion	79
8.1	Contributions	80
8.2	Limitations	80
8.3	Future work	80
9	Appendices	84
A	Communication with NSD	84
B	Questionnaire	85
C	Large figures	91

List of Figures

1	Traditional privacy conceptualization (a), Solove’s privacy conceptualization (b). (Inspired by Solove, 2009)	4
2	Three views on balancing.	5
3	Structure of the literature review.	8
4	Most influential learning analytics literature 2016–2020	10
5	Siemens’ learning analytics model (Siemens, 2013)	13
6	Most popular academic journals for learning analytics literature.	15
7	Subject categories of learning analytics literature.	16
8	Tag cloud of keywords in learning analytics literature	17
9	Illustrative overview of learning analytics components.	18
10	Modified PRISMA 2009 flowchart diagram (Moher et al., 2009)	28
11	Categories of questions in questions asked to students.	33
12	Distribution of privacy principles found in the literature.	34
13	Distribution of privacy principles important to students.	34
14	Likert scale question on a big (a), a medium (b), and a small (c) screen.	48
15	Questionnaire question as displayed on a big (a) and a small (b) screen.	48
16	Balanced (a) and unbalanced (b) Likert scale.	49
17	Norwegian’s degree of concern with privacy.	60
18	Norwegian’s feelings of limited control and powerlessness online.	60
19	Heatmap of questionnaire responses.	62
20	Results Q2: Definition students finds most descriptive of their privacy-view.	62
21	Results Q3: How important students find privacy.	63
22	Results Q4 (a/b): Students’ feeling of lack of control over their personal information in relation to their higher education institution (A). Students’ feeling of powerlessness over their personal information stored at their higher education institution (B).	64
23	Results Q4 (c/d): Students’ desire for privacy self-management (C). Students’ desire for privacy expert-management (D).	65
24	Results Q9: What information students would trade for learning analytics services/benefits	66
25	Results Q5: Learning analytics desirability among students.	66
26	Results Q6: Importance of privacy principles for students (A)	68
27	Results Q7: Likelihood of reading Terms and Conditions agreements.	69
28	Results Q8: Students’ feeling of pressure to consent	69

List of Tables

1	Individual and collective interest in the interest-model (NOU:1997:19, 1997).	6
2	Search queries used for section 3.2	19
3	Screening process for literature in Section Privacy in learning analytics	20
4	Privacy principles in privacy-related learning analytics literature	20
5	Dimensions of ownership and access according to Siemens (Siemens, 2013)	23
6	Search queries used to find literature, in Section 3.3	27
7	Literature included in Section 3.3	29
8	Key information from the literature.	31
9	Questions that are the topic of the secondary analysis.	44
10	Privacy principles asked about in the questionnaire, and their explanations	54
11	Privacy principles important for learning analytics. (A, B): literature review findings, (C): questionnaire findings	70

1 Introduction

We live in the information age. Every day we interact with technologies that generate large amounts of data based on our activity, from the smartwatches we wear, to the library cards we use. We generate data while at home, at work, in school, and probably while asleep as well. What is this data used for? And do we get a say in its use?

Private companies and government agencies are eager to take advantage of the vast amounts of data generated by every-day digital interactions. This type of data is often called *data exhaust* (Technopedia, 2021). It is generated in large amounts, but usually not taken advantage of. It harbors exciting potential, as its rewards are not yet fully explored. Much effort is placed in exploring ways of exploiting data exhaust in all areas of society. One such area is higher education.

Higher education institutions are attempting to make value of data exhaust generated by their students. A research field has formed around creating digital solutions, based on student data, that will improve student learning and learning outcomes. Services that utilizes student-generated-data to improve learning outcomes are known as *learning analytics*. The formal definition of learning analytics, as defined by the Society for Learning Analytics Research (SoLAR)¹, is: “*The measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.*” (SoLAR, 2021). Learning analytics is a young research field, with the first conference held in 2011. As a research field, learning analytics draws on disciplines such as computer science, education, learning science, sociology, law, psychology, and more, as will be explored later. Learning analytics is a fast growing field, and it is becoming more relevant than ever today, as more focus is put on digital schooling, because of the pandemic.

1.1 Motivation

As the learning analytics research field is relatively new, much of learning analytics is still unmapped territory. The research field is also rapidly changing, as it maneuvers in the fast-paced environment of modern technology. Three important but relatively scarcely explored aspects of learning analytics will be the focus of this thesis.

The Norwegian context. Few learning analytics applications have been deployed in Norway. The potential for Norwegian higher education institutions is large, as Norway scores high on digitization². Universities and colleges already use Learning Management Systems (LMS) (such as Canvas³ or Blackboard⁴) to organize their courses, which already aggregate a lot of student data. The aggregated data can be valuable information for exploring student behavioral patterns, risk of failing courses, or other metrics that can enhance students learning and learning outcomes. Furthermore, due to the increase in digital schooling, caused by the pandemic, the potential for information-collection and analysis is greater than ever. Technologically there are no limitations on Norwegian higher education institutions that prevents them from implementing learning analytics. But, technological limitations are not the only gatekeeper for learning analytics. Privacy is of great importance when handling aggregated personal information, as the nature of learning analytics entails.

Privacy. Privacy has received some attention in the learning analytics literature. A search on Web of Science (WoS) reveals that 93/1566⁵ learning analytics articles hosted on WoS, are related to privacy. Privacy is a relevant topic when discussing any type of systems that handles personal information, and learning analytics is no exception. As will be seen in the literature review, privacy is highlighted as a challenge for the learning analytics research field. Despite this, relatively little research is done on the topic. Particularly, little research is done on student perceptions of privacy.

Student perceptions. Students are one of the main stakeholders in learning analytics (Greller and Drachsler, 2012). As stakeholders, they should be included in the development process of learning an-

¹<https://www.solaresearch.org/about>, accessed: 13.05.2021

²According to the EU Index DESI https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60009

³<https://www.instructure.com/canvas>, accessed: 13.05.2021

⁴<https://www.blackboard.com/teaching-learning/learning-management/blackboard-learn>, accessed: 13.05.2021

⁵comparing search results from (“Learning analytics”) and (“Learning analytics” AND “Privacy”). Search is on “Topic” (title, abstract, keywords). Performed: 11.03.2021

alytics applications. This includes decisions regarding privacy. For instance, student privacy concerns should be respected when developing learning analytics applications. There exists very little research that explores student perceptions of privacy for learning analytics. Such research is needed in order to include students in decision processes regarding privacy. To know what students care about when it comes to privacy, they need to be able to voice their opinions. This is the main focus in this research. The intersection between learning analytics, privacy, and Norwegian students can be complex to explore. This is partly because of the lack of experience with learning analytics in Norway. This can pose a challenge, as it is difficult to get empirical data on concepts foreign to the group that is being studied. Furthermore, privacy, as a more general concept, is also challenging to study because of its elusive nature.

The following research explores this intersection between Norwegian students, learning analytics and privacy. The results can give useful insights into the potential for learning analytics in Norway. The research scouts the grounds for learning analytics in Norway and reports important privacy insights that learning analytics implementors will need to keep in mind when considering learning analytics for Norwegian higher education institutions. Later, it will be explored if students want learning analytics and what privacy facets they think are important when faced with learning analytics. The research has a focus on exploring *what* student perceptions of privacy is, and how this affects learning analytics, with less focus on *why* student perceptions of privacy are what they are.

As learning analytics is under-established in Norway, a window of opportunity to shape its future development is open. Particularly development regarding privacy will be explored in this research. There is also room to make an impact on the learning analytics research field, as privacy is not exhaustively explored, thus the relationship between learning analytics and privacy can be impacted before it solidifies into convention. Especially important for this research is the student-perspective, this will be represented through a questionnaire, collecting student perceptions of privacy and their desirability for learning analytics.

This research aims to fill a gap in learning analytics research, as little to none prior research is done on students' relationship to privacy for learning analytics in Norway.

1.2 Research questions

The main goal of this research is to explore student privacy perceptions at a higher education institution in Norway, namely the University of Bergen, and identify the consequences this could have for the use of learning analytics in higher education in Norway. The following research questions of this thesis, are:

1. What privacy principles are most relevant for learning analytics?
2. How do Norwegian students perceive privacy in general, and in relation to learning analytics?
3. What privacy priorities do Norwegian students have for learning analytics?
4. What learning analytics services and benefits are acceptable to Norwegian students?

1.3 Structure of the thesis

The thesis is organized as follows. In Chapter 2, theory is introduced. A discussion of the conceptualization of privacy, and the use of privacy principles to describe privacy, is central in this chapter. Chapter 3 is a literature review where learning analytics, privacy, and student perceptions of these two topics are explored. Chapter 4 describes the research methods used in this research. Secondary analysis, a questionnaire, and descriptive statistics are the main topics of this chapter. Chapter 5 explains the development and implementation of a questionnaire used in this research, to gather students perception of privacy in learning analytics. Chapter 6 displays the descriptive statistics of secondary analysis and questionnaire data. Chapter 7 is the Discussion, revolving around interpreting the findings from the previous results chapter, and answering the research questions. Chapter 8 is the conclusion, where contributions, limitations and future work will be described.

2 Theory

This chapter explores the relationship between different privacy theories and conceptualization. The chapter has focus on privacy conceptualizations, the Norwegian setting, and privacy principles. As privacy is such an elusive and complex concept, a clarification is required before proceeding. The intent of this chapter is to introduce theories to understand privacy and privacy related phenomena. One of the major challenges when working with privacy is that there exists many different conceptions, making it difficult to synthesize and discuss privacy as a unity.

The foremost goal of discussing the following theories are to provide anchoring-points for the understanding of privacy, that will be utilized in this research. The theories are to a smaller degree used as explanation mechanisms of the results generated by the research. The reason for this is that the this research attempts to uncover *what* students think about privacy in relation to learning analytics, and not *why*. The theories are then intended to clarify how privacy is understood, as it has consequences for how it is explored. Hence, it is important to state which privacy conceptualization is put to ground for the following research, as it permeates all aspects of this research.

Solove's privacy theory from 2009 (Solove, 2009) is central for this purpose. His theory builds upon a nontraditional privacy conceptualization that can bridge the gap between different privacy concepts. The Norwegian term *personvern* is also highly relevant. Traditionally there has been a discrepancy between *personvern* and *privacy*. It will be demonstrated that by using Solove's privacy theory, the discrepancy can be limited.

2.1 What is privacy?

There will probably never be agreement on what privacy is. Most people have an opinion about privacy, on whether it is important or not, and what they consider privacy to hold and entail, but these opinions can vary vastly. There exist many conceptualizations of privacy, but there is not consensus.

A conceptualization⁶ is a way of thinking about or envisioning a concept inside the mind. Different people view the same concepts in different ways. For example, an Information Scientist probably views the World Wide Web as a connected network of documents, while others might view it as similar to an online shopping mall; a collection of services located in one place. These mental images help us understand and navigate complex concepts. Conceptualizations are subjective, but some are more widespread than others. Conceptualization matters in the same way that words matter. "Words equals power", as the Norwegian idiom⁷ goes ("Ord er makt"). Likewise conceptualization contains certain power as it entails how we think about concepts, which affects our actions. An example of this is morals. We base our actions on what is considered good. What is considered good depends on how we think about (conceptualize) *goodness*. This makes how we *think* about privacy important. Solove (Solove, 2009) has an arguably fresh take on how privacy should be conceptualized in his privacy theory.

2.2 Solove's privacy theory

Solove describes his privacy theory in his 2009 book: *Understanding privacy* (Solove, 2009). Here he explores the history of privacy theories and documents their flaws and shortcomings, before introducing his own theory of privacy. The common denominator of the shortcomings of older theories is that they are either too broad, too narrow, or too vague (or all of these combined)(Solove, 2009). Solove also criticizes the foundation on which these theories are built. He argues that their conceptualization is flawed. The traditional privacy conceptualization assumes that privacy has a core that can be distilled as an *essence of privacy*. Still, no privacy theory precisely identifies this core. All attempts are imprecise, hence Solove's criticism of broadness, narrowness, or vagueness. Solove argues that perhaps privacy does not have a core. With this backdrop he launches his own privacy conceptualization, arguing that privacy needs to be viewed as a collection of distinct concepts that are all *related* instead of having a common core.

Two different conceptualization are shown in Figure 1 as a visual aid to understanding the difference between the traditional conceptualization and Solove's. This illustration is derived from Solove's ideas of them: a) Shows the traditional conceptualization of privacy, where all aspects of privacy (letters a-z) are

⁶<https://www.oed.com/view/Entry/38150>, accessed: 13.05.2021

⁷perhaps more of a "floskel" (empty phrase)

connected to a common core. b) Shows Solove’s conceptualization. Privacy is defined by loose boundaries capturing different privacy aspects (letters) that are distinct but related (shown by proximity). Some overlap can also exist between the different privacy aspects.

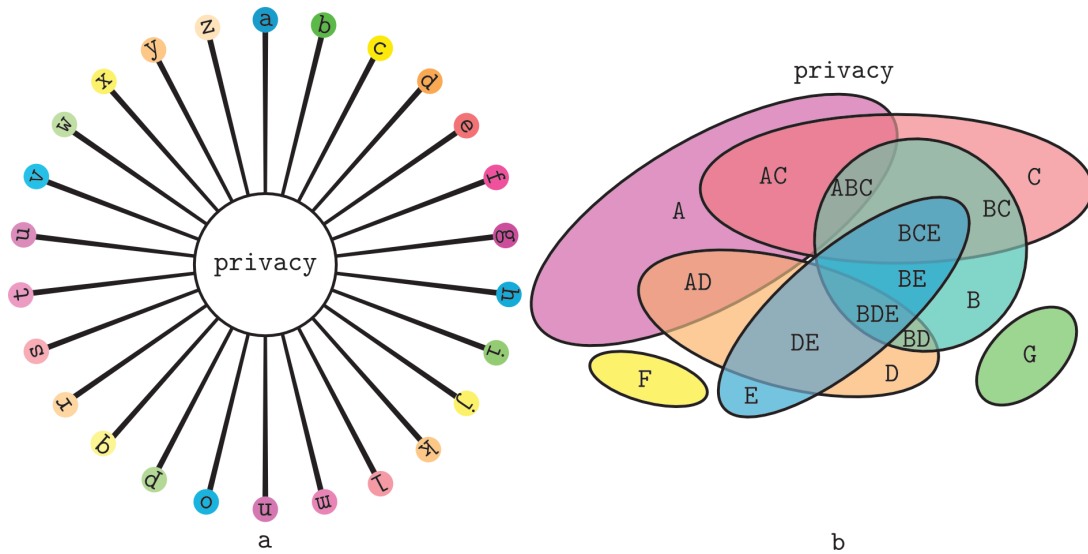


Figure 1: Traditional privacy conceptualization (a), Solove’s privacy conceptualization (b). (Inspired by Solove, 2009)

To give insight into the contents of the traditional conceptualization, two traditional privacy theories are explored. These examples were collected from Solove’s book *Understanding Privacy* (Solove, 2009).

Right to privacy. The theory of *the right to privacy*, also known as *the right to be let alone* (Solove, 2009) is a highly influential privacy theory. This theory has a descriptive title, which describes its main contents. It is similar to the Norwegian expression *privatlivets fred*. This term translates to: having an opportunity of privacy, privacy in this relation, is a life outside the professional and public sphere (Tor Guttu et al., 2021). The heritage of this theory is also included in the *Convention for the Protection of Human Rights and Fundamental Freedoms*⁸ as well being essential in The European Union’s General Data Protection Regulation⁹ (GDPR). Both of these legislative frameworks make up the foundation of privacy legislation in Norway (Regjeringen.no, 2019).

Personhood. *Privacy as personhood* theories are related to personal integrity (Solove, 2009). Personhood¹⁰ is defined as the quality of being an individual. In privacy theories with a focus on personhood; privacy safeguards personhood (Solove, 2009). Privacy as personhood theories are viewed by Solove as supplementary to other privacy theories, focusing on assessing the value of privacy (Solove, 2009). Some view personhood as a variation of autonomy, which is a concept usually kept separate from privacy. While there is however undoubtedly overlap between privacy and autonomy, they are still considered independent concepts.

When it comes to Solove’s privacy theory, he is inspired by pragmatism (Solove, 2009). The goal of his theory is pragmatic, as its intent is to aid the concrete making of law and policy, in a way that treats privacy in a balanced matter. He does not focus on theoretical challenges of privacy, but instead builds his theory upon empirical evidence of privacy issues. These are mostly in the form of court cases. Due to this pragmatic approach, Solove’s privacy theory can be viewed as a tool. A tool that can be used to understand privacy in a balanced way and aid in the solution of privacy challenges.

Balancing privacy against other interests is a central aspect of Solove’s theory (Solove, 2009). He is a believer in balancing as a juridical doctrine, and argues that balancing is fruitful when done properly. Balancing is done properly when the weights have the proper value, which means that the value of privacy, as well as the value of competing interests, needs to be asserted correctly.

⁸Article 8: https://www.echr.coe.int/documents/convention_eng.pdf

⁹<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>

¹⁰<https://www.oed.com/view/Entry/141503>

This relationship is visualized in Figure 2 as a visual aid. The value of privacy is represented as spheres, and other interests as cubes. a) Shows the simplified process of balancing privacy against other interests. b) Shows more complex depiction of the balancing of privacy against other interests. c) Shows a highly complex balancing of privacy against other interests, which is probably the most realistic depiction.

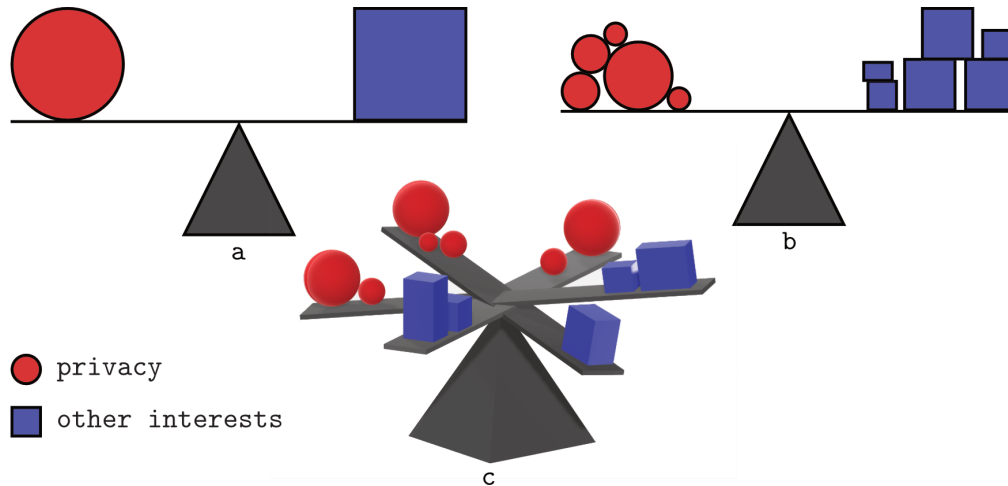


Figure 2: Three views on balancing.

To be able to perform the balancing, the value of privacy needs to be asserted. How is this done in Solove's theory? The value of privacy is, for Solove, social, not highly contextual, possibly culturally dependent, and most importantly, measured by the harms it prevents (Solove, 2009). Thus, the value of privacy is the negation of the privacy harms it prevents.

An example is given to explain his logic: X is a privacy harm causing Z amounts of harm. P is a privacy measure that removes or prevents the amounts of harm (Z) caused by X . The privacy measure P is worth the removal/prevention of Z amounts of privacy harm.

In this way, privacy problems and the harms they cause are central to Solove's theory. Privacy harms assess the value of privacy that is used when balancing privacy against other interests (Solove, 2009).

Privacy harms is caused by privacy problems. A taxonomy of privacy problems is provided by Solove, to systematize privacy problems in categories, and identify the harms they cause (Solove, 2009). The harms are then used to determine the value of privacy for balancing purposes. The privacy challenges are grouped together by the action that causes privacy problems, into the following categories: *Information collection*; *Information processing*; *Information dissemination*; and, *Invasions*.

Not all of contents of the taxonomy are equally relevant for this research. The nature of learning analytics entails vulnerability to some privacy problems, but not all. Of interest, in particular, is which privacy challenges apply specifically to learning analytics. An example of this is information aggregation. An example of a challenge that is not relevant, is, blackmail. While blackmail could happen based on information from a learning analytics application, it is not something that is intrinsic to learning analytics applications. Blackmail can happen anywhere, based on information collected from all types of systems. Thus, as the research focuses on privacy and learning analytics, the categories of information collection and information processing are of particular interest.

Solove's privacy theory is not definitive (Solove, 2009). He argues that privacy is a concept in perpetual motion, that evolves over time. His theory is flexible as to accommodate for future challenges of privacy. This makes it suitable to be used for a variation of purposes, and not only as originally intended; to aid the creation of policy and law.

2.3 Personvern

This section will explore the relationship between Solove’s privacy conceptualization and the Norwegian terms *personvern* and *personopplysningsvern*. *Personvern* is a special Norwegian term (NOU:2009:1, 2009) that is closely related to privacy. In 2009 a Norwegian expert committee formulated a definition of *personvern* (directly translated):

Personvern is about safeguarding personal integrity; safeguarding individuals’ opportunities for privacy, self-determination (autonomy) and self-expression. (NOU:2009:1, 2009, p.32).

Accompanying this definition is the term *personopplysningsvern* (NOU:2009:1, 2009), which is concerned with rules and standards regarding handling of personal information. The goal of these rules and standards is to safeguard *personvern*. A direct translation of how *personopplysningsvern* is described, is:

Personopplysningsvern is about rules and standards for the processing of personal information that have the protection of privacy as the main goal. The purpose of the rules is to ensure individuals’ overview and control over the treatment of information about themselves. (NOU:2009:1, 2009, p.32)

This term is similar to the English term *information privacy*. A *personopplysning* is any piece of information that can be connected to an individual (Datatilsynet, 2019). The translated equivalent of *personopplysning*, *personal information*, will be used in this research to describe what is collected from students by learning analytics, as this term is descriptive of its contents and is more relatable to the Norwegian setting, compared to other terms frequently used in relation to learning analytics, such as, *student data*.

Combined, *personvern* and *personopplysningsvern* capture various privacy related concepts. *Personvern* encapsules multiple concepts known from traditional conceptualizations of privacy. For example, *personhood*, protection of personality’s integrity and the right to privacy, also known as the right to be let alone. *Personvern* is also related to concepts outside the privacy sphere through the mandate to safeguard these, including *autonomy* and *freedom of expression*.

Norwegian *personvern* is not a theory in itself, but it is built upon the interest-model, which can be considered a privacy theory. Modern *personvern* (from 1970) is influenced by the *interest-model* (*interessemodellen*) (NOU:1997:19, 1997; NOU:2009:1, 2009). The interest-model comprise of seven interests, see Table 1, related to *personvern* which have evolved together with society’s progressions on other fronts (NOU:1997:19, 1997). The interests are divided in *individual* and *collective interests*. The individual interests are related to information privacy, that is regarding control and management of personal information (NOU:1997:19, 1997). The collective interests are related to information control on a group level (NOU:1997:19, 1997). The interests are described in Table 1.

Table 1: Individual and collective interest in the interest-model (NOU:1997:19, 1997).

Individual	Collective
Discretion (Diskresjon)	Citizen-friendly administration (Borgervennlig forvaltning)
Completeness (Fullstendighet)	Robust society (Robust samfunn)
Access (Innsyn)	Limited surveillance level (Begrenset overvåkningsnivå)
Right to privacy (Privatlivets fred)	

If these interests contradict each other, the twists are solved by balancing the interests against each other (NOU:1997:19, 1997). Balancing is also the juridical doctrine on which Norwegian law is built (NOU:1997:19, 1997). This entails that privacy interests are balanced against other interests. Parallels can be drawn from this to Solove’s privacy theory, where balancing is a central part (Solove, 2009).

2.4 Privacy principles

Dividing privacy into smaller pieces and anchoring them in principles can make privacy more manageable. This is done in both theories discussed above, although in different ways. Solove splits privacy into *privacy challenges* and uses these to navigate the privacy terrain, while *personvern* is based on a division

of privacy into *core values*. The latter is maybe the most familiar to people in general as it is prevalent in all aspects of society. Core values and principles are found everywhere and used as cognitive centering points for handling complex subjects.

The use of principles is also prevalent in the privacy sphere. An example of this is the GDPR which has the following core principles: Lawfulness, fairness and transparency; Purpose limitations; Data minimization; Accuracy; Storage limitations; and Integrity and confidentiality (GDPR, 2016). In the literature review (Chapter 3) other frameworks (principle collections) for privacy, related to learning analytics, will be described.

Dividing privacy into principles is helpful for working with privacy, as it becomes more manageable. This makes it possible to examine smaller parts of privacy, without having to account for privacy as a whole. Examining privacy as a collective unity is difficult as it contains a large number of different facets, where some can be contradicting. This is the root of older privacy conceptualizations' problems, according to Solove (Solove, 2009), namely that fitting all privacy related aspects into one term is difficult or possibly, impossible.

Privacy principles will be focused throughout this research. Privacy principles will be the anchoring points for discussions, literature, results, and implications. Privacy principles are also an important part of the methodology (Chapter 4), as it will be used as indicators to measure the importance of privacy among students.

The use of privacy principles to measure privacy fits into Solove's privacy theory with some slight modifications. The main modification is how the value of privacy is assessed. Solove argues in favor of using *privacy harms* to assess the value of privacy, and this works well when having a large source of empirical data at hand, such as court cases. This is not equally helpful when empirical data is lacking, as is the case for *student perceptions of privacy in learning analytics, in Norway*. Empirical data will be collected using a questionnaire and thus, privacy principles are used to assess the value of privacy, instead of *privacy harms*. The value of privacy is determined based on how many respondents find the individual privacy principles important. This also tailors privacy importance to the specific setting (students at a Norwegian higher education institution), as the students dictate the importance of privacy. This will be taken into consideration when designing the questionnaire (see Chapter 5).

2.5 Summary

Personvern and Solove's privacy theory have a lot of similarities. Both operate in the same domain and describe many of the same concepts. Their difference lies in how they assess the value of privacy. Solove gives privacy its value based on what *privacy harms* it prevents, while personvern is more straightforward, as it uses *core values* to assess its value.

The theoretical foundation for this research draws on Solove's privacy conceptualization, with a focus on privacy principles. The rest of his theory is not actively utilized, as it is outside the scope of this research. Solove's privacy conceptualization allows for the co-existence of all privacy principles without conflicts, and the focus on privacy principles allows for a suitable way to measure privacy importance.

Solove's process of balancing, which is also shared by personvern, is a large part of his theory, but outside the scope of this research. It can be interesting to balance privacy with other interests, but at a later stage. Balancing needs to be done when a complete picture of privacy and learning analytics is mapped. Privacy can then be balanced against the benefits of learning analytics when considering implementing learning analytics applications in higher education institutions. This is as mentioned outside of the scope of the following research and will not be done.

Privacy is always in motion, as pointed out by Solove (Solove, 2009). Every time we research privacy we add to the pool of knowledge and privacy evolves. This research will not be able to cover the whole domain of privacy, and this is not the goal either. The goal of this research is to explore student privacy perceptions at a higher education institution in Norway and identify the consequences this could have for the use of learning analytics in higher education in Norway.

3 Literature review

The following literature review is divided into three parts. The structure of the review can be visualized with an funnel-shape (Figure 3). The goal of this literature review is to establish a background of learning analytics, privacy in learning analytics, and student perceptions of privacy in learning analytics.

A collection of questions guide the literature review. “*What is learning analytics?*”; an overarching question that aids in finding the essence of learning analytics. “*How is privacy discussed in learning analytics?*”, and “*What privacy principles are important in learning analytics?*”. These questions help map the privacy debate in learning analytics. “*What privacy principles are important for students?*”, and “*What is students’ view on privacy in learning analytics?*”. These questions aid exploration of research done on the student perspective of privacy in learning analytics. But first, an exploration of learning analytics by examining the most influential learning analytics literature.

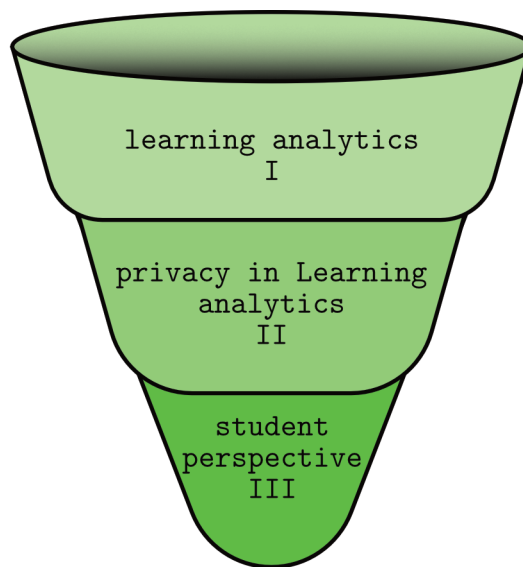


Figure 3: Structure of the literature review.

3.1 Learning analytics

To answer the guiding question: “*What is learning analytics*”, the most influential literature of learning analytics will be explored. The most influential literature or *seminal works* can be found based on citation-count. Although this gives an indication of the general impact a piece of literature have had, not the impact on a specific research field, as is desired for this analysis. To find the most influential literature in the learning analytics research field, one needs to look at what pieces of literature are the most cited by *learning analytics authors*. This is slightly different from finding the most cited learning analytics literature. To find out which pieces of literature are the most cited by other learning analytics authors, Complex Network Analysis (Zinoviev, 2018) was utilized. After the literature has been identified, their contents will be described and discussed.

3.1.1 Review method

Web of Science was searched with the query: “*TS= (‘Learning Analytics’ OR ‘learning analytics’)*” to get all the results categorized as learning analytics literature. The search was performed on the 12th of November 2020. The search yielded a total of 1426 results. The results were then divided by year, focusing on the last five years. The metadata for the papers were downloaded. The most important information was the reference lists of the literature. The result of this process was a dataset composed of all learning analytics literature that was hosted on Web of Science, including all of their references.

The dataset was loaded into the Science of Science (Sci2) Tool (Sci Team, 2009). Sci2 Tool was used

to generate networks from the dataset. A *paper citation network*¹¹ was generated for each year between 2016 and 2020. The networks were then loaded into Gephi (Bastian, Heymann, and Jacomy, 2009), which was used to analyze and manipulate the networks.

Each paper is represented as a node in the directed networks. The nodes are rated by how many outgoing edges they have to other nodes in their network. Outgoing edges represent a reference. The node that is the *source* of the edge, has been referenced by the *target* of the edge. The size of the nodes was made to reflect how many times they have been referenced. The nodes that have been referenced just once, were removed. The top three most referenced nodes for each year was given a number to reflect their position. Some years had a tie for the third most referenced node. In such cases all candidates were included. Each paper were color coded and lines were drawn from the legend to the networks for readability.

The final result of the network analysis, was an illustration of the learning analytics literature that have been referenced, most, by other learning analytics authors. The illustration shows literature divided by year, and is presented in Figure 4. The literature covered in the illustration will be the foundation of this part of the literature review.

3.1.2 Defining literature for the research field

Ferguson explores, in her highly influential paper, the history of learning analytics (Ferguson, 2012). The paper can be viewed as an account of learning analytics' origin story. Ferguson separates learning analytics from other related fields of research. She also identifies driving factors behind learning analytics. In addition to this, she addresses future challenges for learning analytics.

The driving factors for learning analytics are three unsolved *challenges*, according to Ferguson (Ferguson, 2012). One challenge is: mining value from big sets of learner data. Ferguson names this the *technical challenge*. Another is the challenge of improving online learning, which she dubs the *educational challenge*. The last is the challenge of optimizing students results, which she terms the *political/economic challenge*. These three challenges make up the driving forces behind analytics research, according to Ferguson. The challenges have at different times become motivational factors behind learning analytics, they have had varying importance and have been given varying amounts of attention during the (short) history of learning analytics.

Ferguson (Ferguson, 2012) describes different directions learning analytics has taken since the origin of the research field. She describes a shift in focus, from a focus on the technical aspect of generating learner data, to a focus more centered on the learner itself. Ferguson argues that after 2010 learning analytics distinguished itself as a research field. It separates itself from Academic analytics and cultivates a focus on the second of the driving analytics challenges: the educational challenge. The other two challenges, technical challenge and political/economic challenge are addressed by Educational data mining and Academic analytics respectively. Although there is still overlap, as Ferguson argues.

Ferguson points out four tasks that learning analytics needs to overcome in future work. The tasks she has for the future are (Ferguson, 2012):

- (1) Build strong connections with the learning sciences;
- (2) Develop methods of working with a wide range of datasets in order to optimize learning environments;
- (3) Focus on the perspectives of learners;
- (4) Develop and apply a clear set of ethical guidelines.

Ferguson together with Shum (Shum and Ferguson, 2012) (influential in 2016) addresses the 1st of learning analytics' future tasks. This elaborate paper details the concept of Social Learning Analytics (SLA). Social Learning Analytics is launched as a new and distinct learning analytics direction that has tight connections with learning theory. Greller and Drachsler (Greller and Drachsler, 2012) also recognize having a relation with learning theories as one of the key questions of learning analytics.

While exploring a generic learning analytics framework, Greller and Drachler highlights Ferguson's second future challenge. Greller and Drachler discusses the limitations caused by the difficulty of working with heterogeneous datasets. Their framework is intended as a guide for deploying learning analytics applications. Greller and Drachler also announces six pillars of learning analytics. These pillars need to

¹¹Inbuilt feature in Sci2.

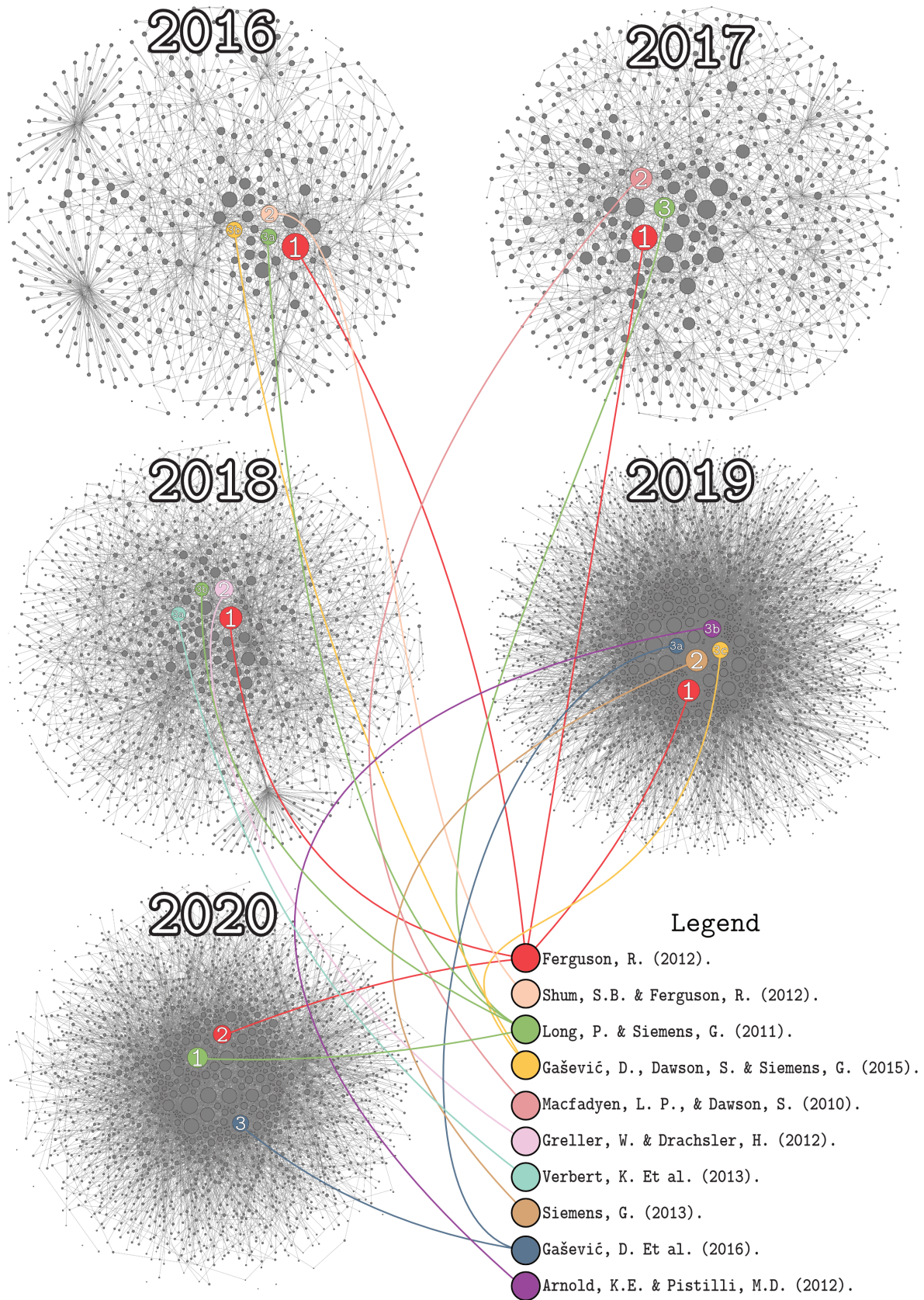


Figure 4: Most influential learning analytics literature 2016–2020

be addressed in the design of learning analytics applications. The six pillars, called critical dimensions in

the paper, are: stakeholders, objective, data, instruments, external limitations and internal limitations (Greller and Drachler, 2012). These make up the core dimensions of learning analytics, according to Greller and Drachler. These pillars resonate well with the narrative of Ferguson (Ferguson, 2012), when it comes to the potentials, opportunities, and challenges of learning analytics.

Greller and Drachler (Greller and Drachler, 2012) do not provide a clear-cut solution for the 2nd challenge put forward by Ferguson. This challenge remains largely unsolved in the present day as well, at least a universal solution to the problem of working with heterogeneous datasets. This is not a learning analytics specific challenge and can be found in a large selection of research fields, most notably perhaps in Big Data and Machine Learning.

The paper by Gašević et al. (Gašević et al., 2016) (influential in 2019/2020) can possibly be interpreted as an indirect critique of Greller and Drachler's (Greller and Drachler, 2012) idea of having a general approach to learning analytics applications. The title summarizes their point: "learning analytics should not promote one size fits all". The authors warn of the practice of using generalizes models for learning analytics, and stresses the importance of customizing the applications to the data and courses they are applied in (Gašević et al., 2016). Greller and Drachler's work sketch out a foundation all learning analytics applications should have, while the paper by Gašević et al. warns of practices on a much more specific level, not within the scope of Greller and Drachler's early work.

The 3rd and 4th of Ferguson's challenges are the most relevant for this research. In the 3rd challenge, Ferguson argues that we need to focus on the perspectives of learners (Ferguson, 2012). By focusing on the learners, learning analytics can provide meaningful experiences for the learners that operates at a deeper level that just improving grades, is Ferguson's key point. To achieve this, Ferguson argues that personalized visual analytics, that are easy to comprehend, is necessary. This stipulates the future development of learning analytics dashboards that will become more popular as the research field evolves.

Gašević et al. (Gašević, Dawson, and Siemens, 2015) reminds the learning analytics community about this learner-focus in their influential paper titled: "Let's not forget: learning analytics are about learning" (influential in 2016/2019). This paper is similar to others described in this review, in terms of giving an historical account of learning analytics as well as pointing out future challenges the research field faces. The authors highlight the importance that learning analytics applications should propel educational research and how education is practiced. They prescribe the same solution as Ferguson (Ferguson, 2012): that learning analytics needs to build upon learning research.

In the 4th challenge, Ferguson (Ferguson, 2012), calls for a clear set of ethical guidelines for learning analytics, and, mentions that data ownership is a pressing question. How to handle consent is also a question that Ferguson raises, as is the question of being able to opt-out. Privacy is here viewed as a subcategory of ethics. This emphasizes the ethical grounding of privacy, putting it into a larger context. This challenge and privacy for learning analytics is the most relevant topic for the following research and this topic is dedicated a whole section later in the review.

The focus on learners is crystallizing itself as a core value of learning analytics. Learning analytics is centered around the learner, and decisions made for learning analytics should reflect this position. Based on this student-centric view, learning analytics can be characterized as; analytics concerned with facilitating better learning for learners.

The Society for Learning Analytics Research (SoLAR) defines learning analytics as:

learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (Ferguson, 2012; SoLAR, 2021).

This definition, combined with the identification of motivating factors and future tasks of learning analytics, courtesy of Ferguson Ferguson, 2012, a synthesis of what the learning analytics research field is comprised of, emerges. A picture is drawn of a research field that has the learner at the center and learners' interest are important to maintain. In learning analytics, different techniques and methods are used to further the interests of the learners. But there are also challenges in learning analytics that needs to be solved in order to not damage the central interests of the learners. One of these challenges are

related to privacy.

There are a lot of similarities between the key points of Ferguson, and, Siemens and Long's highly influential paper from 2011 (Siemens and Long, 2011). In this popular science style paper, the authors address the potential benefits of learning analytics. They also make an effort of separating learning analytics as a research field from Academic analysis, similar to what Ferguson does. Without using the same terms as Ferguson, Long and Siemens points to similar driving factors behind learning analytics.

Big data is mentioned as an important facilitator, and the missed opportunity of not extracting value from Big Data is one of the main arguments for putting increased efforts into learning analytics (Siemens and Long, 2011). This resonates well with Ferguson's technical challenge (Ferguson, 2012). Long and Siemens also talks about how the general technological development is a facilitating factor for learning analytics (Siemens and Long, 2011). Long and Siemens provides a wide range of examples of what value learning analytics will bring to higher education. The value learning analytics generates, benefits both learners and faculty. This demonstrates learning analytics' focus on the learner and optimizing learning. This is recognized in Ferguson's educational challenge, although Ferguson has a focus on online learning, while Long and Siemens seem to talk generally about learning; online learning included (Ferguson, 2012; Siemens and Long, 2011). The authors of the two papers agree that one of the next steps for learning analytics is to move away from the heavy reliance on Learning Management System (LMS) data (Ferguson, 2012; Siemens and Long, 2011). They seem to agree that moving beyond this can give rewarding results for learning analytics, and ultimately the learners.

There is a lot of consensus between these two papers, giving growing evidence of what the learning analytics research field is concerned with, what core values it has, and what the potential benefits it can provide.

Siemens (Siemens, 2013) supplies Ferguson's origin story of learning analytics in his seminal work. It is published a few months later than Ferguson's paper and cover much of the same ground. Siemens' paper is more detailed than Ferguson's, covering specific techniques that are used in learning analytics and provides a variety of figures describing internal learning analytics processes. In addition to providing a short historical account of contributions to the research field. Siemens provide a Learning Analytics Model (LAM) that can be viewed as a model organism of learning analytics applications. The model is reproduced in Figure 5.

This illustration sheds light on the learning analytics process. Here it is depicted where data is collected, how it is analyzed and what results it produces. Siemens Siemens, 2013 has produced a conceptual model of learning analytics that gives an overview of the whole process. As this is a model from 2013, new methods can have emerged, but the core processes are believed to be the same as depicted in this illustration.

Most notable from a privacy perspective on figure is the *collection & acquisition* step. This step entails data aggregation, which in itself can entail a lot of privacy challenges (Solove, 2009). The challenges it can introduce varies, but all have their origin in that the data becomes more condensed and *meaningful* when aggregated, as described by Solove.

Information-aggregation challenges can be, among others, related to: (1) context: people share different types of information dependent on the context, not necessarily aware that their information are connected to other pieces of information through aggregation (Solove, 2009). (2) Decision foundation: aggregated information about individuals are often used to make decisions, this is also one of the cornerstones of learning analytics. As this aggregated information does not always provide a representative picture of the individual it is related to, automatic decisions from a system can be made based on faulty information causing harm (Solove, 2009).

The distinction between the papers by Siemens (Siemens, 2013) and Ferguson (Ferguson, 2012) is the scope and depth. Ferguson gives more of an overview and a theoretical account of learning analytics, with a focus on the goals and core values of learning analytics (Ferguson, 2012). Siemens is more technically oriented, in the sense that he gives account for the tools, techniques and applications that learning analytics utilizes (Siemens, 2013). Both papers include mentions of the Course Signals Tool, as an example of a learning analytics application.

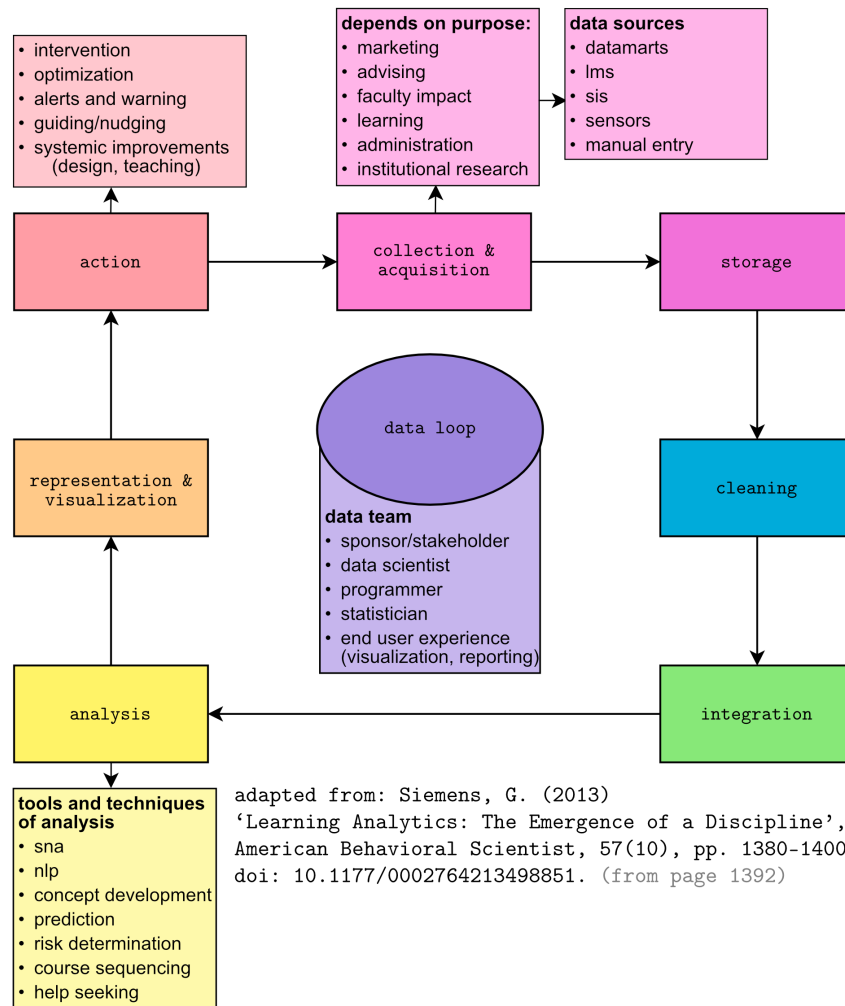


Figure 5: Siemens' learning analytics model (Siemens, 2013)

Arnold and Pistilli (Arnold and Pistilli, 2012) wrote about their Course Signals project in 2012 and as seen in Figure 4 was highly referenced by other learning analytics papers in 2019, and highly regarded in general as a successful learning analytics application (Ferguson, 2012).

Course Signals is a tool that is intended to promote student's success (Arnold and Pistilli, 2012). It uses different sources of learner data coupled with analytics and statistical methods to create an image of which of the students are at risk of failing courses. It also offers opportunities for instructors to give feedback to students through the Course Signals system as explained by the authors. The students are provided with a signal, like a traffic light, for each of their courses (that utilizes this system). The color of the light; green, yellow, or red, indicates how much at risk the students are to fail a class. The instructors are provided different means of reaching out to students to provide feedback.

The Course Signals application is an apparent epitome example of what learning analytics applications should be. It maintains multiple stakeholders' interests, it is designed to prevent students from failing courses, which is both in the higher education institution's interest as well as the students', and it uses diverse sources of data, which addresses one of the future challenges pointed out by Ferguson – Although they do not use any of Ferguson's suggested datatypes: mood data, biometric data, and mobile data (Ferguson, 2012). The main source of data in Course Signals seems to be LMS data (Arnold and Pistilli, 2012).

The use of LMS data have been criticized earlier by other fellow learning analytics researchers (Ferguson, 2012; Siemens, 2013). Although possibly improving students learning outcome further when using more sources of data, by doing so it also strengthens the privacy concerns related to aggregation.

Course Signals receives positive feedback in Arnold and Pistilli's own paper (Arnold and Pistilli, 2012). There is little mention of privacy concerns or ethical framework in their paper, so it hard to assess how well these challenges are addressed. The Course Signals system can be categorized as an early warning system, or a risk identification system which goes under Siemens *Trend analysis and prediction (application)* category, in his categorization of learning analytics approaches (Siemens, 2013). Macfadyen and Dawson's (Macfadyen and Dawson, 2010) (influential in 2017) early warning system, is also an example of such an application.

Learning analytics dashboards are not mentioned by Siemens (Siemens, 2013) nor Ferguson (Ferguson, 2012), suggesting it is a more recent phenomenon. Although not mentioning explicitly dashboards, both of these authors discuss visualization as a key quality. Learning analytics dashboards is another category of learning analytics application. Verbert et al. (Verbert et al., 2013) documents and evaluates multiple learning dashboards in their influential paper on the topic. Their paper is highly influential in 2018 and it impresses the importance of dashboards in the field of learning analytics. Learning analytics dashboards answers to Ferguson's 2nd challenge for the future of learning analytics, the challenge related to learners' perspective.

As related earlier, Ferguson argues that for learning analytics to be meaningful to learners, the analytics need to be personal and visualized in an explainable manner (Ferguson, 2012).

Dashboards can fulfill both criteria, although Ferguson seems to think that the summarization of previous assessments as a way of giving meaningful value for learners, is limited (Ferguson, 2012). LMS have proven to be a prominent source of learner data. This can also be deducted from Verbert et al. (Verbert et al., 2013) analysis which shows widespread use of LMS data. Macfadyen and Dawson also demonstrates the use of LMS data in their previously mentioned learning analytics application (Macfadyen and Dawson, 2010).

Returning to Siemens' seminal work (Siemens, 2013). After discussing different tools and applications, he addresses some future challenges of learning analytics. Data ownership is addressed as a future challenge in both Ferguson's (Ferguson, 2012) and Siemens' (Siemens, 2013) work, as well as in Long and Siemens' (Siemens and Long, 2011). Ferguson puts this under the category of developing an ethical framework while Siemens calls it a privacy challenge (Ferguson, 2012; Siemens, 2013).

Siemens discusses privacy as a challenge (Siemens, 2013), this is supported in Ferguson's paper, although Ferguson categorizes it as an ethical challenge (Ferguson, 2012). But it is not explicitly mentioned. Privacy is also briefly mentioned in Long and Siemens influential paper, as an unanswered question (Siemens and Long, 2011). Greller and Drachsler recognizes privacy (and ethics) as a challenge for learning analytics. They have it categorized under one of their six pillars of learning analytics. It is located under the pillar titled *External constraints* (Greller and Drachsler, 2012).

3.1.3 Insights from networks

A visual representation of the learning analytics research field is possible to produce using Complex Network Analysis tools. The same tools used to create Figure 4, are used for this task. Following are two networks and a tag-cloud based on the dataset from Web of Science, documented earlier. This dataset consists of 1426 learning analytics papers. These are all the learning analytics papers kept in Web of Science. This is by no means all learning analytics papers in existence, only the ones curated by WoS. The first network gives an overview of the most popular academic journals for learning analytics literature. The second network gives an overview of the subject categories of the literature. The tag cloud gives an overview of the most frequently used keywords used to describe the contents of learning analytics literature.

Figure 6 gives an overview of most popular journals for learning analytics literature. The most popular journals are *Computers in Human Behavior* (74), *Journal of Learning Analytics* (71) and *British Journal of Educational Technology* (59). *Computers in Human Behaviour* is concerned with the junction of computer-use and psychology¹². *Journal of Learning Analytics* is the official journal of Society for

¹²<https://www.journals.elsevier.com/computers-in-human-behavior>

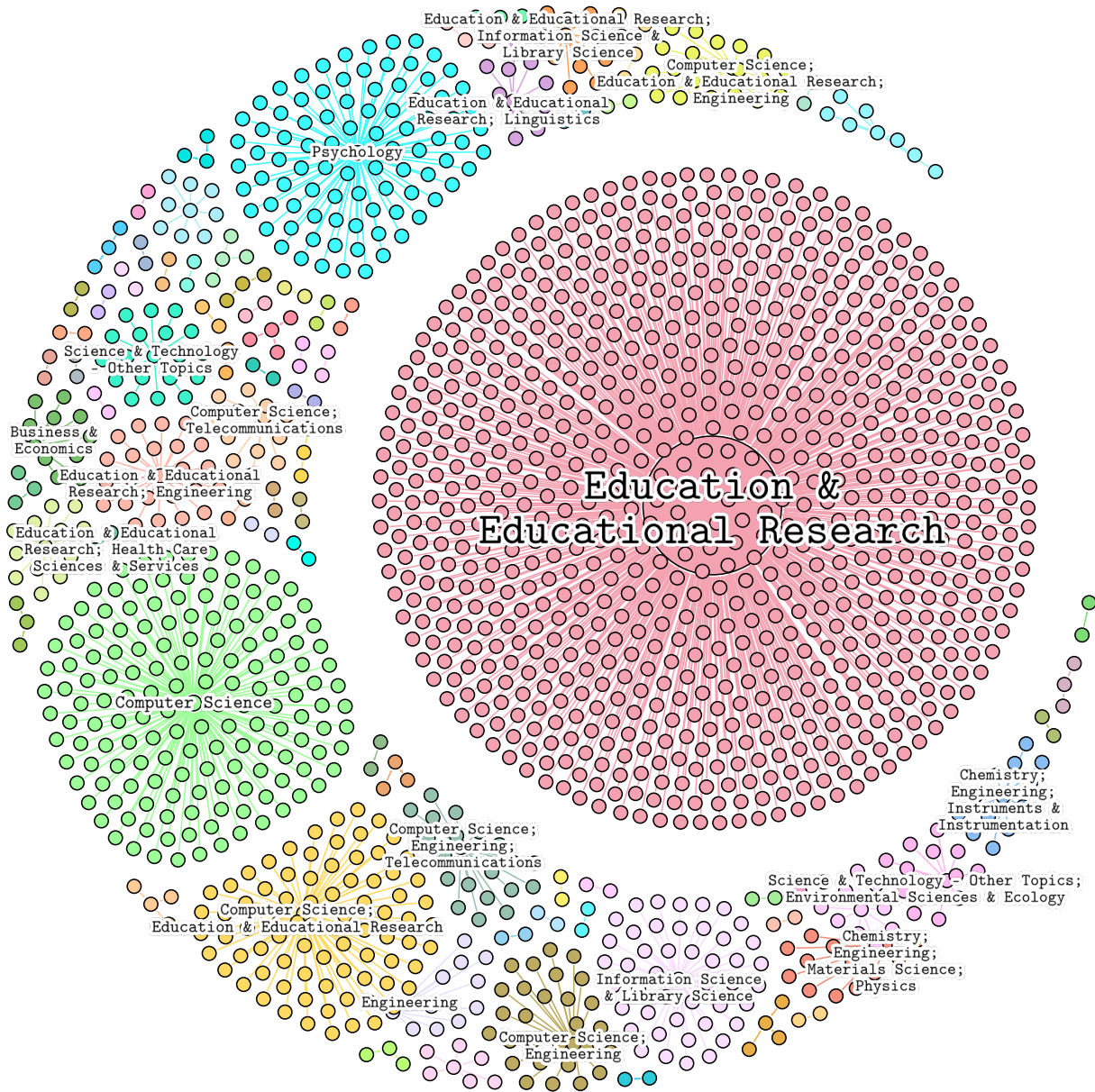


Figure 7: Subject categories of learning analytics literature.

3.1.4 Summary

The review of the most influential and defining learning analytics literature, for the last five years, sought to give an account of what learning analytics is. The core of the learning analytics research field is discovered through the synthesis of characteristics found in the selected literature. The core value of learning analytics seems to be the learner. Learning analytics is concerned with different methods of synthesizing data into valuable information that can benefit the learners, and those surrounding them. Ferguson (Ferguson, 2012) has addressed three motivating factors behind learning analytics. Other learning analytics authors seem to agree with these. These motivating factors determine in large part what characterizes learning analytics. Learning analytics use data about learners to improve their learning, and to a certain degree their faculty. This description covers Ferguson’s motivating factors, which she titled: *the technical challenge, the educational challenge, and the economic/political challenge*. Ferguson argues that learning analytics is concerned mostly with the educational challenge. Evidence from other literature supports this position. Still there exist multiple foci in learning analytics.

The specific methods that learning analytics applications utilize are not given a lot of attention in this review. It is a common understanding that methods play a large part in defining a research field. For this

research field?”. These are questions guiding the following review of literature.

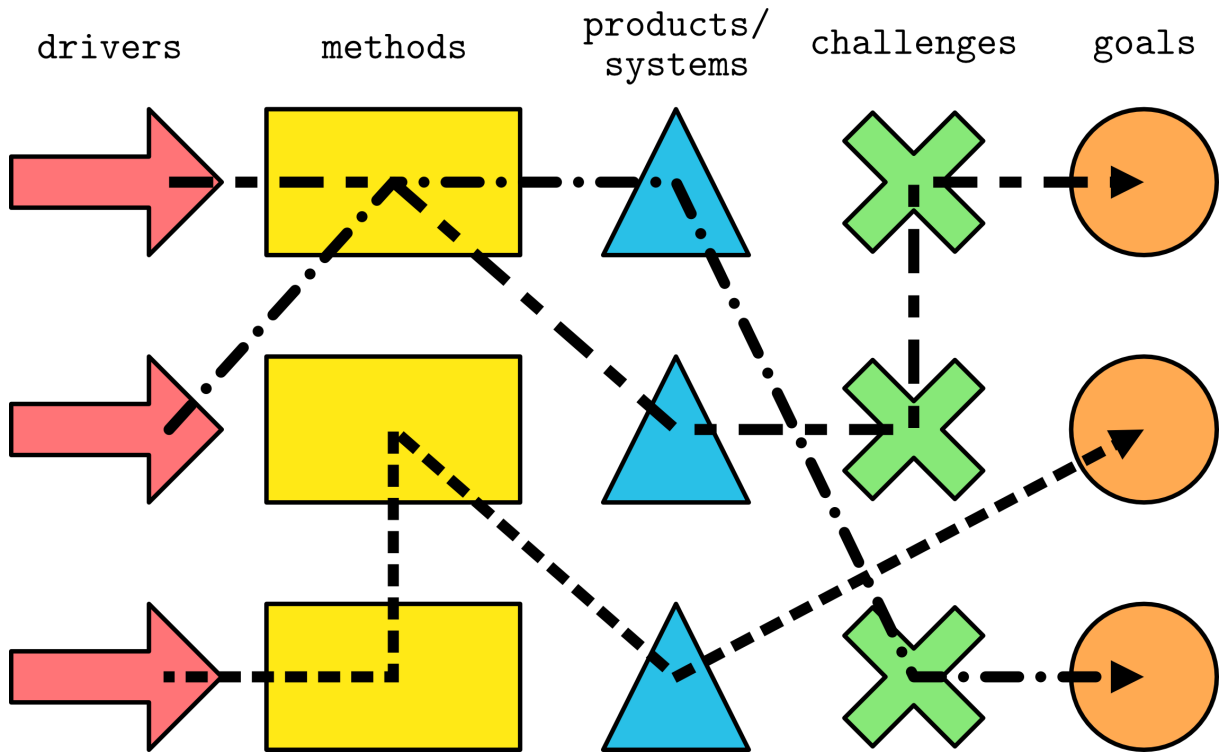


Figure 9: Illustrative overview of learning analytics components.

3.2 Privacy in learning analytics

This section will explore the privacy debate in the learning analytics field of research. The guiding questions, mentioned, will guide the review to explore privacy concepts and principles, as they are discussed in learning analytics literature. Mapping privacy principles and privacy concepts, is a way of quantifying the privacy debate in learning analytics, this gives an overview of what is important in the debate. This is based on the assumption that the most important topics gets the most attention from researchers. The process of gathering and screening literature is inspired by systematic literature review techniques as described by Bryman (Bryman, 2015), but the literature review in itself is not a systematic literature review, but narrative.

3.2.1 Review method

Web of Science and ACM Digital Library were queried for literature. The query-strings and number of results are described in Table 2. Web of Science was used as it is a high-quality curated database. ACM Digital Library was used as it contains conference proceedings for the biggest learning analytics conference; Learning Analytics and Knowledge (LAK)¹⁵.

Table 2: Search queries used for section 3.2

Date	Database	Search query	Results
09.09.20	Web of Science	TS=(“Learning Analytics”) AND TS=(“Privacy”)	80
23.03.20	ACM Digital Library	AND [All: “privacy”]	299

The screening process is described in Table 3. The screening process consisted of three steps: First a title and keywords screening, followed by an abstract screening, and lastly, a full paper screening. The steps have their own criteria that needs to be fulfilled in order for the paper to be considered for the review, these are also described in the table.

29 papers remained after the screening. These were considered for the review. It was too comprehensive to include detailed accounts of all 29 papers. Instead, all papers are included in an analysis of privacy principles and concepts. The privacy principles and concepts that are discussed in the papers have been extracted and quantified. The most discussed principles and concepts will be explored and select papers, from the pool of 29, will be used to highlight the principles and concepts.

3.2.2 Privacy principles

Privacy principles were *mined* from the literature in a systematic but informal way. There is no guarantee that all principles discussed in all of the literature have been recorded, but this does not weaken the purpose of the collection notably. The purpose was to get an insight into what is discussed when privacy is the topic of learning analytics literature. The result of the quantification is presented in Table 4.

The most frequently discussed privacy principles in the literature were: **consent**, **transparency**, **data ownership**, **access**, and **student control**, in descending order. These will be further explored. Consent in learning analytics is generally related to the practice of asking students their permission prior to collecting, processing, and analyzing their personal information. Transparency in learning analytics is related to higher education institutions’ openness surrounding how they collect, process and analyze student’s personal data. Data ownership is about who owns the data generated by students. Access is related to what stakeholders can view of the collected, processed, and analyzed data. Student control is related to how much power students should have over impactful processes of learning analytics, such as the collection, processing, and analysis of their data. It is also related to, if students should be able to decide themselves what data is collected of them, and what it is used for.

The explanations of the privacy principles and concepts vary between the different pieces of literature, making it challenging to compare the discussions without introducing ambiguity. Some authors are more explicit with what principles and concepts they discuss, while other are more vague. This introduces some

¹⁵<https://www.solaresearch.org/events/lak/>

Table 3: Screening process for literature in Section Privacy in learning analytics

Steps	Criteria	Screened	Discarded	Remaining
1. Title and keywords screening	Needs to be related to learning analytics and privacy.	379	250	129
2. Abstract screening.	Large part needs to be dedicated to privacy or privacy principles. Ignore Big Data. Ignore Blockchain. Ignore Academic Library.	129	82	47
3. Full paper screening	Degree of relevance to the guiding questions: “What privacy principles are discussed in learning analytics literature?” and “How is privacy discussed in learning analytics literature?”	47	18	29

Table 4: Privacy principles in privacy-related learning analytics literature

Accuracy	Student control	Consent	Autonomy	Access	Student access
8	14	19	11	14	7
Purpose	Fairness	Transparency	Awareness	Trust	Openness
12	1	18	8	8	2
Data ownership	Anonymity	Opt-in/Opt-out	Vulnerability	Data sharing	Data preservation
15	11	9	1	6	9
Scope of data collection	Data security	Accountability	Assessment	Willingness to share	Identity protection
10	8	5	2	2	1
Relevance	Lawfulness				
1	1				

uncertainties concerning the correctness of principle and concept identification. Despite the limitation of term ambiguity, the review of literature will uncover insights into the privacy debate in learning analytics.

Consent. Slade and Prinsloo (Slade and Prinsloo, 2013) mentions the importance of informed consent in a paper addressing ethical issues and dilemmas in learning analytics. They argue that students are becoming more aware of online surveillance and data harvesting by companies online, but they are not certain that students are aware that educational institutions are starting to do the same.

The same authors explore consent further in a later paper about MOOCs (Massive Open Online Course), and the opt-out mechanism; being given the opportunity of not being handled by learning analytics applications (Prinsloo and Slade, 2015). Prinsloo and Slade are influenced by Solove, when discussing the importance of consent. Consent is here portrayed as important for privacy as it can make privacy violating action legitimate through the act of obtaining consent. The authors go on to discuss six challenges with consent, indicating that this is an intricate topic. The greatest of these challenges is arguably what the authors call “*The problem of scale*” (Prinsloo and Slade, 2015) where they argue that it is almost infeasible for anyone to understand all the dimensions of what they consent to, when faced with for example a learning analytics system or similar systems.

Another challenge with consent is the apparent lightness with which it is given (Khalil, Prinsloo, and Slade, 2018). This is highlighted by Khalil et al. (Khalil, Prinsloo, and Slade, 2018) in a paper analyzing among other aspects, how four MOOCs address student consent in their platforms. The telling title “*The unbearable lightness of consent [...]*” highlights the challenge. Khalil et al. express that consent is often given as an entry-ticket to a services, where little effort is put into reading the Terms and Conditions agreements prior to accepting. This is similar to discussions by Prinsloo and Slade (Prinsloo and Slade, 2015). This lightness of which consent is given can pose a challenge of the validity of the consent. For consent to be valid, it needs to be sincere. Which it arguably is not when given with the lightness these authors describe.

Consent is also the “C” in Drachsler and Greller’s DELICATE checklist (Drachsler and Greller, 2016). DELICATE is a checklist for ethical development and use of learning analytics applications. The authors argue the importance of informed consent by showing to the historical development of research ethics. Using as examples the Nurnberg Code, The Helsinki Declaration and the Belmont report; all related to ethical research involving humans. Drachsler and Greller have an ethical foundation and argumentation for consent, while Prinsloo and Slade (Prinsloo and Slade, 2015) uses privacy as their main argument.

Consent is undoubtedly important for the prompting part, as it can circumvent difficult privacy challenges. However, it is important to research how students regards consent as well. If consent is not important for student, or if they view it with the lightness previously pointed out, there is a problem with the validity of the consent. It is also interesting to research if students feel pressured to give consent when asked by their higher education institution; the unevenness in power between the parts in this relation can possibly incline such an outcome. There are many challenges with consent, and many of these cumulate into the challenge of consent-validity.

Transparency. Pardo and Siemens argue that one of the most important privacy principles for learning analytics is transparency (Pardo and Siemens, 2014). This principle together with others (student control over data, right of access, accountability, and assessment) makes up what Pardo and Siemens consider a foundation of privacy for learning analytics. This foundation is meant as a guidance for educational institutions that already use learning analytics applications; for them to improve their handling of privacy. Pardo and Siemens also argue that transparency possibly is the most challenging privacy principle. Full transparency seems like an ideal for Pardo and Siemens, where all information about the inner workings of the learning analytics applications is open and explained to all stakeholders. Transparency for Pardo and Siemens works as a means of legitimating the learning analytics applications, as well as possibly promoting trust between students and their higher education institutions.

Slade and Prinsloo (Slade and Prinsloo, 2013) have transparency as one of their core principles in their ethic-centered learning analytics framework for higher education institutions. Transparency of purpose is highlighted as important, as well as transparency of access, and transparency surrounding data security. In this way the transparency principle stretches out connections to other principles as well.

Transparency is also part of the DELICATE (Drachsler and Greller, 2016) checklist. Her Transparency is also connected with trust. For Drachsler and Greller, transparency works as a trust-promoter by informing stakeholder of the intents of the learning analytics applications. Transparency becomes a way of showing that there are no dubious intents in the learning analytics applications.

West et al. (West et al., 2020) discusses the lack of a student perspective on learning analytics, including on the principle of transparency. West et al. explains that lack of transparency is previously identified by other researchers as having great impact on student views on analytics systems, indicating that not enough transparency can be damaging for student support of learning analytics. This resonates well with the views of Pardo and Siemens (Pardo and Siemens, 2014), and, Drachsler and Greller (Drachsler and Greller, 2016).

Bellini et al. (Bellini et al., 2019) have a different understanding of transparency. In their three-point list of categories of open questions regarding privacy post GDPR they introduce the concept of “Student’s right of transparency”. This is part of a collection of four open questions that needs to be resolved by

privacy policy. These questions contain a lot more than what other authors understand transparency to accommodate, including students' access to data, as well as data portability and right to deletion of data. Other authors treat these as separate subjects.

Transparency is important for multiple reasons; to show the inner workings of the learning analytics applications, how the data-collection is done, how the data is handled, etc. This can show the students and other stakeholders that data is kept safe and that it is not shared with third parties. As well as being important for higher education to show their intent with their learning analytics applications, to show that their intentions are pure, directed at the students, and designed to promote learning. Researchers seem to agree upon the importance of transparency in learning analytics applications, although they have slightly different conceptualizations of what transparency contains and entails.

Data ownership. Greller and Drachsler (Greller and Drachsler, 2012) is early to point out data ownership as a challenge for learning analytics. They categorize it as one of the “external constraints” of learning analytics in their generic framework, mentioned earlier. They argue that the challenge surrounding data ownership intersects with informed consent, where the former threatens the latter. They argue that an increasing amount of personal data are gathered of individuals without their approval and by solving the challenge of data ownership, clearer boundaries of what is allowed can be set. They seem to impose that the legislation that regulate data ownership do not keep up to speed with the technological development, legislation that worked earlier is now insufficient for a new reality with increasing amount of sensor data and other data generating technologies. This makes room for uncertainties in the legislation. Prinsloo and Slade (Prinsloo and Slade, 2015) discuss the same argument; that legal frameworks can have challenges caused by the speed of technological development. Pardo and Siemens (Pardo and Siemens, 2014) also have an issue with current legislation, expressing that data ownership is not a unique challenge to learning analytics and that it has existed in other research fields before learning analytics. This has been a challenge for learning analytics from the start, and the same points have been highlighted by Siemens (Siemens, 2013), earlier. Siemens argues that data ownership has not been solved culturally (norms) or legally (legislation), making it persistently an open problem.

Drachsler and Greller (Drachsler and Greller, 2016) also elaborate on the data ownership, stating that at present, it is unclear which learning analytics stakeholder owns the collected data. They also state, as does Pardo and Siemens (Pardo and Siemens, 2014), that the question becomes increasingly challenging after the data is processed. Drachsler and Greller pseudo-solve the challenge of data ownership in their DELICATE framework by asking consent and monitoring the data access.

Haythornthwaite (Haythornthwaite, 2017) argues that data ownership should be addressed on an institutional level with policy. This indicates an optimism that the challenge of data ownership can be solved by the higher education institutions themselves and not by national legislation, introducing a slightly different take on the challenge.

A common denominator in the literature is the call for clarity surrounding who owns the data that is collected using learning analytics applications. This is a general problem, not isolated to learning analytics. Other questions branch out from this, for example, who owns the data after its been aggregated and analyzed? (Siemens, 2013). The dichotomy: individual ownership – collectors' ownership, is at the center of this discussion. This can be tied to the concept of privacy self-management as will be explored later. The outcome of this discussion can have a big impact on learning analytics practices, as, along with ownership comes rights of management, distribution, etc.

Pardo and Siemens (Pardo and Siemens, 2014) rightly foresaw the increasing relevance and importance of data ownership. The data ownership challenge is a gordian knot that will not be solved by learning analytics research alone, but the future of learning analytics is largely dependent on a solution to have stable future and it will be greatly shaped by the outcome of this discussion.

Access. Access to personal data, as well as, access to analytics is discussed by Siemens (Siemens, 2013) when relating privacy challenges for learning analytics. In Siemens' work, the principles of access and data ownership intersect. There are two dimensions to these principles as well. Each of these dimensions have open questions associated with them, that Siemens argues that educators need to deliberate. A lot

more questions can spring out of the four core questions in Table 5.

Table 5: Dimensions of ownership and access according to Siemens (Siemens, 2013)

	Data	Analytics
Ownership	Who owns the data?	Who owns the analytics? (results)
Access	Who gets access to the data?	Who gets access to the analytics?

At the center of this challenge is the power-relationship in the dichotomy of student – higher education institution, which is frequently found in learning analytics challenges, for example, as displayed by Slade and Prinsloo (Slade and Prinsloo, 2013). Which of these two stakeholders should own and have access to the data and the analysis of the data? – This is still an open question.

Slade and Prinsloo (Slade and Prinsloo, 2013) place the handling of access under the principle of transparency in their ethical framework for learning analytics. They argue for restricted access for learner’s data, and protection against unauthorized access. They also propose that students should have access to their own data and be given an overview of who has access to their data. From this vision of how access should be handled, parallels can be drawn to the journal systems used by public health care institutions. Slade and Prinsloo connect access not only to transparency, but also to data security, and in some extent to student control and privacy self-management, which will be explored later. Although access cannot be said to be the same as control, they are closely related.

Rubel and Jones (Rubel and Jones, 2016) argue for different levels of access in learning analytics systems, as they seem to argue that context of access plays an important part. They agree with Slade and Prinsloo, that students should have access to their data but uses a different line of argumentation. Rubel and Jones propose that data collected of students might be beneficial for the students when looking for jobs, as a functional part of their academic record, but also indicate that students can be pressured by the job market to provide such information about themselves (Rubel and Jones, 2016). This touches on the subject of data portability as well.

Access seems to be a *soft* challenge for learning analytics. It seems possible to solve access on the institutional level, unlike the dominant perception of data ownership which might need to be solved on a higher level. The question of access is more nuanced than the word indicates. It is interconnected with other privacy principles such as transparency, ownership, and control, making it difficult to address alone, as it can entail unforeseeable consequences without careful consideration.

Student control. Student control, learner control or just control is a frequently mentioned, highly regarded, privacy principle. Rubel and Jones (Rubel and Jones, 2016) raise the question, if students should have decision power over collection and analysis of their data. They argue that control, power to make choices, is important for student autonomy. They offer different suggestions to achieve this, for example by implementing an opt-out mechanism. Another suggestion is to provide the students with a data management dashboard. Rubel and Jones also argues that higher education institutions should, or at least could, go further in providing student control, more so than what is required by legislation. They refer to Family Education Rights and Privacy Act (FERPA) as the foundation of privacy preservation, not the limit. They indicate by this, that higher education institutions should go beyond the regulation of FERPA to preserve students’ privacy.

Pardo and Siemens (Pardo and Siemens, 2014) regard access and the right to correction as the core of the student control principle. They describe how control is related to access, and access is tightly connected with transparency. Pardo and Siemens include a form of student control in their proposed framework by allowing access to data. But they are more hesitant in including right of rectification as it entails difficult challenges. Siemens (Siemens, 2013) also argues, that, what he calls learners control, is important, and he ties it together with the principle of data ownership. In this sense access, ownership and control fits together and can possibly not be viewed in isolation while discussing privacy principles.

Brown and Klein (Brown and Klein, 2020) have in recent research examined over 200 policy documents from universities and other institutions, related to student privacy, and responsible use of data. They

find that the policies give an impression that they include the students in their processes, as well as talk about the students as informed. Despite this, policies give no account of in what manner students can administrate their data. They argue that students have some options to control their data, but they seem difficult to access. As a consequence, it limits student actual control. The authors conclude that the institutions' policies are outdated and need to be updated to accommodate current data use. This indicates a discrepancy between learning analytics researcher's expectations of privacy policies and how the policies work in practice. This can pose a challenge for the legitimacy of learning analytics in the future if not important privacy principles are accommodated in privacy policies.

3.2.3 Privacy concepts

Multiple privacy concepts frequently surface in the learning analytics privacy discussion. Some of these will be highlighted in the following section. *Ethical and juridical frameworks* are first discussed. Followed by a discussion of *the privacy paradox*, *privacy self-management* and *contextual integrity*. These concepts will show to be highly interconnected with each other and with the principles discussed previously.

Ethical and juridical frameworks. Frameworks are prominent in the learning analytics literature and plays an important part in privacy research in learning analytics. Frameworks can be regarded as the steppingstone for solving difficult privacy challenges or to facilitate pragmatic solutions to privacy issues for learning analytics applications.

Frameworks are not legislation. They are guides that attempts to make complex topics manageable and implementable. They are also pragmatic in nature, often created as a way of communicating requirements across professions. The guidelines set out by the framework need to speak to developers as well as product owners, together with other stakeholders. The attractiveness of frameworks is distinct, working as a form of general heuristic for complex problems. Still there are multiple challenges with frameworks. They need to cover enough to fulfill the requirements that they seek to solve, but be vague enough to be generalizable. It can be hard to find a satisfactory balance in this dichotomy of narrow–broad. There is also a challenge of implementing requirements from frameworks as Botnevik et al. (Botnevik et al., 2020) demonstrated when attempting to implement the ethical principles from the Norwegian National Strategy of AI in a kindergarten allocation system. One of the issues with implementation was the contradictions and vagueness in the principles.

Frameworks for learning analytics development and privacy frameworks are popular topics in the literature. From early in the learning analytics research field the creation of ethical frameworks has been called for. Ferguson (Ferguson, 2012) argues in her seminal paper from 2012 that, at the time, no thorough ethical framework for learning analytics existed. Calling it a pressing need, she calls for the creation of an ethical framework for learning analytics. Ethical frameworks such as DELICATE (Drachsler and Greller, 2016) and the framework proposed by Slade and Prinsloo (Slade and Prinsloo, 2013) has since then been developed. Drachsler and Greller's (Drachsler and Greller, 2016) DELICATE framework is possibly the most elaborate of the two as it draws upon Slade and Prinsloo's (Slade and Prinsloo, 2013) work. There is also discussion of other frameworks, guidelines, principles and tools for facilitation ethical and privacy preserving conduct in the learning analytics literature. This can indicate a need for clarity when faced with ethical and privacy considerations in learning analytics. The development, improvement, as well as, the value of frameworks, is persistently up for discussion in the literature. These frameworks are often called ethical frameworks, but they address privacy as well, making them relevant for this thesis.

Another type of frameworks that are frequently discussed in learning analytics are legal frameworks such as the GDPR and FERPA. This bridges the research field of learning analytics with Law. These frameworks cannot be directly compared with the frameworks mentioned above, as the legal frameworks has to be followed. While frameworks created by learning analytics researchers are more for guiding purposes if not ratified by higher education institutions. FERPA and more recently GDPR is discussed and highly relevant. FERPA is prominent in literature discussing the American higher education environment, while GDPR is prominent in literature with a European point of view. The latter is most relevant for this research, as it studies the Norwegian higher education setting. Although Norway is not part of the EU, it is part of the Economic European Economic Areas which commits to the implementation of the GDPR. The GDPR is important as it needs to be respected by Norwegian higher education institutions. It does

also accommodate many of the principles earlier related, as it builds on the following principles: *Lawfulness, fairness and transparency; Purpose limitations; Data minimization; Accuracy; Storage limitations; and Integrity and confidentiality* (GDPR, 2016). Higher education institutions will have to respect these principles if implementing learning analytics applications.

Privacy self-management. Privacy self-management is tightly connected with the privacy principles of student access, data ownership, student control, and consent. Prinsloo and Slade (Prinsloo and Slade, 2015) goes into detail of the challenges with privacy self-management. Privacy self-management is about having responsibility and control over your own privacy. This is made possible through rights provided for example by the GDPR where individuals are given rights to access and correct personal data about themselves (GDPR, 2016). As a consequence of the GDPR individuals are also asked consent prior to having their data collected by, for example websites. Here one is given the opportunity to opt-out of data collection in the form of denying cookies for other purposes than the functionality of the website. Initially this can seem as a good way of giving individuals autonomy and control over their own data sharing, but there are challenges with this as Prinsloo and Slade (Prinsloo and Slade, 2015) discusses. They largely build their argumentation with inspiration from Solove, which is skeptical to the practice of opt-in/opt-out (Prinsloo and Slade, 2015).

Opting-out is a prominent theoretical solution for learning analytics applications for students that do not want to participate. Privacy self-management critique by Prinsloo and Slade (Prinsloo and Slade, 2015) problematizes this solution among other aspects of privacy self-management. The question of privacy self-management boils down to a question of responsibility. Should handling privacy be the responsibility of the individual or the collective – in the form of higher education institutions or higher up institutions such as the state. A common political statement is that the public sector is created to solve problems that are too big for individuals to handle alone, this prompts the question; is privacy management a challenge too big to handle alone?

The privacy paradox. Tsai et al. (Tsai, Whitelock-Wainwright, and Gašević, 2020) discuss the privacy paradox and how it affects learning analytics. They describe the privacy paradox as individuals' representation of outspoken privacy concerns but with a behavior that do not indicate such concerns. Following from this, individuals say that they are concerned with their privacy, but their behavior shows little sign of this.

The privacy paradox is interesting to research as it is an apparently open contradiction, which is intriguing. It also promotes an interesting question related to handling privacy. If the privacy paradox can be stated to occur among student when dealing with privacy for learning analytics applications, it imposes the question of how much weight one should put on students' outspoken opinions on privacy. Should one listen to what they say, or what they do? As they are the largest stakeholder in learning analytics, it is argued that their opinions need to be respected. Although if they are incompetent to display their true opinions through words, this might need to be reevaluated. This is possibly to push the question to the extreme but moderated questions needs to be answered on how to accommodate for the privacy paradox, if recognized.

Contextual integrity. Prinsloo and Slade (Prinsloo and Slade, 2015) mentions contextual integrity as a great concern in relation to the study of surveillance and privacy. The authors state that “*Data and information that are collected and/or shared in one context lose contextual integrity when shared or used out of context*” (Prinsloo and Slade, 2015, p.89). Derived from this, data usage and sharing are determined by the context. This can also entail challenges in regard to the principle of consent, among others, as minor changes in data usage or sharing can cause consent to become invalid, assuming that one acknowledges the contextual integrity theory. Drachsler and Greller’s (Drachsler and Greller, 2016) DELICATE checklist or ethical framework is also influenced by contextual integrity theory.

3.2.4 Summary

In this section, important privacy principles and concepts, for learning analytics, has been highlighted. The principles and concepts discussed has shown to be highly interconnected, working more similar to a unity than as individual entities. This corresponds with Solove (Solove, 2009) conceptualization of privacy. Privacy principles and concepts are interconnected and can overlap as previously shown in Figure 1(b).

Only a small selection of principles and concepts has been explored. The ones explored still suffice to give an overall impression of what characterizes the privacy discussion in learning analytics. The discovered dichotomies can also help to understand the privacy discussion. Individual ownership – collectors’ ownership, in relation to data ownership. The relation of power between student – higher education institution. The scope of frameworks and their limitations captured in narrow–broad. Individual – collective responsibility when discussing how privacy should be handled.

As discovered in the first part of the literature review, the goal of learning analytics is to improve learning. A corresponding general goal of privacy in learning analytics, can be to maintain student’s privacy to a satisfactory degree while being able to use learning analytics applications in higher education institutions.

Trying to achieve this goal entails a lot of different challenges that needs to be addressed. Some of the challenges are highlighted in the discussion of principles and concepts, while there exists numerous more. Solutions to some of these challenges can potentially be found by looking into student perceptions of privacy in learning analytics, as will be explored in the next section.

3.3 Student perceptions of privacy principles in learning analytics

A compressed version of this part of the literature review was published in proceedings of the European Conference on Technology Enhanced Learning (EC-TEL 2020)(see Botnevik, Khalil, and Wasson, 2020). The following version is a revised and expanded account, where more emphasis is put on the methodology and results of the reviewed literature.

Little research is done on the student perspective of privacy, in learning analytics. Arnold and Sclater (Arnold and Sclater, 2017) did a literature search in 2017 and found only three papers on the topic. When searching for literature for the following review, 11 papers were found. This lack of papers highlights the need for more research on the topic, as this research is.

With limited research on the topic, it is important to look at the existing work that is done. In the following sections, research that is related to student perceptions of privacy in learning analytics, will be explored. First the literature gathering process will be described before going into discussion of the literature. Two questions guided the review:

- What privacy principles in learning analytics are most important for students?
- In relation to student perceptions, what privacy principles are explored, most, in learning analytics research?

3.3.1 Review method

The literature was collected by searching three digital libraries with similar search queries. The small variation in search queries is because of the digital libraries variable quality and depth. Curated databases like Web of Science yields fewer results even with broad search terms, while non-curated databases like Google Scholar returns a wider set of results. A more specific search query is therefore necessary in the case of the latter. The queries used to search for literature are displayed in Table 6.

Table 6: Search queries used to find literature, in Section 3.3

Date	Database	Search query
20.03.20	Web of Science	ALL=(“Learning Analytics” AND “Privacy” AND (“student” OR “students”))
20.03.20	Google Scholar	With all the words: Privacy. With the exact sentence: Learning Analytics. With at least one of the words: Student Students.(In the title.)
23.03.20	ACM Digital Library	[All: “privacy”] AND [[All: “student”] OR [All: “students”]]

The criteria for the screening process was:

- The literature needs to be in English.
- The literature needs to be in the research field of learning analytics.
- The literature needs to be about privacy (preferably privacy principles).
- The literature needs to represent the perspective of students in higher education institutions.

The last criterion implies that the research described in the literature needs to give some inquiry into student perceptions. This could be in the form of interviews, focus groups, questionnaires, or other ways of getting the opinions of students. The search, screening and eligibility process resulted in 11 papers that fulfilled the criteria. This process is displayed in, Figure 10, as a PRISMA 2009 flowchart diagram (Moher et al., 2009).

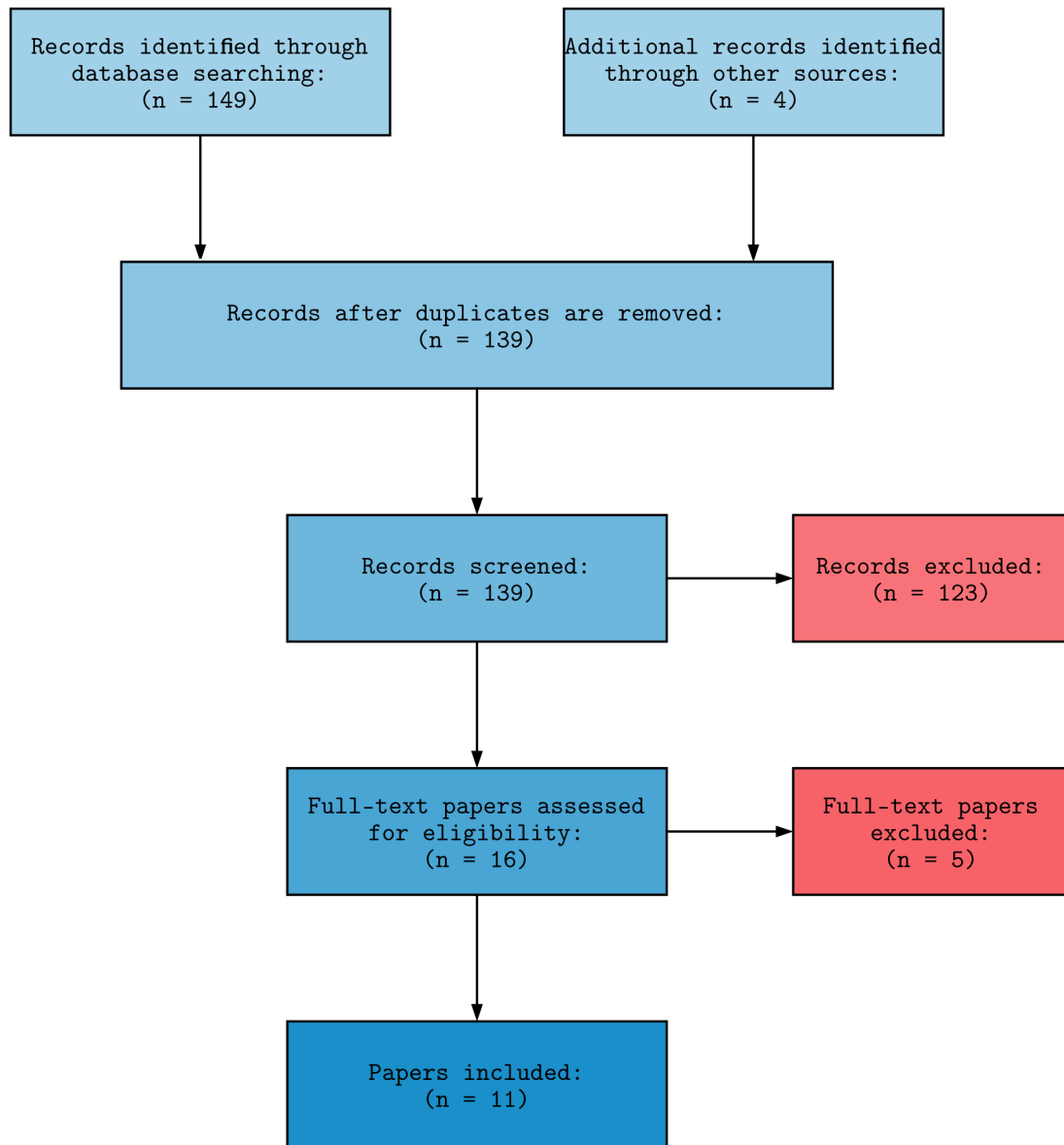


Figure 10: Modified PRISMA 2009 flowchart diagram (Moher et al., 2009)

Table 7: Literature included in Section 3.3

Reference	Title
(Whitelock-Wainwright et al., 2020)	Assessing the validity of a learning analytics expectation instrument: A multinational study
(Vu, Adkins, and Henderson, 2020)	Aware, but Don't Really Care: Student Perspective on Privacy and Data Collection in Online Courses
(Slade, Prinsloo, and Khalil, 2019)	learning analytics at the intersections of student trust, disclosure and benefit
(Adejo and Connolly, 2017)	Learning Analytics in a Shared-Network Educational Environment: Ethical Issues and Countermeasures
(Falcão et al., 2020)	Perceptions and expectations about Learning Analytics from a Brazilian Higher Education Institution
(Tsai, Whitelock-Wainwright, and Gašević, 2020)	The privacy paradox and its implications for learning analytics
(Roberts et al., 2016)	Student Attitudes toward Learning Analytics in Higher Education: "The Fitbit Version of the learning World"
(Whitelock-Wainwright et al., 2019)	The Student Expectations of Learning Analytics Questionnaire
(Ifenthaler and Schumacher, 2016)	Student perceptions of privacy principles for learning analytics
(Arnold and Sclater, 2017)	Student Perceptions of Their Privacy in Learning Analytics Applications
(Sun et al., 2018)	Taking Student Data for Granted? A Multi-Stakeholder Privacy Analysis of a Learning Analytics

The literature included in this review is listed in Table 7. Key information from the literature has been gathered in Table 8 for easier comparison of the literature. Factors regarding the key information in the table will be highlighted and discussed.

The categories of key information were chosen with the intent to cover the most important privacy related aspects of the literature. In addition to exploring how research on student perceptions of privacy for learning analytics is conducted. Arguably, the most important category is captured by the question: "What privacy principles are students asked to give their opinion on?". This question is expressed in the table-column *Privacy principles explored*, which is a quantification of the questions asked to students, documented in the literature.

By doing this quantification, it is possible to deduct what is considered important to get student opinions on, according to learning analytics researchers. This in turn reflects what privacy principles are important for the field of learning analytics. This also gives insight into what privacy principles students consider important for learning analytics.

In addition to the mentioned information of *Privacy principle explored*, there following information is also recorded: *Method*, which describes what research methods was used to gather student opinions; *No. Students*, which indicated the amount of students included in the research; *Country*, describes where the study was conducted; *Questions*, is the amount of questions asked to students, also type of questions where this is relevant; *Results relating to privacy*, describes privacy findings from the research; and, *Values identified*, which are the privacy values or principles that the researchers identify to be important to students. This information is derived from the results presented by the researchers of the respective literature.

The full papers have not been mapped. There are a lot of elements in the literature that are not about privacy and these are ignored. Only privacy related elements are given attention. Some of the papers do not have an initial focus on student perceptions of privacy principles, but are still included because of their relevance. An example is the research by Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020), here results exploring student perceptions is a byproduct of the research, as the authors describe the development process of a questionnaire framework called the Student Expectation to Learning Analytics Questionnaire (SELAQ). The framework is intended to be

used to explore student expectations for learning analytics applications. These two papers mainly focus on the development and testing of the framework. The gathering of student opinions is to test and assist in validating the framework. Because of this, the research is still relevant, although the relevancy is slightly reduced.

3.3.2 Research methodology

Six of the papers contain research utilizing questionnaires (Adejo and Connolly, 2017; Ifenthaler and Schumacher, 2016; Slade, Prinsloo, and Khalil, 2019; Vu, Adkins, and Henderson, 2020; Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020), three describe research that uses focus groups or interviews (Falcão et al., 2020; Roberts et al., 2016; Sun et al., 2018), and two documents research where a combination of these are used (Arnold and Sclater, 2017; Tsai, Whitelock-Wainwright, and Gašević, 2020). While Arnold and Sclater document the use of interview (at Jisc) (Arnold and Sclater, 2017), the interview is composed of three questions and is more similar to a face-to-face questionnaire. Despite this, the authors' term (interview) is used.

For the focus groups and interviews, between 20 and 41 students participate. An exception to this is Arnold and Sclater's (Arnold and Sclater, 2017) research, which has 402 interviewed students (short interviews). The questionnaires range between approximate 200 and 1650 student participants, while the majority is between 200 and 600.

The majority of questionnaire-based studies have few items in their questionnaires. Except the questionnaires by Slade et al. (Slade, Prinsloo, and Khalil, 2019) and Ifenthaler and Schumacher (Ifenthaler and Schumacher, 2016) which has 57 and 30 items (including subitems), respectively. There are advantages and disadvantages of having few questionnaire items. With fewer items the chance is probably higher that students will answer the questionnaire. On the other hand, having more questions in the questionnaire gives the opportunity of gaining more insight into the opinions of students. This has the possible downside that fewer students will answer the questionnaire, possibly evident in the research by Slade et al. as they have a response rate of 2.7% (Slade, Prinsloo, and Khalil, 2019).

A different questionnaire style is adopted by Arnold and Sclater (Arnold and Sclater, 2017). This questionnaire has few items – three polar (yes/no) questions. Insights from the questionnaire is limited because of the few questions, but they have a higher response rate (approximately 70% on the questionnaire they have documented the response rate (Arnold and Sclater, 2017)).

Another approach, in between the two questionnaires above, if represented on a scale is demonstrated by Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020) They down scaled their original SELAQ questionnaire after a pilot study, from 39 to 19 items. This was later reduced to 12 items, five of which is related to privacy. The reasons for the reduction in questions varied, but it resulted in a concise and compact questionnaire.

It is worth mentioning that the questionnaires mentioned serves different purposes, and this can be the reason for their differences. Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020) are interested in developing a questionnaire that will be reused as a tool, which entails that it needs to be solid and thoroughly tested. Other questionnaires are more exploratory, possibly made to fit one particular setting and without the aim of being reused.

Mostly polar and Likert scale questions are used in the questionnaires of the selected papers. There are few open questions in the questionnaires, but in the focus groups / interviews there are more use of open questions.

Analysis methods. All of the questionnaire-based research use some form of descriptive statistics as part of their analysis (Adejo and Connolly, 2017; Arnold and Sclater, 2017; Ifenthaler and Schumacher, 2016; Slade, Prinsloo, and Khalil, 2019; Tsai, Whitelock-Wainwright, and Gašević, 2020; Vu, Adkins, and Henderson, 2020; Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020). In addition to this, some also use more complex analysis, like correlation analysis (Slade, Prinsloo, and Khalil, 2019) or regression analysis (Ifenthaler and Schumacher, 2016) to accompany their descriptive statistics, but this represents a minority. Cronbach's alpha is also a frequently used method as part of analysis. Cronbach's

Table 8: Key information from the literature.

Ref	Privacy principles explored	Method	No. Students	Country	Questions	Results relating to privacy	Values identified
(Whitelock-Wainwright et al., 2020)	Consent, Data security, Third party data usage, Transparency.	Questionnaire.	161, 543, 1247.	Estonia, Spain, The Netherlands.	12 items (5 on privacy).	Students expect a high degree of data protection.	Data security, Restricted access of third parties.
(Vu, Adkins, and Henderson, 2020)	Awareness, Collection / use of personal data (HEI), Anonymity.	Questionnaire.	1647.	United States of America.	6 items.	Students are aware of the possibility that institutions may use their data, but they are not very concerned.	None.
(Slade, Prinsloo, and Khalil, 2019)	Awareness, Collection / use of personal data (HEI and 3rd party), Data security, Control, Anonymity.	Questionnaire.	286.	United Kingdom.	28 items.	Not very high levels of concern regarding their privacy. Students show high levels of trust in their universities.	Control.
(Adejo and Connolly, 2017)	Collection / use of personal data (HEI), Security, Awareness, Trust, Accuracy, Access.	Questionnaire.	209.	Scotland.	6 items.	Inadequate awareness among students.	Withdrawal of data. Access.
(Falcão et al., 2020)	Purpose of data usage (and an open question on privacy/ethics).	Interview and focus groups.	5 (interview), 17 (in focus groups).	Brazil.	7 items.	The students have little concern about privacy.	Anonymity.
(Tsai, Whitelock-Wainwright, and Gasević, 2020)	Consent, Data security, Collection / use of personal data (3rd party), Transparency.	Questionnaire and focus groups.	674 (questionnaire), 26 (in focus groups).	United Kingdom.	12 items (5 on privacy).	Purpose, access and anonymity are important to measure ethics/privacy integrity.	Anonymity, Purpose, Access, Consent, Data security, Control.
(Roberts et al., 2016)	None (privacy topics was voiced by students through open questions).	Focus groups.	41.	Australia.	Videos and prompt questions (uncertain number).	Students have privacy concerns regarding learning analytics.	Restricted access, Informed consent, Control (opt-in, opt-out).
(Whitelock-Wainwright et al., 2019)	Consent, Data security, Third party data usage, Transparency.	Questionnaire.	191 (study three).	England.	12 items (5 on privacy).	None.	Data security.
(Ifenthaler and Schumacher, 2016)	Access, Control, Collection / use of personal data (HEI), Transparency.	Questionnaire.	330.	Germany.	Three questionnaires (68 items in total, 30 on privacy).	Students have a willingness to share educational data for learning analytics purposes.	Control.
(Arnold and Sclater, 2017)	None	Interview, questionnaire.	402 (interview), 435 (questionnaire).	United Kingdom (interview), United States of America (questionnaire).	3 items in common.	The majority of students do not mind data usage as long as it helps them achieve better grades.	None.
(Sun et al., 2018)	Awareness, Collection / use of personal data (HEI).	Semi-structured interviews.	32 (20 students).	United States of America.	No documentation of questions.	Students want to be informed and have a say when the university uses their data.	Awareness, Consent, Control (opt-in, opt-out).

alpha is used to assess the reliability of responses.

Thematic analysis (Roberts et al., 2016; Sun et al., 2018) and content analysis (Falcão et al., 2020) is used by research utilizing a qualitative research methodology. This method is based on grouping findings in categories. This shares some similarity with clustering as used by Slade et al. (Slade, Prinsloo, and Khalil, 2019) to present their questionnaire results.

3.3.3 Explored privacy principles

To explore what privacy principles students are questioned about, all the documented questions used in questionnaire related research were categorized and quantified. Questions were categorized loosely on what privacy principles they address. Some categories are not directly related to privacy, but are included as they are relevant. Questions can have multiple categories as they can address multiple privacy principles. Research using questionnaires but lacks documentation of questions have been excluded. This applies to the paper by Ifenthaler et al. (Ifenthaler and Schumacher, 2016). Papers included in the analysis were: (Adejo and Connolly, 2017; Arnold and Sclater, 2017; Slade, Prinsloo, and Khalil, 2019; Tsai, Whitelock-Wainwright, and Gašević, 2020; Vu, Adkins, and Henderson, 2020; Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020). The idea behind the quantification is that the magnitude of questions addressing a privacy principle can indicate its importance for learning analytics.

The most explored principle is *collection / use of personal data* for higher educational institutions. This is expected as the goal of learning analytics is to utilize learner data in order to assist their learning. To be able to do this, student data needs to be collected and used. The collection and use of personal data can be connected to information aggregation, as discussed previously. Student's support for the use of learning analytics is interesting to explore, although it is controversial how much weight should be put on students' opinion.

Student support of learning analytics. Although *collection and use of personal data* (HEIs and third party) is not directly identified as a privacy principle, it entails information aggregation and other privacy related topics, and is therefore highly relevant to privacy. The collection and use of personal data constitutes the foundation of learning analytics.

Ifenthaler and Schumacher (Ifenthaler and Schumacher, 2016) finds that students can be prone to sharing higher education institution related data for use by learning analytics applications. Slade et al. (Slade, Prinsloo, and Khalil, 2019) discovers similar results, they report findings that indicate that 79% of students are comfortable with the university using their personal information for purposes like improving services and support for students. This again is supported by evidence from Tsai et al. (Tsai, Whitelock-Wainwright, and Gašević, 2020), who finds, in their focus group research, that students accept the use of their data for improvement of educational services, among other reasons. It is important to highlight that this related to sharing data with their higher education institution and not third parties. The research by Tsai et al. indicate that students are uncomfortable with sharing data with third parties.

Other researchers find that students are skeptical. Based on the key themes of Roberts et al. (Roberts et al., 2016), some students seem skeptical of learning analytics applications, questioning the value and impact as well as having privacy concerns. Sun et al. (Sun et al., 2018) also finds that students think that the learning process consists of more than a learning analytics system can measure, questioning the usefulness of learning analytics. These results are similar to the ones identified by Roberts et al. as a key theme in the research finding; "more than just a number" (Roberts et al., 2016). This can be tied to the 2nd challenge of information aggregation related earlier, the challenge related to the data foundation used by systems to make decisions about individuals.

Other researchers find little privacy concerns among students in regard to learning analytics applications. Vu et al. (Vu, Adkins, and Henderson, 2020) finds that students know that their data is being collected, but they are not concerned with their higher education institution's use of this data. Similarly, Falcão et al. (Falcão et al., 2020) finds that students do not really have any privacy concerns in relation to use of their personal data. They find that instructors are more worried for students' privacy than the students are themselves.

It is evident from some of the literature that researchers are interested in finding out if student data can be exchanged for certain benefits (e.g., Arnold and Sclater, 2017; Slade, Prinsloo, and Khalil, 2019). Improved grades are an example of a benefit students can receive for their data (Arnold and Sclater, 2017). As Arnold and Sclater shows, most students want to trade their data for improved grades. They also suggest that more research is needed to explore what students want their data to be used for. This will be pursued in the empirical data collection of this research.

After the collection / use of personal data, the most explored privacy principles are *consent*, *awareness* and *data security*. The full overview of the most explored privacy principles, based on this quantification, is shown in Figure 11.

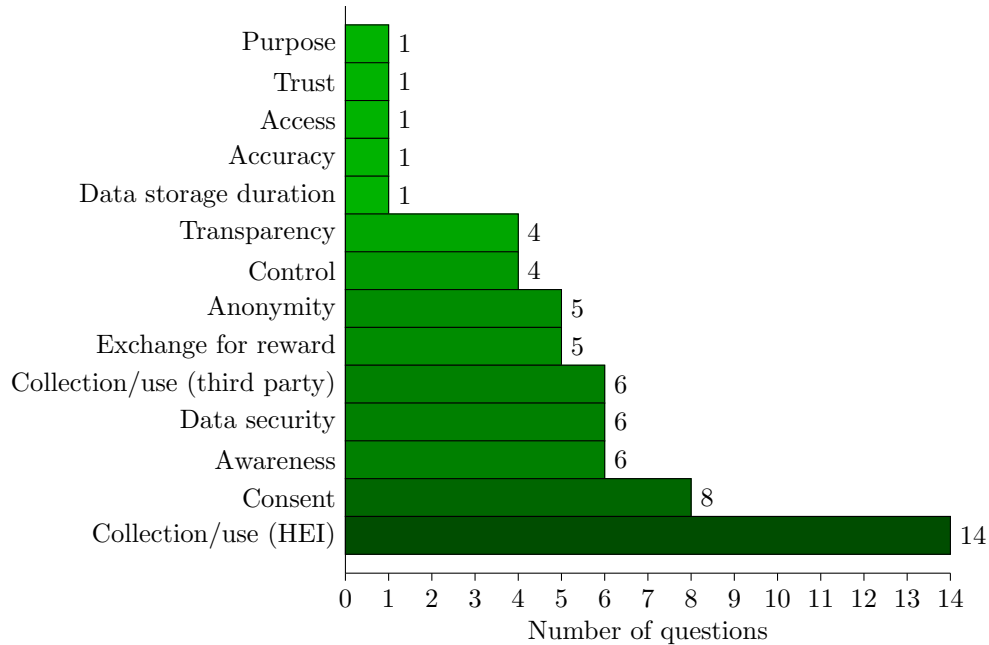


Figure 11: Categories of questions in questions asked to students.

The principles that are least asked about in the literature are: trust, access, accuracy, storage duration and purpose. The only research that goes into these questions explicitly (except for purpose) is by Adejo and Connolly (Whitelock-Wainwright et al., 2019). Their paper was published in 2017 and the principles they ask about aligns with some of the privacy principles of the GDPR (published in 2016, implemented in 2018), which can give a plausible explanation for why these principles are explored. The reason for why these principles are not asked in questions to students can be because of their relevance. For example, for a learning analytics system, it is important to have a purpose in order to collect student data, this is regardless of whether students think this purpose is necessary or not. Similar explanations are also possible to formulate for the other least asked about principles.

Another way of highlighting what privacy principles are explored in the literature, is to look at what privacy principles are asked about on a per-paper basis. The quantification above (Figure 11) evaluates each individual question on its own, while the following (Figure 12) evaluates each paper on its own. An example of the difference is given; in Figure 11 consent is the topic of 8 questions. In Figure 12 consent is the topic of question(s) in 3 papers. This means that the 8 questions consent is the topic of, is distributed over 3 papers. This changes the view slightly on what privacy principles have been explored the most, but this variation is so small that the analysis is not changed noticeably, and hence is not given any attention.

3.3.4 Important principles for students

In an attempt to find what privacy principles the students find most important; the privacy related results of the literature have been analyzed. This data is represented in the *privacy principles explored* column of Table 8. This includes all research in the review, not only research concerned with questionnaires as

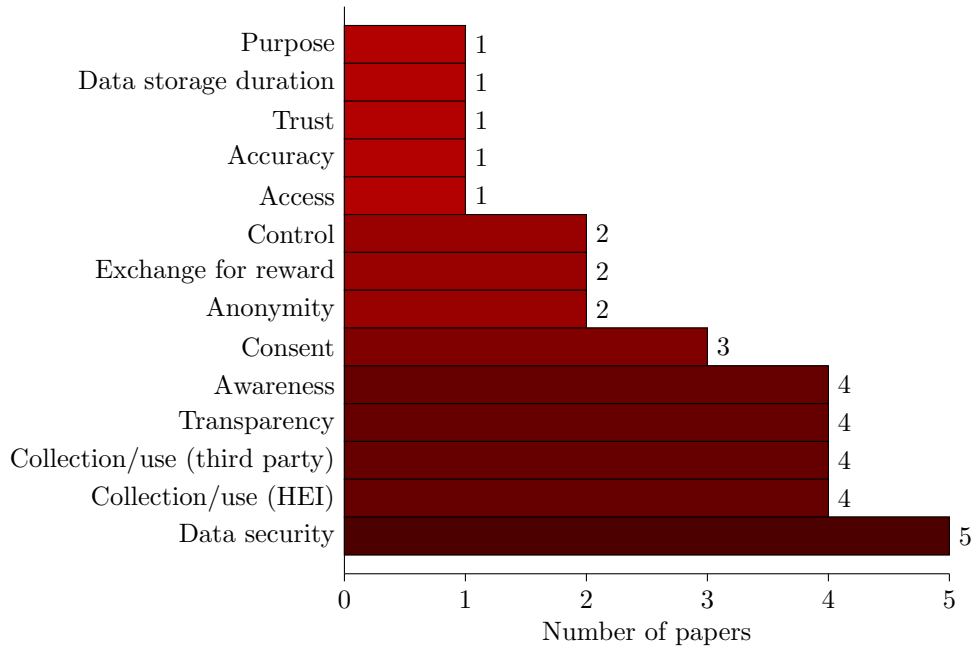


Figure 12: Distribution of privacy principles found in the literature.

the questionnaire question analysis above. Findings are categorized based on the privacy principles they address. The result of this analysis is presented in Figure 13.

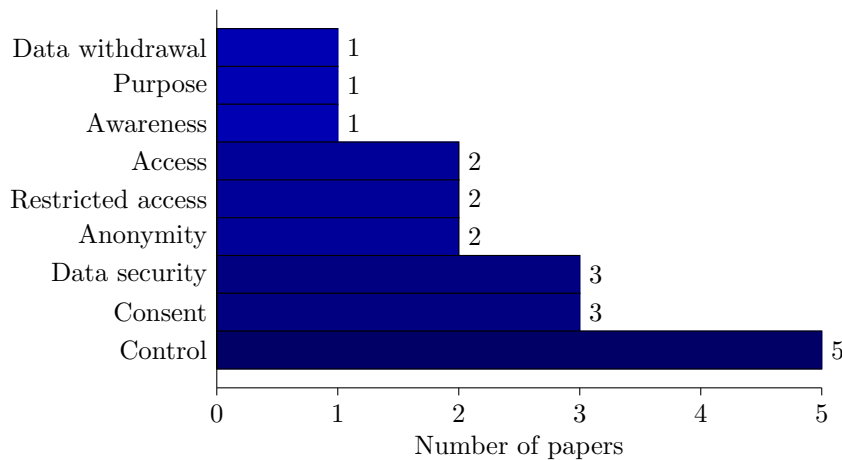


Figure 13: Distribution of privacy principles important to students.

Control is the privacy principle that seems to be the most important to students across all the research subject to this review. Slade et al. (Slade, Prinsloo, and Khalil, 2019) finds that close to 70% of students finds it important to control access to their information. They also find that students want to control what information is collected of them. Some students in the research by Tsai et al. (Tsai, Whitelock-Wainwright, and Gašević, 2020) voice that ideally, they would want to have control over their own data, but understand that it is unrealistic.

The results of the research by Sun et al. (Sun et al., 2018) show a different side. Here a few students would prefer not to have any control when it comes to their own data. Their opinion is that everyone should be included in the analytics and that individual students should not have the rights to reserve themselves from participating. This entails an alternative view of privacy self-management.

Consent has a special position in privacy because it harbors the potential to neutralize privacy, as discovered previously (Prinsloo and Slade, 2015). An understanding of consent in relation to privacy is that

a violation of privacy is not a violation if consent has been obtained. This makes consent a powerful and interesting principle to explore. This is reflected in the amount of consent-related questions asked to the students. Why consent is explored is not always clear from the literature, as the reason is not always explicitly stated, or that it is hard to deduce from the research.

Tsai et al. finds that students highly regard consent (Tsai, Whitelock-Wainwright, and Gašević, 2020). This is supported with evidence from the research by Roberts et al. (Roberts et al., 2016), they find that students have a need for informed consent when it comes to sharing their data, students want to know what their data is used for. This needs also to be seen in connection with awareness, access, and control. This is specially related to control as Roberts et al. finds that students want an opt-in/opt-out option. As consent is frequently asked about in the questionnaires (Figure 11), this is a possible explanation for why consent is found to be important to students.

Data security. Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2020), finds that students have high ideals when it comes to data security. This is true for all the countries they perform their questionnaire in. This is also found in the additional research that uses the SELAQ, 12-item-questionnaire (Tsai, Whitelock-Wainwright, and Gašević, 2020; Whitelock-Wainwright et al., 2019). Not many comments are made on these findings by the researchers, outside of stating them. This makes it difficult to discuss the principle. More research is needed to highlight these findings.

It is interesting that many of the most frequently asked about principles are also the ones that are most important to students. This can indicate that there is an imbalance in representation of principles. It would be interesting to research if students were presented with questions about a broad selection of privacy principles, that it would produce different results. By doing this, it can also give a fuller view of how students perceive privacy principles, which can also say something about their privacy perception.

It is important to highlight that even if these principles are found to be important to students in multiple studies, they are not found important in all of them. There can be multiple reasons for this, for example, as discussed above, that not all studies have questions that give students opportunities to voice their opinion about particular privacy principles.

3.3.5 Other insights

SELAQ & SHEILA. Four of the papers in this review has connections to the SELAQ framework or the SHEILA framework. SELAQ is a tool to measure student expectations in regard to higher education institution's safeguarding of privacy, and ethical conduct, in relation to learning analytics, as well as students' expectations to learning analytics features (Whitelock-Wainwright et al., 2020). SELAQ was created to support the empirical data gathering related to learning analytics among students that have little to no prior experience with learning analytics applications (Whitelock-Wainwright et al., 2020).

The SELAQ – questionnaire framework, seems to have been incorporated into the wider SHEILA framework. SHEILA has protocols for how to gather empirical information surrounding expectations and needs for learning analytics applications. The protocols are designed to fit questioning of multiple different stakeholders (student and staff), and multiple forms of empirical data gathering techniques (survey and focus groups)¹⁶.

Out of the four papers with connection to these frameworks: Tsai et al. (Tsai, Whitelock-Wainwright, and Gašević, 2020) conduct the SELAQ questionnaire; Falcão et al. (Falcão et al., 2020) uses protocols from SHEILA for focus groups; and, the other two are Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020), who discuss the development and assessment of the framework.

As SELAQ is designed to support empirical data gathering from students with little experience with learning analytics, it has a high potential to be used in the Norwegian setting.

¹⁶<https://sheilaproject.eu/sheila-framework/>

Geographical coverage. The research covered by the literature indicates that most research of the research related to students perceptions of privacy principles, takes place in Europe, America and Australia. The United Kingdom and the United States of America is most highly represented. There is a lack of research in the context of Africa, Asia, Latin America and the Nordic countries.

Comparative studies between countries. Few papers document research conducted in more than one country. There is little comparison between how students of different nationalities perceive the privacy principles of learning analytics. The exception to this is Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020), who are some of few researchers that conduct multinational studies on this topic. The research by Falcão et al. (Falcão et al., 2020) indicates that cultural aspects can affect how privacy principles are viewed by students. The research by Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2020) also indicates that there can be differences based on nationality.

Generalizability. Few if any of the studies report that their findings are generalizable. Many of the studies have a small population, often only from one higher education institution, where the students are from a narrow set of study-directions. This calls for surveying larger amounts of students at multiple higher education institutions within different kinds of study-directions. Alternatively, a larger amount of small scale studies can be conducted in similar ways as of facilitating for comparison of results.

3.4 Summary

Most of the research exploring student perceptions on privacy uses questionnaires to gather empirical data, frameworks for this collection also exists. When analyzing the data, descriptive statistics is most frequently used. The minority of research is based on a qualitative research methodology, employing interview and focus group, and analyzing their data using thematic analysis and content analysis.

The collection and use of student data is a highly explored topic in the reviewed literature. This is important as it includes student opinions in the data foundation of learning analytics knowledge. Surveying student limits in relation to the collection and use of their personal information, is an important step in discovering and respecting their boundaries. Many students are interested in sharing their personal information for learning analytics benefits, such as improved grades.

Based on the analysis in this review, students regard control as their most valued privacy principle, closely followed by data security, and consent. Because of the large overlap between privacy principles that are asked about, and the privacy principles students find important, there exist uncertainty surrounding the representativeness of these results. Privacy principles are possibly explored in an imbalanced way, entailing imbalanced results. This can be addressed by representing more privacy principles when questioning students about privacy in learning analytics.

4 Methodology

For this research, a quantitative research design will be adopted. Student perceptions of privacy will be explored using a questionnaire. The questionnaire needs to be developed, and this is the focus of the next chapter (Chapter 5). The questionnaire will be a self-administered web-based questionnaire. In addition to the questionnaire, secondary analysis will be performed on questionnaire data for the Norwegian Data Protection Agency¹⁷ (NDPA) 2019/2020 questionnaire¹⁸. The secondary analysis will give background to how the Norwegian population in general view privacy.

The results of both the secondary analysis and questionnaire will be analysed using descriptive statistics. Descriptive statistics is chosen as it is a prominent method of analysis in similar learning analytics research that explores students perceptions of privacy, prominent methods of analysis for learning analytics is explored using desk research.

Prior to the conduction of the quantitative research, desk research is necessary. This needs to be conducted in order to gather information that will be used to develop the questionnaire. The desk research takes the form of a literature review. How the literature review is conducted is related in Chapter 3. The chapter begins with describing the desk research and how the research questions will be answered.

4.1 Desk research

In this research, desk research in the form of a literature review, will be administered in order to provide answers to the research questions. The literature review will be divided in three parts, all exploring different topics within the learning analytics research field, with focus on privacy and the student perspective. Part one will introduce the learning analytics research field, while the two others quantify the privacy debate in learning analytics literature.

The quantification of the privacy debate in learning analytics will be used to identify privacy principles that are central to the learning analytics research field. These principles will then be adopted as indicators of privacy for learning analytics, when measuring student perceptions in the questionnaire. This will quantify students' view of privacy, and also, provide a nuanced view of how students perceive privacy for learning analytics. This will be used to answer the research questions.

4.2 Addressing the research questions

All of the research questions will be answered in the Discussion (Chapter 7). The different questions will be answered based on insights from different processes, how this will be done is described next.

The first research question: **What privacy principles are most relevant for learning analytics?** will be answered based on the theoretical positioning (Chapter 2) and the literature review (Chapter 3), a quantification of the privacy debate in learning analytics is central to this answer.

The second research question: **How do Norwegian students perceive privacy in general, and in relation to learning analytics?**, will be answered in the following way:

A questionnaire will be developed based on privacy principle findings in the literature review, as well as with inspiration from the NDPA's 2019/2020 privacy questionnaire, containing data on how the general Norwegian public view privacy. The questionnaire will be implemented at the University of Bergen, gathering student opinions of privacy in general, and, in related to learning analytics. The questionnaire results will be analyzed using descriptive statistics, and the results will be interpreted to answer the research question.

The third research question: **What privacy priorities do Norwegian students have for learning analytics?**, will also be answered based on the questionnaire results. Questions dedicated to exploring different privacy principles will be central to answering this question.

¹⁷Datatilsynet

¹⁸<https://www.datatilsynet.no/regelverk-og-verktoy/rapporter-og-utredninger/personvernundersokelser/personvernundersokelsen-20192020/>, accessed: 09.04.21

The fourth research question: **What learning analytics services and benefits are acceptable to Norwegian students?**, will also be answered based on the questionnaire results. Questions dedicated to measuring the general desirability of learning analytics, as well as specific questions on services and benefits will be central to answering this question.

The questionnaire results are essential to answering the research questions. The questionnaire is conducted within a quantitative research design, which entails a lot of positions, these positions are described next. The methodological discussion is mostly based on Bryman's (Bryman, 2015) methodology book; *Social Research Methods*, with input from other methodology authors.

4.3 Quantitative research

Some authors refer to quantitative research as a research strategy (Bryman, 2015; Denscombe, 2017), others call it a research family (Blaxter, Hughes, and Tight, 2010), others again call it a research style (Bell and Waters, 2018). Regardless of the term used, quantitative research can be understood as being related to the structured planning of research of a particular genre. The research methods used, can vary, but are within a certain collection of related methods. In addition, the research strategy it is often accompanied by a paradigm or a certain way of viewing the world. Social research strategies are often divided between quantitative and qualitative research strategies (Bryman, 2015). The relationship between these two dominant categories are controversial. Some mix them, others keeps them separate. There exist different views on how one should approximate to these categories.

One perception is that research strategies exist on a scale instead of as contradictions, as Grønmo (Grønmo, 2016) suggests. Others voice a stricter view on the separation of research strategies. Bryman (Bryman, 2015), is an example of this. He argues that there exists multiple strong distinctions between quantitative and qualitative research strategies. Although Bryman has a stricter view than Grønmo, he also argues that one has to be careful not to completely separate the strategies.

Bryman (Bryman, 2015) argues that quantitative research is concerned with *generalization*, *causality*, and *measurements*. Based on his descriptions; generalization entails that one is able to apply research findings to a greater context outside of the small area which is subject to the research, causality is related to drawing connective lines between different factors or phenomena, and measurements are to a degree self-explanatory, but in social research measurements has a tight connection to quantification and concepts.

Quantification is a central focus of quantitative research. Quantification¹⁹ involves determining the quantity of something measurable. *Something measurable* is often *concepts* in social research, and the practice of *measuring* has a central position in this type of research (Bryman, 2015).

Concepts, as used in social research, can be described as categorizations of meaningful patterns or features that affect the world, as concepts are used to understand, categorize, and describe the world we live in (Bryman, 2015). Baxter et al. (Blaxter, Hughes, and Tight, 2010) describe concepts as a broad term relating to the classification of objects.

To measure a concept, indicators are needed. Indicators are smaller parts of a concept that are measurable. The indicators for the measurement of concepts are not predefined and have to be identified, and there are numerous ways of doing this. One way is to use a number of questions in a questionnaire, where each answer to a question becomes a measurable indicator, and using a Likert scale for questions gives a multiple-indicator measurement (Bryman, 2015).

Bryman (Bryman, 2015) also connects quantitative research to *positivism*, *objectivism*, and *deduction*. A brief account of these positions is explored as they entail premises for the quantitative research. Positivism is related to an epistemological position – what knowledge is adequate for a discipline. Objectivism is related to an ontological position – social ontology is related to how social objects are conceptualized (Bryman, 2015).

¹⁹<https://www.oed.com/view/Entry/155919>

Positivism links social research to the natural sciences (Bryman, 2015). Positivism is inspired by the methods used in natural sciences. It entails a view of the social world as something that can be explored and measured similarly to the physical world (Denscombe, 2017). An example of this is the use of quantification.

Objectivism implies that social phenomena exists outside the minds of people. This entails that social phenomena can be studied as entities independent of people (Bryman, 2015), and objective facts can be stated about these social phenomena.

Deduction describes the relationship between theory and social research (Bryman, 2015). It is related to how theory evolves. In its simplest form: (1) a hypothesis is formed based on theory. (2) Hypothesis proves right. (3) Theory is revised to accommodate the new findings. The opposite of deduction is induction.

Deductive and inductive approach. This research leans more towards an inductive approach than a deductive approach. The reason for this is that there is little theory that exist for privacy in learning analytics. In learning analytics, there are some frameworks of how privacy should be handled, but these have yet to become “theory” through broad consensus. Without established theory a deductive approach is not very useful, as the key element of testing theory is missing. Thus, induction is more natural direction to lean. In induction, findings are used to generate theory (Bryman, 2015). Findings from this research will contribute to the pool of knowledge for learning analytics and potentially be used to develop theories. The goal of this research, however, is not to develop new theories, but to gain new insights. How these insights are utilized, is dependent on future research.

Epistemological position. The learning analytics research field does not seem to have a dominant epistemological position. Related research is split between qualitative and quantitative research methods, entailing different epistemological positions. This can indicate that multiple positions are viewed as equally acceptable in the learning analytics research field, or that not a dominant position has taken form yet. The field of learning analytics is complicated as it is an interdisciplinary field with a wide research focus (Misiejuk and Wasson, 2017).

Ontological position. The ontological position of learning analytics researchers is rarely documented. Perhaps it remains to be decided what the convention will be as the research field is relatively young. It can also indicate that learning analytics does not have a strong connection with social research, even though it uses social research methods. This research leans towards constructionism, rather than objectivism. Constructivism is the position that social phenomena, and the meanings associated with the phenomena, are constructed by people (Bryman, 2015). Despite Bryman’s connection of quantitative research to objectivism denoted above, constructionism is the position that harmonizes best with this research. This is because the theories used in this research do not correspond well with objectivism. Solove’s (Solove, 2009) privacy theory is an example of this. As discussed, he views privacy as a concept in constant motion, changing together with society, and influenced by the opinions of the people. This does not fit the ontological position of objectivism.

The topic of epistemological and ontological positions is not dominant in learning analytics research, thus little weight is put on these considerations. Other methodology authors (Bell and Waters, 2018; Blaxter, Hughes, and Tight, 2010; Denscombe, 2017), besides Bryman, only briefly mention epistemological considerations, but do not go into detail. Ontological considerations are not explored actively by these authors either. This can limit the importance of these positions. Bryman (Bryman, 2015) argues that it is impossible to separate social ontology and social research as it entails the nature of the research. For this reason, the positions are discussed but not given a lot of attention.

The discussions on different theoretical positions must be seen in the context of classical social sciences. Learning analytics, as a new research field, draws on different disciplines (Misiejuk and Wasson, 2017) such as educational research and computing, and does not seem settled on its position in the social research terrain. Perhaps it is a “guest”, borrowing methods that fits their use, without the commitment to philosophical positions. Conventions are still in development, which allows for more free exploration and combinations of methods. In respect to this, this research has exploratory aspects as well. Grønmo (Grønmo, 2016) argues that selection of research strategy should be strategic opposed to being anchored

in principles. The research strategy adopted for this research and the methods that comprise it, were chosen based on the belief that they will provide important insight into the topic at hand. The research strategy is also similar to the strategies used by comparable learning analytics research. In this way the research fulfills the expectations of the research community to which it belongs (Denscombe, 2017).

Based on the section above, *quantification*, *concepts*, and *indicators* are important concepts for quantitative research. *Privacy* is the concept that will be explored in this research. *Privacy principles* will be the indicators that measures privacy. The results of the measurements will be quantified and presented. This is done using a questionnaire and descriptive statistics. A questionnaire is a method utilized in survey research, and survey research is a subcategory of quantitative research.

4.4 Survey research

Denscombe (Denscombe, 2017) describes survey research with three key aspects: *Empirical research* – collecting peoples’ opinion; *Temporal* – limited area in time; and, *Comprehensive coverage* – exploring all relevant aspects.

There are multiple advantages to the use of surveys, some of which can be deducted from the key aspects of Denscombe. For example, comprehensive coverage can be an attractive trait. Not all advantages of survey research are relevant for this research. The main advantage in using surveys for this research, is that conducting surveys does not entail fieldwork (Blaxter, Hughes, and Tight, 2010). This is a great advantage when doing research during a pandemic, as this research is. Furthermore, using a questionnaire makes it possible to reuse the questions in future research (Blaxter, Hughes, and Tight, 2010).

Denscombe (Denscombe, 2017) argues that the most used survey research methods (in social research) are related to asking questions. Blaxter et al. (Blaxter, Hughes, and Tight, 2010) also supports this position. Still, there exist methods of surveying that obtains information from other sources than explicitly asking people, such as documents, official statistics, or structured observation (Bryman, 2015; Denscombe, 2017).

Both of these survey techniques will be used in this research. A questionnaire will be used to gather the opinion of students (asking questions), and a secondary analysis of previously collected questionnaire data will be conducted to measure how the Norwegian population relates to privacy (using (semi-)official statistics).

4.5 Questionnaire

Questionnaires are used to gather the opinions of people. It is one of the most popular research techniques in social science (Blaxter, Hughes, and Tight, 2010). The goal is to collect data that can later be analyzed (Denscombe, 2017). Distribution of questionnaires can take many forms, for example through the post, face-to-face, or as web-based questionnaires.

Denscombe (Denscombe, 2017) identifies three criteria for a good questionnaire: *response rate*, which is related to how many from the invited population, responded to the questionnaire; *Completion rate*, related to how many of the respondents completes the questionnaire; and, *Validity*, which is related to validity of responses (if they are truthful or not). Many factors need to be considered to fulfill these criteria in a satisfactory way. In addition to these criteria, there exists a plethora of other advice and *best practices*, when it comes to the development and conduction of questionnaires. The best practices can depend on what type of questionnaire is used.

For this research, a web-based self-administered questionnaire will be administered. **Web-based** means in this context that the questionnaire will be created using a web-application and the responses will be recorded digitally through the webpage. **Self-administered** entails that respondents have to fill out the questionnaire independently without supervision (Bryman, 2015; Denscombe, 2017).

There are certain advantages with conducting a web-based self-administered questionnaire. The amounts of resources needed to conduct a questionnaire that is self-administered is low, as it is inexpensive to

administer (Bryman, 2015; Denscombe, 2017).

Web-based questionnaires are perhaps most importantly accessible (Denscombe, 2017). Web-based questionnaires can be accessed by a link to a webpage and because of this, is easily available to large number of persons. Combined with being self-administered, the accessibility increases further. The questionnaire can be answered whenever respondents wants. This can also have a potential negative affect if respondents postpone answering the questionnaire, and ultimately forget about it. Reminders for answering the questionnaire could, in this case, be sent out to participants.

Bryman (Bryman, 2015) highlights the convenience self-administered questionnaires entails for participants. In addition, he points to the advantage of not having a researcher overlook the answering process. As opposed to interviews or face-to-face questionnaires, the respondents do not have to deal with another person when answering the questionnaire. This is valuable for the truthfulness of the answers, as interviewers overlooking questionnaire-answering can potentially affect the answers respondents give.

Bryman (Bryman, 2015) also lifts forward a related advantage web-based questionnaire have over telephone administered questionnaires. He points to evidence indicating that respondents are more truthful when reporting on sensitive topics in web-based questionnaires, compared to telephone-based questionnaires.

There are many reasons for choosing a questionnaire for the gathering of empirical data, as has been described above. For this research, a questionnaire was first and foremost selected because it is believed to yield valuable results for the research, that can be used to answer the research questions. In addition to the other advantages mentioned, ethical considerations is of importance for choosing the method of web-based self-administered questionnaire.

Ethical considerations. When designing a questionnaire, certain ethical considerations has to be made. This is important when designing the questions, conducting the questionnaire, and when handling the responses. These considerations are related to the sensitivity of the topics, respondents' informed consent, and the anonymity of respondents (Denscombe, 2017).

The most safe option for the anonymity of respondents is to opt for a totally anonymous questionnaire. This will be done for the questionnaire in this research. Total anonymity guarantees that no personal information about respondents are recorded, this includes not recording the IP-address of respondents. It also entails that no background information about the respondents is recorded, which can be a limitation.

Respondents might feel more comfortable with voicing their true opinion when guaranteed anonymity, this is also related to validity of responses, as truthful responses prevent skewed answers (Denscombe, 2017). This can be done by assuring anonymity, as well as eliminating unnecessary questions and potentially reducing the sensitivity of question topics (Denscombe, 2017). Handling of data is also easier when it is anonymous, as one is not under the responsibility of handling personal information which limits handling to web-applications or university hardware (NSD recommendation). Conducting a totally anonymous questionnaire can entail limitations of the findings in relation to categorization, as no background information is recorded. There is also the possibility of duplicate responses with the use of total anonymity.

Obtaining informed consent for participation in a questionnaire is considered good practice (Bell and Waters, 2018). In Norway, this is required if the questionnaire administration involves any form of personal information handling. When not handling personal information, the use of a consent form is still recommended (Figure A.1, appendix). Providing students with information about the questionnaire is also viewed as an ethical responsibility (Denscombe, 2017).

Even if all formal requirements for ethical research is covered, the topic or single questions can entail unforeseen consequences or skewed responses (Denscombe, 2017). This can then jeopardize the validity of the results. Making the respondents feel comfortable when answering the questionnaire is important (Denscombe, 2017). This will be discussed in Chapter 5, as it describes all choices taken when developing the questionnaire.

4.5.1 Questionnaire development

To understand the considerations that need to be invested in the development of the questionnaire, its scope needs to be fresh in mind. As mentioned previously, to measure concepts, they need to be broken down into measurable pieces called indicators. The indicators used to measure privacy, is privacy principles. The principles will be discovered through the literature review, as described previously.

The questionnaire will be designed in accordance with principles of questionnaire design. A plethora of questionnaire design principles exist. The principles used for this research is based on the most prominent ones from multiple questionnaire-methodology sources (e.g., Bell and Waters, 2018; Blaxter, Hughes, and Tight, 2010; Bryman, 2015; Denscombe, 2017). The following topics will be elaborated on in detail in the next chapter; anonymity, introductory information, language, length, clear presentation, types of questions, formulation of questions, justification of questions, expert review, sampling, distribution, reliability, and validity. The tool that will be used for the development of the questionnaire is called SurveyXact.

Questionnaire tool. SurveyXact²⁰ will be used to create the questionnaire. The tool supports totally anonymous questionnaires and is endorsed by the University of Bergen²¹. SurveyXact support distribution and analysis in addition to the creation of questionnaires. Only limited functionality of the tool will be used. Distribution of the questionnaire will be handled by other software-solutions and the need for a more customizable analysis requires other analysis software.

4.5.2 Expert review

The questionnaire will be reviewed by three experts from the learning analytics research field. The expert review will be used as a pilot-test of the questionnaire. Pilot testing is used for a multitude of purposes (Bryman, 2015). Arguably the most important is to get feedback on the questionnaire regarding the flow, correctness, and understandability of questions (Bryman, 2015). Experts can check this and also check the theoretical foundation of the questions. Expert review and a pilot questionnaire differ because pilot questionnaires should use individuals that are comparable to the sample which is going to be studied (Bryman, 2015). The reason for choosing an expert review is opportunistic.

The reason for using a pilot-test or an expert assessment is to assess the reliability and validity of the questionnaire. Reliability and validity are both big topics within social research methodology, and there are multiple different ways of promoting reliability and validity.

4.5.3 Reliability and validity

Reliability relates to consistency (Bryman, 2015), this is that the same results are found when redoing the research. For a measurement to be useful it needs to measure consistently. A number of ways exist of measuring reliability. None of Bryman's (Bryman, 2015) three types of reliability; stability, internal reliability, and inter-rater reliability, will be tested for in this research. Bryman highlights the importance of reliability testing in social research although he acknowledges that stability-testing is often not done as it is complicated and requires a lot of resources.

Validity is a concept closely connected to reliability, but not the same. Validity is related to the question: "Does a measurement correctly measure the concept it attempts to measure?" (Bryman, 2015). Validity is also related to measurements. There exist multiple measurement-validity-tests. In its simplest form, measurements are reliable if they correctly measure what they are meant to measure (Bryman, 2015). Bryman (Bryman, 2015) describes validity as dependent on reliability, making the concepts interconnected. Measurement validity is relevant for this research, as the questionnaire will measure privacy. Measurement validity for this questionnaire is then related to what degree the questionnaire will be able to measure the privacy concept. Face validity, the simplest form of validity, according to Bryman, will be assessed for the questionnaire, through the use of the expert review.

²⁰<https://www.surveyxact.no/>, accessed: 13.03.21

²¹https://it.uib.no/Sp%C3%B8rreunders%C3%B8kelsen_-_verkt%C3%B8y_for_utvikling,_utsending_og_analyse, accessed: 13.03.21

Internal validity is usually weak in survey research designs according to Bryman (Bryman, 2015). Internal validity is related to causality. Causality is present in a relationship of two (or more) measurements/variables if one causes the other (Bryman, 2015). Internal validity is then concerned with the correct assessment of causality between different measurements or variables. Causality relationships will not be a topic of the following analysis and are therefore not relevant. External validity is generally strong for survey research designs if a representative sample is used (Bryman, 2015). External validity is concerned with whether the results of a study is generalizable beyond the participants it studies, as Bryman describes. Generalizability and having a representative sample are thus important keywords for producing valid findings.

Another topic related to reliability and validity is sampling. Sampling is selecting a subgroup of the population, and is necessary when it is unfeasible to survey a complete population (Bryman, 2015). *Population* refer to the group of units one wants to research and make statements about (Bryman, 2015). For this research the population will be students at the University of Bergen. No sampling technique will be used as the whole population will be invited to answer the questionnaire.

Avoiding using a sample does, however, not eliminate potential sampling errors. Bryman (Bryman, 2015) gives examples of three types of sampling errors: (1) lack of random sampling usage, (2) insufficiency of sampling-frame, and (3) refusal of participation. The 3rd is the most relevant for this questionnaire. Refusal of participation entails that a certain type of respondents does not wish to answer the questionnaire, skewing the results, according to Bryman.

Generalizability is the main reason for being concerned with representative samples, reliability and validity. Generalization, as mentioned earlier, entails that one is able to apply research findings to a greater context outside of the population subject to the research (Bryman, 2015). For this research, the population is as mentioned, students at the University of Bergen. To get generalizable results is considered one of the biggest challenge with administering a questionnaire, because of the low expected response rate (Bryman, 2015). Measures will be taken when designing the questionnaire to increase the chase of a good response rate, this will be discussed in Chapter 5. Generalizable results are easier to accomplish when performing secondary analysis.

The scope and time frame of this research prevents the use of reliability and validity testing outside of assessing face validity, through an expert review.

4.6 Secondary analysis

Secondary analysis is analysis based on data previously collected by others (Bryman, 2015). Secondary analysis will be used in this research. The data that will be used for the analysis was collected for the NDPA's privacy questionnaire 2019/2020²². Not all of the data is relevant for the following research, the relevant questions have been extracted and are presented in Table 9. While NDPA has published a report²³ based on the data from their questionnaire, this will not be read, in order to be neutral in the analysis of the data. The data was collected by Opinion for the NDPA. Opinion is a well-known and respected company, specialized in administering questionnaires²⁴.

There are certain advantages with the secondary analysis method, one of the biggest advantages is the quality of the data (Bryman, 2015). Professionally collected data fulfils the requirements of reliability and validity that is hard for a lone researcher to match. Another advantage it that the questionnaire answers are a source of empirical data. As there exists little research in Norway on student perceptions of privacy and learning analytic, a supplement of empirical data from related research is valuable. The secondary analysis gives additional background for the results provided by the administered for this research, and because of design decisions of the questionnaire, it will also be a foundation for comparison. Still, there are some limitations with using secondary analysis.

²²<https://www.datatilsynet.no/regelverk-og-verktoy/rapporter-og-utredninger/personvernundersokelser/personvernundersokelsen-20192020/>, accessed: 13.03.21

²³<https://www.datatilsynet.no/link/f6ecf9dfc5e04569a595f727b3269b39.aspx/download>, assecced: 13.05.2021

²⁴<https://opinion.no/selskapet/>, accessed: 13.03.21

Table 9: Questions that are the topic of the secondary analysis.

Id	Question
Q1	To what degree are you concerned with privacy?
Q2	To what extent do the following descriptions fit your view?
Q2.1	- I feel I have little control over how personal information about me is stored and used on the internet.
Q2.2	- I feel powerless when it comes to having control over personal information about me on the internet.

4.6.1 Limitations

Bryman (Bryman, 2015) argues that a lack of control over data quality can be a limitation with secondary analysis. This is of little concern in this research, as the data was collected by a professional company. Other limitations relate to the lack of variables, unfamiliarity with the data, and the complexity of data, according to Bryman. Lack of variables is a relevant challenge as most of the data is not relevant for this research. Complexity of data is not a challenge as the data is non-complex. The simplicity of the data on the other hand, can be a limiting factor, as the data is in an aggregated format (percentages), and therefore makes an analysis limited. Having aggregated data is limiting compared to having complete records of respondents' answers, as patterns in responses cannot be highlighted. Although a limitation, this will not be considered a challenge for this research as the data analysis will take the form of descriptive statistics, which is a non-complex analysis method.

4.7 Method of analysis

In this section the method of analysis is described. Descriptive statistics will be used to analyze the questionnaire data and the secondary analysis data. Descriptive statistics is selected as the results of this type of analysis are easy to comprehend and can be presented in understandable charts.

4.7.1 Descriptive statistics

Descriptive statistics is a low-level analysis method for quantitative data (Blaxter, Hughes, and Tight, 2010). The descriptive statistics for this research will consist of variable frequencies (percentages) and averages (arithmetic mean). The averages will be displayed in plain text while the variable frequencies will be displayed in charts. Standard charts for descriptive statistics (Bryman, 2015) will be used, such as: pie chart and bar chart. In addition, two more complex charts will be used: polar bar chart and heatmap.

The descriptive statistics will be produced, mostly, using openly available software solutions, which is heavily reliant on the Python programming language (Python 3.8). The results of this analysis will be presented in Chapter 6. The analysis requires the use of various software tools.

Tools PyCharm Professional, Pandas, Plotly, and Affinity Designer, are the tools that will be used for the descriptive statistics. Python libraries (Pandas and Plotly) are the main tools used for the descriptive statistics, based on both the questionnaire data, and the secondary analysis data. In addition, a digital vector graphic tool (Affinity Designer) will be used to customize the visualizations produced by the descriptive statistics. This provides more freedom of analysis and visualization, compared to the features offered by SurveyXact. This also provides the opportunity to use the exact same methods on the two datasets, as the same tools will be used. This prevents having to use SurveyXact to analyse the questionnaire data and the other tools only for the secondary analysis data. Details of the tools and what their purposes are described below.

PyCharm Professional is a Python Integrated Development Environment (IDE) provided by JetBrains²⁵. PyCharm Professional is a powerful tool for working with the Python programming language. It has all the necessary features for the relevant analysis tasks. PyCharm Professional will be used because of its graphical user interface, its integrated Git functionality, and its clear directory handling, which is important for handling the charts that will be generated by Plotly.

Pandas is a Python package that is focused on data analysis²⁶. Pandas will be used to generate descriptive statistics of the questionnaire data and secondary analysis data. The `pandas.DataFrame.describe()`²⁷ function will be the most important feature. Other aggregation functionality will also be utilized, such as counting the different categories of answers in Likert scale questions. It was decided to use Pandas for the analysis as it provides fitting features that are easy to implement.

Plotly²⁸ is a Python library designed for creating charts. Plotly will be used to generate charts for each of the questions in the questionnaire data and the secondary analysis data. It will be used to generate charts based on the statistics generated by Pandas. Plotly relies on other Python packages for features such as exportation of charts. These dependencies are accounted for in Plotly's documentation, which will be the main source of guidance in generating and exporting charts.

Affinity Designer²⁹ is a vector graphic design tool (it also supports raster graphics). Affinity Designer will be used to customize the charts generated by Plotly. This will be done to give the charts a uniform design and to make a clearer presentation of the charts. Affinity Designer will be used to resize the charts, change fonts, and change colors. Other small visual tweaks will also be fixed using this tool. Affinity Designer was chosen as it is an affordable alternative to Adobe Illustrator.

4.8 Summary

This chapter has presented the research strategy and the methods of empirical data gathering, in addition, methods of analysis, for this research. The methods that will be used for empirical data; questionnaire and secondary analysis, are well known methods for social research. Questionnaires are also a prominent method of gathering student opinions of privacy in learning analytics. The analysis methods that will be used, descriptive statistics, also has a prominent position in related learning analytics research.

The questionnaire will be used to gather student opinions on privacy in learning analytics. These results will be supplemented by results of secondary analysis. The secondary analysis will be performed on data from the NDPA's 2019/2020 privacy questionnaire. The results this generates, will be analyzed using descriptive statistics. The descriptive statistics will be represented in textual form as well as in charts. This results of the analysis is described in Chapter 6, while in the next chapter, Chapter 5, the questionnaire development is described.

²⁵<https://www.jetbrains.com/pycharm/>, accessed: 09.04.21

²⁶<https://pandas.pydata.org/>, accessed: 09.04.21

²⁷<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>, accessed: 09.04.21

²⁸<https://plotly.com/python/>, accessed: 09.04.21

²⁹<https://affinity.serif.com/en-us/designer/>, accessed: 09.04.21

5 Questionnaire development

In this chapter, design decisions related to the development of the questionnaire will be discussed. Decisions related to questionnaire design choices was taken based on questionnaire design principles from multiple methodology authors. As the decisions entail both advantages and disadvantages, the decisions are discussed as they are described. The chapter will start by exploring the many design decisions taken for the questionnaire development. A web-based self-administered questionnaire containing **9 items, with 37 subitems**, was developed for this research. The nine questions included will be described and discussed. The questions addresses privacy in general, privacy in learning analytics, and services and benefits of learning analytics, in addition to questions concerning consent and Terms and Conditions agreements. The questionnaire was distributed to all students at the University of Bergen via SMS. The final questionnaire-product is reproduced in Figure B.1-B.7 (Appendix B), as it was distributed. The implementation of the questionnaire will also be accounted for, before highlighting some limitations to questionnaires is general, and specific limitations to this questionnaire. The chapter ends with a summary.

5.1 Design

This section discussed design decisions for the questionnaire. All decisions have been inspired by design principles discussed by methodology authors, as described in Chapter 4. Decisions related to the following topics will be explained: anonymity, cover letters, introductory information, language, length, clearness of presentation, types of questions, Likert scale. In addition formulation of questions including all the decisions this entails and display and justification of all the nine questions included in the questionnaire. But first, why it was important for this questionnaire to be totally anonymous despite the limitations this entails.

5.1.1 Anonymity

The questionnaire was developed to be totally anonymous. This entails that no background information was recorded, this can have limitations in relation to the value of the results. These disadvantages are related to not knowing precisely who is behind the responses to the questionnaire, which allows for duplicate responses as well as the loss of opportunity to examine background information. Based on the amount of subitems (37) and the clicks needed to finish the questionnaire (more than 40), it is unlikely that respondents maliciously answer to the questionnaire multiple times, also, as this research do not intend to categorize findings based on background information (gender, socio-economic factors, etc.), loss of background information is of little concern for the results. A totally anonymous questionnaire also entails some advantages.

Because of the pandemic there were great uncertainty related to access of university facilities during the course of this research. This jeopardized the use of university-hardware to analyze respondents' data, which is a recommended practice by the Norwegian Centre for Research Data (NSD). This reason weighted heavily when deciding to opt for a totally anonymous questionnaire.

As a consequence of this choice, in conversations with NSD, it was ruled unnecessary to apply for approval of the questionnaire, as it did not record any personal information (Figure A.1, Appendix A).

Because the questionnaire explored the topic of privacy in learning analytics, total anonymity also supported the integrity of the questionnaire. A questionnaire on privacy that is not "private" or anonymous might seem hypocritical and deter respondents from answering. This was not a determining factor, but still an interesting consideration.

5.1.2 Cover letter

Information on the project were given to all participants as a cover-letter before the questionnaire started. The cover letter is documented in Figure B.1 (Appendix, A). The cover letter was formed based on an NSD-template³⁰. In this cover-letter, contact information to the project supervisor and responsible

³⁰Template can be downloaded from: <https://www.nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/sjekkliste-for-informasjon-til-deltakerne/>, accessed: 13.03.2021

researcher was included, as well as information about the questionnaire. This included what research project the questionnaire was associated with, as well as information about the scope of the questionnaire.

5.1.3 Introductory information

To give respondents the knowledge needed to answer the questionnaire (Bryman, 2015), a brief clarification was given prior to the questionnaire start. Many of the questions revolve around the collection, processing, and analysis of personal information. Thus it is required that the respondents know what is meant by *personal information*. Without this information, respondents might be uncertain of what the question really is about or have misguided conceptions of what personal information is. A distinction was made between *personal information* and *sensitive personal information*, underscoring that the questionnaire is only related to the former, and not the latter. (see Figure B.2, Appendix B, for this clarification). The clarification is based on information from NDPA (Datatilsynet, 2019).

5.1.4 Language

The questionnaire was developed in English, because of the technical terms used in the questionnaire, and with the possibility of translation loss, this was ruled necessary. When it comes to terms it is always difficult to know if individuals understand them in the same way; this is independent of language. Because of this, questions including technical terms are accompanied by an explanation (see Table 10). The practical reasons for having the questionnaire in English outweighed the potential benefits of having it in Norwegian. Still, it can be considered a limitation to conduct an English questionnaire in a Norwegian speaking country, as it will probably impact response rate in a negative way.

5.1.5 Length

The questionnaire was kept as short as possible. This was to heighten the chance of getting respondents to fill out the questionnaire (Denscombe, 2017). The length of the questionnaire is comparable to similar questionnaires used by other learning analytics researchers. Test runs of the questionnaire indicated that it could be completed within 12 minutes, with a medium reading speed. This included the reading of the attached invitation and introductory information. As mentioned, the questionnaire ended up containing 9 items with 37 subitems. 37 subitems are not few, but the subitems are fast to answer as they are connected to a main question and answered in an identical manner. An example of this is question 4/9, which is displayed in Figure 14, and Section 5.3.4. Here a main question is asked: “*To what extent do you agree with the following statement?*”, this is followed by four sub-items containing different statements. The statements are all answered in the same manner; by indicating how much one agrees with the statement, using a radio-button beside the question. This provides a fast and easy design to answer multiple similar questions. When wrapped in a nice presentation it possibly enhances the simplicity and flow of questionnaire execution.

5.1.6 Clear presentation

The questionnaire was designed to have a clear presentation. Not having this can deter respondents from answering, resulting in even lower response rates (Bryman, 2015). The questionnaire utilized a SurveyXact-template made for the University of Bergen. This provided a clear presentation of questions. The design includes the university’s logo and color scheme, this provided a clear indication of whom the questionnaire was associated with, in addition to providing a pleasant design. The questionnaire webpage was dynamic, making the presentation fit both big and small screens (e.g., desktop-computers, and smartphones). A screenshot of the design is given in Figure 14 and 15. The example is given for big, medium, and small screens.

Having a dynamic webpage design supports accessibility and convenience. Dynamic webpage design is also a highly regarded feature when distribution of the questionnaire is via SMS. Because of this, most students will answer the questionnaire on their smartphone. This feature is expected from most questionnaire service providers and can be considered standard today. Still, it is worth mentioning the value of this feature. Bryman argues that, at the time, smartphones had shown promising results in relation to self-administered questionnaires (Bryman, 2015).

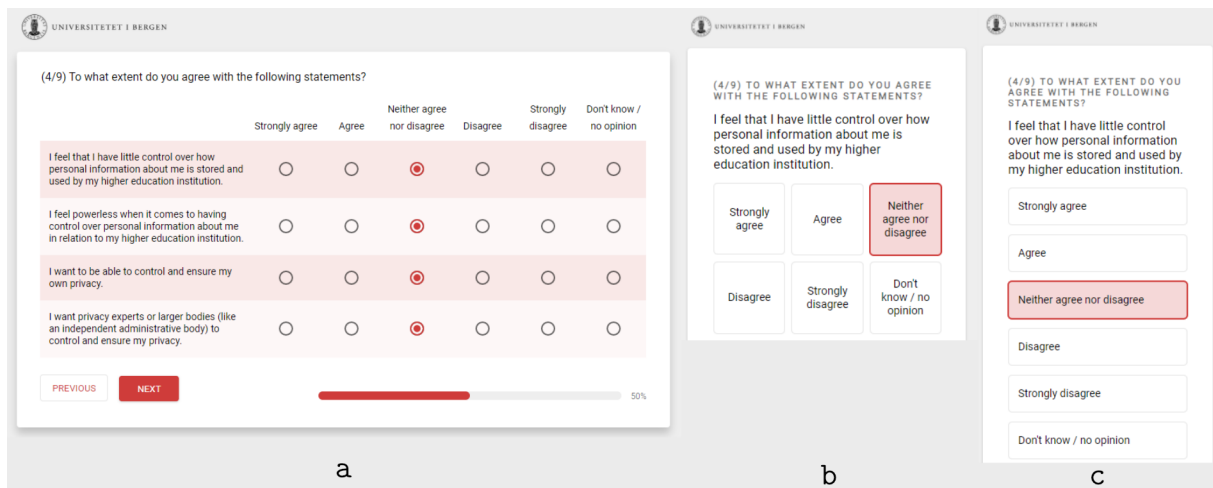


Figure 14: Likert scale question on a big (a), a medium (b), and a small (c) screen.

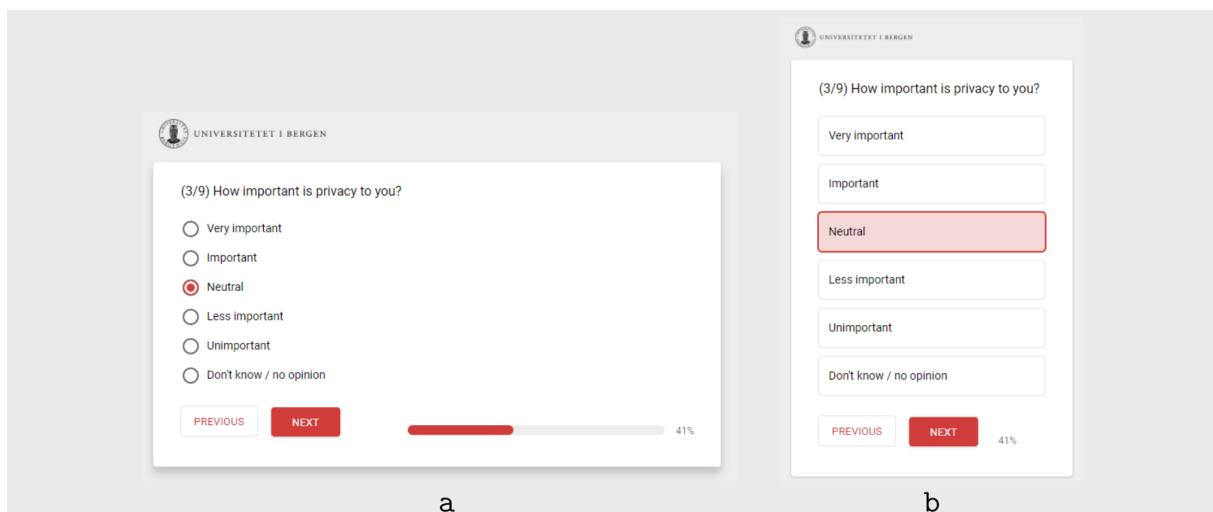


Figure 15: Questionnaire question as displayed on a big (a) and a small (b) screen.

5.1.7 Question formats

The questionnaire contained questions about attitudes and beliefs. These are two categories of questions explained by Bryman (Bryman, 2015). Attitudes are most frequently recorded using Likert scale type questions, according to Bryman. The Likert scale will be discussed later, but in order to use the Likert scale; closed questions are required.

Closed questions are questions with a selection of predetermined answers that the respondents can choose from (Bryman, 2015). Bryman argues that there are advantages and disadvantages with this closed questions. A disadvantage can be that respondents are not able to use their own words, which more precisely can express their opinion, while an advantage is that processing and analyzing responses is highly efficient when answers are pre-coded. Getting answers to complicated questions is possibly easier with closed questions. Question 2/9 (Section 5.3.2) from the questionnaire is used to illustrate this example.

If this question was in the format of an open question, it would be the following: *“How do you understand privacy?”*. In a self-administered questionnaire it is too much to expect a well considered answer to such a difficult question. Based on Solove’s (Solove, 2009) arguments, even privacy scholars are not able to create a good comprehensive definition, thus, it is better to present this as a closed question. In the questionnaire, this question had the following form: *“Which of the following statements is closest to how you understand privacy?”* followed by alternatives comprised of well-known and widely used privacy definitions, in addition to providing an opportunity to give an open answer if the pre-determined ones

not fit their preference. Providing an open alternative can prevent false results.

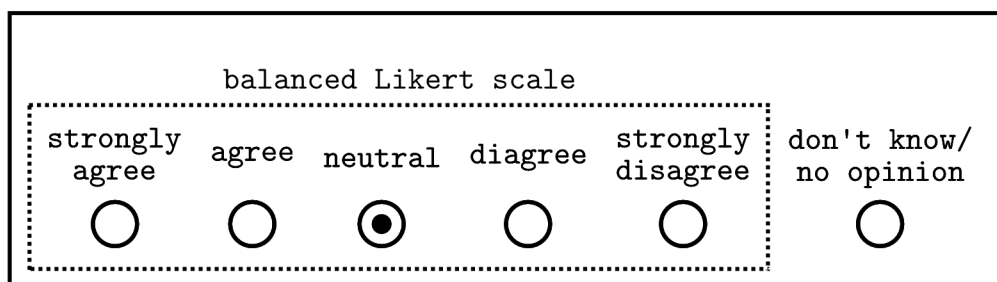
False results can appear when the respondents' opinion is not covered as an alternative in a closed question. This forces the respondent to select an option that is not in line with their view. Thus, another preventative measure that was used in the questionnaire was the addition of a *don't know / no opinion* option to all questions, with one exception. The exception is the question mentioned above, where the respondent can approximate to one of the alternatives or opt for the open answer option.

Having a *no response* option is controversial as it provides an easy way out for answering complicated questions (Bryman, 2015). This can result in passive responses, containing neutral answers. Still, it needs to be considered that the true opinions of respondents can also truly take the form of passive responses. Perhaps the respondents have never thought of these questions and do not have an opinion on them. When weighing the alternatives for the questionnaire, the potential prevention of false results weighted heavier than the potential of getting passive responses. This discussion is similar to the discussion of what the middle alternative should be in questions using scales, such as Likert scale (Bryman, 2015).

5.1.8 Likert scale

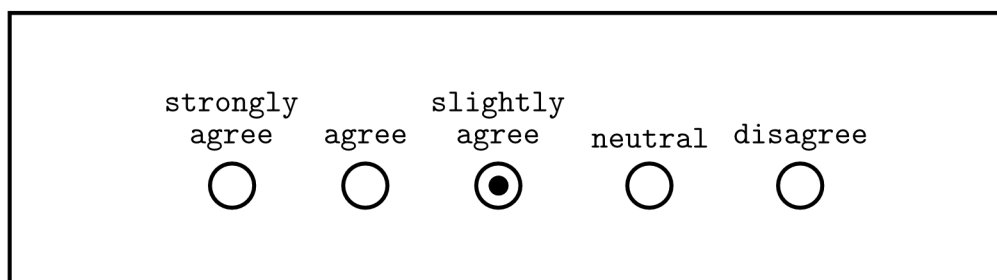
Likert scale questions were used to a great extent in the questionnaire. A "Likert scale question" is a question accompanied with a Likert scale. Respondents provide their answer by placing it on the scale (Denscombe, 2017). The scale is comprised of a set of predetermined answers. The predetermined answers exist between two extreme and opposite ends (Denscombe, 2017). An example of a Likert scale used in the questionnaire, is displayed in Figure 14. It is important that the scale is balanced (Bryman, 2015). Figure 16 shows a conceptual model of the Likert scale used for multiple questions in this questionnaire. It also shows an example of an unbalanced scale to highlight the difference.

balanced Likert scale with "don't know / no opinion" option



a

unbalanced Likert scale



b

inspired by: Bryman, Alan. Social research methods. Oxford university press, 2016. (p. 256) & Denscombe, Martyn. The good research guide: for small-scale social research projects. McGraw-Hill Education (UK), 2014. (p.196).

Figure 16: Balanced (a) and unbalanced (b) Likert scale.

The questionnaire deployed a five-point Likert scale for most of the questions (3–8/9). The scale was balanced. The answers was recorded using a radio-button, preventing the selection of more than one alternative. It also contained an option of *I don't know / no opinion* outside of the scale. In addition,

the scale included a neutral middle alternative.

The discussion of middle alternative is about whether to include a neutral middle alternative on the predetermined answer scale (Bryman, 2015). The concern of including a neutral middle alternative is that it might cause respondents to go for the easy solution and choose the neutral alternative (Bryman, 2015). This can be viewed in connection with questionnaire fatigue.

Questionnaire fatigue describes a state respondents can be in where they are mentally tired of answering questions and opt for faster solutions to get out of the questionnaire (Denscombe, 2017). This can for example be to drop out of the questionnaire, answer randomly, or go for the passive opinions (I don't know / neutral) (Denscombe, 2017). Measures can be taken to prevent questionnaire fatigue, as was done in the development of this questionnaire.

The advantages of having a neutral middle alternative in the Likert scale is similar to the advantages of having an "I don't know" option. The choice of a neutral option just might represent the true opinion of a respondent, without this option it can generate false results (Bryman, 2015). The decision to include a neutral option was a well-considered choice for this questionnaire. It was decided that the potential for getting truthful responses was worth the trade-off of possibly getting more passive responses. The number of passive responses is difficult or impossible to assess prior to receiving the answers to the questionnaire, and needs to be considered in light of questionnaire fatigue. The higher the chance of questionnaire fatigue, the higher the chance of passive responses. As measures are taken to reduce this phenomenon, it was ruled beneficial to include a neutral middle alternative in the Likert scale, in the closed questions, as well as an *I don't know* option.

In the same debate, we find the question "*should respondents be forced to supply an answer?*", for this questionnaire, the answer was yes. None of the questions in the questionnaire were optional. There is a theoretical possibility that this can induce questionnaire fatigue, make respondents irritated, and result in false answers as described above. Considering the neutral alternative to the scale, and the *I don't know* option, this is unlikely. Obligatory answers to all questions was used to get full responses from respondents. This circumvents the problem of missing data (Bryman, 2015) caused by partially completed responses. Obligatory answers is a preventative measure against partial completion. Ironically, making answering all questions obligatory, can result in partially completed questionnaires, as respondents can be deterred from completion if they face a question they do not know how to answer. Again, as a neutral alternative, as well as a *I don't know* option was included, this can be prevented, but there is no guarantee that it works as planned. The only guarantee is that the respondents that complete the questionnaire provide answers to all questions, making them eligible for analysis. The trade-off between higher numbers of partially completed questionnaires and guaranteed fully completed questionnaires is worth it, as the latter strengthens the analysis.

The decisions regarding the structure of the questionnaire and the design of questions were made in order to make the questionnaire as easy as possible to answer, as this can prevent questionnaire fatigue (Denscombe, 2017). Although some questions handle complex subject matters, they were made as easy as possible. Terms are explained in simple language and the complexity has been reduced as much as possible. This was a balancing between easiness of completion and asking the questions that are interesting to the research. Respondents were also informed of the scope of the questionnaire, in the cover-letter, as this can also prevent questionnaire fatigue (Denscombe, 2017). Each question was labeled with its position (1/9–9/9), and a progress bar was at all time visible to the respondents.

Question 1/9 (Section 5.3.1) and 9/9 (Section 5.3.9) are of polar (yes/no) format. In case of the first, this question is only required a simple yes or no. For the latter a Likert scale is not used as the nuances it entails are not needed for this particular question. A *maybe* alternative was added to the polar accept/deny alternatives to allow for some uncertainty possibly preventing the frequent use of *I don't know* / *no opinion* option.

5.2 Formulation of questions

There exist many principles of good practice when formulating questionnaire questions. Bryman provides an extensive list of such guiding principles (Bryman, 2015). Other methodology authors also provide corresponding guiding principles (Bell and Waters, 2018; Blaxter, Hughes, and Tight, 2010; Dencombe, 2017). There is a lot of overlap in the principles discussed by these four authors. Bryman's list is the most comprehensive and was therefore used to guide the formulation of questions. Not all of the principles apply to this questionnaire, as they are related to specific forms of questions that are not utilized. Only the relevant principles are discussed. First of all is the importance of the research question.

The research questions guided the formulation of all questions in the questionnaire.

1. What privacy principles are most relevant for learning analytics?
2. How do Norwegian students perceive privacy in general, and in relation to learning analytics?
3. What privacy priorities do Norwegian students have for learning analytics?
4. What learning analytics services and benefits are acceptable to Norwegian students?

All of the nine questionnaire questions played a part gathering information that can be used to answer these research questions. All questionnaire questions relates to privacy or learning analytics, as discussed earlier, the only exception to this is the first question which asks the respondents to confirm that they are students at a higher education institution, in order to confirm that they are in the target group of the questionnaire, this was necessary because of the lack of background information recorded.

One of the questionnaire questions, contained technical terms, which is not recommended by Bryman (Bryman, 2015). This concern question 6/9 (Section 5.3.6). To be able to use privacy principles as indicators of privacy, the terms of the principles needs to be used. There is a high probability that not all respondents are familiar with the terms used in the questionnaire, and because of this, each term was accompanied with an explanation of what it means in this context. This eliminates the negative effect of using technical terms, but it makes the question-text much longer. These terms are discussed later (Section 5.3.6), when question 6 is discussed.

Long questions are another negative quality of questionnaire questions, according to Bryman (Bryman, 2015), and this especially applies to questions about attitude. The trade-off between explaining technical terms and having a long question had to be taken. Other ways of formulating the question 6/9 was explored, but was not possible without hurting the integrity of the question. The consequence of this was the possibility of fewer respondents, or more partially completed responses as respondents could drop out when they were faced with this long and bulky question. This is regarded a limitation of this particular question. The other questions were reasonably short and do not face the same challenge.

All the questions in the questionnaire have been checked for bias in the formulation. Avoiding leading questions is important for allowing respondents to answer truthfully (Bryman, 2015). One question in the questionnaire could be considered a borderline case, as it is possible to read in a negative way. This concerns question 4/9 (a and b) as it contains negatively charged phrases. If this question was negated, for example, by putting *not* in front, it would undermine another of Bryman's principles. He argues that having negatives of questions can make respondents miss the word *not* and wrongfully answer the question (Bryman, 2015). In addition to this, the question is derived from the NDPA's latest privacy questionnaire ³¹ (highlighted in Section 4.6), which gives some authority to the question, as it is designed by professionals.

This demonstrates the use of another questionnaire development principle: to use previously designed questionnaire questions (Bryman, 2015). A lot of effort goes into creating questionnaire question, therefore it can be a good idea to reuse previously designed questions (Bryman, 2015). This topic was also touched upon in the literature review. Here the SELAQ and SHEILA frameworks were highlighted as potentially useful to explore learning analytics in the Norwegian setting, as no experience with learning analytics is needed to answer the questionnaire they provide. SELAQ was considered utilized but the questions

³¹<https://www.datatilsynet.no/regelverk-og-verktoy/rapporter-og-utredninger/personvernundersokelser/personvernundersokelsen-20192020/>, accessed: 13.03.21

proved limited in privacy exploration for the depth desirable for this research, and were therefor not used.

5.3 Justification of questions

In this section, the questions included in the questionnaire are documented. The section is divided in nine subsections, each dedicated to one questionnaire question. The subsections follows the same structure. First the question is reproduced accompanied with its subitems, if relevant, this if followed by an explanation of why the question was included. The final questionnaire-product is reproduced in Figure B.1-B.7 (Appendix, B), as it was distributed. As mentioned previously, the questionnaire consisted of 9 questions with 37 sub-items.

5.3.1 Question 1 Student confirmation

Are you a student enrolled in a Norwegian higher education institution?

This question was included as a consequence of utilizing a totally anonymous questionnaire, as it does not collect any background information. If this question is answered with a *no* the respondent is taken to the end of the questionnaire and not allowed to answer any of the following questions. This is the only purpose of this question.

5.3.2 Question 2 Privacy understanding

Which of the following statements is closest to how you understand *privacy*?

- a) I understand privacy as the right to be let alone.
- b) I understand privacy as the state in which one is not watched or disturbed by others.
- c) I understand privacy as the right to limit access to myself as a person.
- d) I understand privacy as the right to decide when, how, and to what degree information about me is shared with others.
- e) I understand privacy as the right to safeguard my personal integrity; safeguard my opportunity for privacy life, self-determination (autonomy) and self-expression.
- f) I understand privacy as the state or condition of being alone, undisturbed, as a matter of choice or right; seclusion; freedom from interference or intrusion.
- g) Other - please enter (English or Norwegian).

This question was included to understand what understanding students have of privacy in general. This was done by questioning what definition of privacy lies closest to their understanding. This is important to examine, as it makes up a big part of how they conceptualize privacy. The answer alternatives (privacy definitions) were inspired from multiple sources. Solove's (Solove, 2009) explanations of historical privacy theories was the main inspiration, but the personvern definition (NOU:2009:1, 2009) was also included.

The list of definitions is not exhaustive as there exists an immeasurable number of privacy definitions and perceptions. The focus of this question was on the most known definitions.

5.3.3 Question 3 Privacy importance

How important is privacy to you?

This question was included to explore how much students care about privacy, in general. The results of this question will be compared to the view of the general population (secondary analysis). This can give interesting insights to whether students are more or less concerned with their privacy, compared to the general population.

5.3.4 Question 4 Privacy statements

To what extent do you agree with the following statements?

- a) I feel I have little control over how personal information about me is stored and used by my higher education institution.
- b) I feel powerless when it comes to having control over personal information about me in relation to my higher education institution.
- c) I want to be able to control and ensure my own privacy.
- d) I want privacy experts of larger bodies (like an independent administrative body) to control and ensure my privacy.

This question was included for comparison purposes in connection with the secondary analysis. These subitems a) and b), are inspired (slightly adapted) versions of questions that were part of the NDPA's 2019/2020 privacy questionnaire. They are reproduced in this questionnaire so as to be able to compare findings across the two questionnaires.

Subitems c) and d), were included to uncover how students view privacy self-management and privacy expert-management. The subitems will uncover if students want to control their own privacy or if they want other, larger, administrative bodies to handle their privacy. Or both, as these two can co-exist without any contradiction.

5.3.5 Question 5 Learning analytics desirability

How desirable is it for you to have your personal information collected, processed and analyzed by your higher education institution in order for them to provide services that has the potential to improve your learning outcome?

This question was included to examine students' desire for learning analytics applications. This is seen in connection with the *collection/use of personal data (HEI)* principle found in the literature review. This question will uncover if students want learning analytics and what this can implicate for the future of learning analytics in Norway.

5.3.6 Question 6 Privacy principles

Imagine that your higher education institution is collecting, processing and analyzing your personal information in order to improve your learning outcome. How important are the following principles to you?

This question contained 17 sub-items. Each of the sub-items address one privacy principle. An explanation of the each of the privacy principles are included in the questionnaire. Table 10 displays the different privacy principles asked about, and how they were explained to the respondents.

The principles were collected from learning analytics literature, through the literature review. The main inspiration for the principles were the quantification of the privacy debate. The results from this quantification was displayed in Table 4 and in Figure 11, 12, and 13.

Each principle is captured in a term. The terms were explained in the questionnaire, using definitions from the Online Oxford English Dictionary³² (OED). This was done because the definitions from the learning analytics literature was often lacking, and if present, their meanings could vary dependent on the literature. By defining the terms, it is made sure that the students have the same understanding of the principles they are asked about.

The definition of the principles was adapted to fit the context of learning analytics. The respondents are asked to visualize that their higher education institution is collecting, processing, and analyzing their

³²<https://www.oed.com/>. Terms searched: Accountability, accuracy, anonymous, awareness, consent, ownership, relevance, aware, control, preservation, purpose, security, and trust. accessed: 14.03.2021

Table 10: Privacy principles asked about in the questionnaire, and their explanations

	Principle	Explanation
a)	Accountability	That your higher education institution is liable to account for and answer for the collection, processing and analyzing of your personal information.
b)	Accuracy	That the information about you that is represented in your higher education institution's systems are accurate, precise and exact.
c)	Anonymity	That the information about you, in your higher education institution is represented in such a way that it is indistinguishable from others of its kind so that it cannot be directly connected to you.
d)	Autonomy	Liberty to follow your own will; control over your own affairs; freedom from external influence, personal independence.
e)	Awareness	To be informed. To know what personal information about you is collected, processed and analyzed by your higher education institution.
f)	Consent	That you are asked to agree to a request, before your personal information is potentially collected, processed and analyzed by your higher education institution.
g)	Data ownership	That you own the data that is represented in your higher education institution's systems about you.
h)	Data security	That your personal information is being protected from unauthorized access, the risk of being intercepted, decoded or tapped.
i)	Data sharing	That your higher education institution does not shares your personal information with others internally, with other companies or with other service providers.
j)	Data preservation	That your personal information stored in your higher education institution's system is deleted periodically.
k)	Limited access	That as few as possible are able to view, process and analyze your personal information.
l)	Opt-out	To have the option to choose to not participate in having your personal information collected, processed and analyzed by your higher education institution, without this affecting your learning outcome.
m)	Personal access	That you can view your personal information and how it is processed and analyzed by your higher education institution.
n)	Personal control	That you have authority over, and can determine what personal information about you is collected, processed and analyzed by your higher education institution.
o)	Purpose	To know the reason for why your information is collected, processed and analyzed by your higher education institution.
p)	Relevance	That only your personal information that is relevant to improving your learning outcome is collected by your higher education institution.
q)	Trust	That you have firm belief in, can rely on, and have confidence in your higher education institution to protect your privacy.

personal information (the core processes of learning analytics). With this backdrop, they are prompted with questions regarding the importance of different privacy principles. The results of this question will indicate students' attitude towards privacy in the context of learning analytics. In this way learning analytics was asked about without students needing to have any experience with it.

17 principles were included to have a broad representation of principles, as to get a nuanced view of how students view privacy in relation to learning analytics. Not all principles from the literature review were included, as not all principles made sense to ask the students about. An example of this is the principle of lawfulness, to ask if they think lawfulness is important is unnatural, as there is an intrinsic expectation of lawfulness associated with higher education institutions' practices. Transparency is another example. Transparency is shown to be an important quality for learning analytics, and is frequently discussed in learning analytics literature. But, the term is from the perspective of the learning analytics implementors and not the students. Transparency faces inwards, while benefits of transparency face outwards to the students, such as for example awareness. Hence, transparency is implicitly accounted for in the awareness principle. The list of privacy principles included in the questionnaire is not exhaustive, partly because of the reasons discussed, and partly because of the nature of the collection of the principles. It is possible that some privacy principles went unnoticed when reviewing literature if principles were not explicitly stated. Still, a comprehensive number of principles are included, and they cover much of the privacy aspects relevant for learning analytics in relation to students.

5.3.7 Question 7 Terms and Conditions

If you were asked to give consent to have your personal information collected, processed and analyzed, would it be likely that you would read a Terms and Conditions declaration before accepting/rejecting?

This question was included to measure student opinion on consent. As discussed earlier, consent has a special position within privacy, and is therefore given extra attention. The question attempts to examine how likely it is for students to read Terms and Conditions agreements. This was based on the impression from the literature review findings, suggesting that few students care about reading such agreements (Khalil, Prinsloo, and Slade, 2018; Prinsloo and Slade, 2015).

5.3.8 Question 8 Pressure to consent

If you were asked to give consent to have your personal information collected, processed and analyzed, by your higher education institution { would you feel pressured to give consent?

This question was also included to measure consent. Here it was explored if students feel pressured to provide their consent, if prompted by their higher education institution. This was based on the impression from the literature review that students might be more agreeable in relations to their higher education institutions.

5.3.9 Question 9 Learning analytics services and benefits

What services would you accept/deny your personal information to be used for, by your higher education institution?

- a) Improving your learning outcome.
- b) Improving your grades.
- c) Giving you consecutively information about how well you are doing in your subjects.
- d) Giving you an overview of how well you are doing in your courses compared to other students.
- e) Giving you personalized feedback and suggestions related to your courses.
- f) Improving the courses offered by your higher education institution.
- g) Increasing graduation rates of your higher education institution.
- h) Economic gain through sale of your personal data.

This question was included to identify what services/benefits students would accept that their personal information is used for by their higher education institution. The sub-items were based on findings from the literature review. Some services are derived from Siemens learning analytics model Siemens, 2013, reproduced in Figure 5, and (Siemens and Long, 2011) was also an inspirational source for the subitems.

As the other questionnaire subitems in this questionnaire, this list is not exhaustive either. There exists a large amount of services and benefits associated with learning analytics, and not all could be included. One of learning analytics most notable features is student at risk detection, as exemplified in the work by Arnold and Pistilli (Arnold and Pistilli, 2012).

This service is not included in the questionnaire as sub-item for similar reasons stated previously. student at risk detection is a service targeted at other learning analytics stakeholders. The service is in interest of the higher education institutions. Students are of course the subject of this service, and they are affected by it, but for students it is not desirable to drop out. The important aspects of student at risk services are implicitly included in other subitems, like a), c), d), and e).

Some subitems (f, g, h) are targeted at benefits not directly related to students personal benefit, but rather a benefit for their higher education institution. This was included to explore if there is a pattern in students answers, for example that students only want services/benefits that will result in their personal gain.

5.4 Implementation

The questionnaire was distributed in connection with the DigiTrans project ³³, led by Barbara Wasson at the Centre for the Science of Learning & Technology (SLATE) at the University of Bergen.

DigiTrans is a project tasked with researching changes the University of Bergen experience during the transition to a fully digital learning and teaching environment as a response to the pandemic. This research is part of DigiTrans theme D: Digital Student Behaviour.

The questionnaire does not depend on a sampling technique as the whole population was invited to answer the questionnaire. For the questionnaire there could be an unevenness in respondents that care about privacy and the ones that do not. There exists the possibility that the ones that do not care about privacy avoid answering the questionnaire. If this is the case, it can hurt the representatives of the sample. It is not possible to know the true reason for why members of the population refuse to participate, as there is no way of recording this. Still, an effort has been put into examining the partially completed responses, to look for traces of unevenness in respondents' attitude towards privacy. This is described in Section 6.2.1.

The Face validity was assessed for the questionnaire by a set of experts from the field of Learning Analytics; Professor Paul Prinsloo, Dr. Sharon Slade, and Dr. Mohammad Khalil. and the feedback they provided was implemented in the questionnaire design. The questionnaire was not tested thoroughly for validity and reliability, only face validity was assessed, but validity and reliability were protected as best as possible through the questionnaire development process.

5.4.1 Distribution

The anonymity of respondents was, as mentioned, central to the questionnaire. A consequence of this was that the questionnaire had to be accessed via a general link. The link forwarding to the questionnaire was distributed to all students (18 500³⁴) at the University of Bergen via SMS. Accompanying the link was an invitational text describing what the questionnaire was about. The Studieadministrasjon Department, at the University of Bergen, was responsible for the distribution of the text messages. If the questionnaire was distributed via E-mail, in addition to SMS, it is possible that it could have yielded more responses, as some of the questions are complex and might look clearer on a large screen in comparison to a small one. The SMS was distributed on the 24th of February 2021, with a response deadline set to the 15th of March. The questionnaire got 403 responses, this is discussed more in detail in Section 6.2.1.

³³<https://slate.uib.no/projects/it-takes-a-community-the-digital-transformation-of-uib>, accessed: 12.03.2021

³⁴based on numbers from: <https://www.uib.no/om>, accessed: 13.03.2021

5.5 Limitations

Limitations of using a questionnaire to gather empirical data comes from many sources. First, questionnaires have some known weaknesses, these can be called general limitations. Some limitations stem from design decisions, these can be called specific limitations, and other limitations become apparent post-questionnaire-conduction. The two first will be discussed next, while the latter will be described in Chapter 8.

5.5.1 General limitations

Low response rate is a general limitation with questionnaires (Bryman, 2015). Questionnaires are easy to ignore or easy to forget. The special situation from March 2020–present, have prompted the extensive use of questionnaires by higher education institutions, locally, nationally, internationally, as well as other organizations, to measure aspects of daily life during the pandemic. For example, questionnaires measuring students' mental health ³⁵ during the pandemic. Another example is the mandate of the DigiTrans project mentioned above, where other surveys have been set to students. This can strengthen low response rates as students are saturated with questionnaires. This combined with another concern, questionnaire fatigue, which can cause low response rates and a considerably high rate of partially completed questionnaire-responses. This ultimately leads to the absence of a representative sample, which is the next round leads to nongeneralizable results.

Reliability and validity were maintained through the design and development process of the questionnaire. This was done by following established guidelines and design principles as detailed above. Special attention was put into fulfilling Denscombe's (Denscombe, 2017) good practices for ensuring validity of questionnaire responses. This entailed that the questions were clear and in a logical order, sensitive topics were minimized, and anonymity was ensured. Except for this and the assessment of face validity through expert review, no additional reliability or validity testing were done for the questionnaire.

5.5.2 Specific limitations

Some limitations originate from specific decisions that was made for the questionnaire, as has been discussed in detail in the previous sections. As limitations are already mentioned, they are not further elaborated on, but they are summarized below:

- The questionnaire was in English (not Norwegian).
- The questionnaire was only distributed by SMS (not also E-mail).
- The questionnaire was totally anonymous.
- The questionnaire had many subitems.
- It was easy to provide passive responses to the questions.
- One question included technical terms.
- No reminders were sent out.
- Respondents had to supply answers to all questions.
- It is not possible to conclude that the respondents constitute a representative sample.
- It is not possible to conclude that the results of the questionnaire will be generalizable.

³⁵<https://www.uib.no/sa/138583/psykiske-plager-og-livskvalitet-hos-norske-universitetsstudenter>, accessed: 13.03.2021

5.6 Summary

In this chapter, the development and implementation of the questionnaire used in this research, have been discussed. Decisions in connection with the development have been explored in detail, questions have been documented and their purposes have been justified. The questionnaire was conducted as a web-based self-administered questionnaire containing 9 items with 37 subitems. The main focus of the development process was to develop a questionnaire that can yield responses relevant to answering the research questions. Measures were taken to prevent a low response rate, which is often regarded as the biggest limitation of questionnaires. This was done by attempting to limit questionnaire fatigue. Still, after many measures to avoid this, some of the design decisions could result in increased chance of questionnaire fatigue.

The questionnaire was implemented at the University of Bergen, where all students (approximately 18500) received an SMS containing an invitation to the questionnaire. Limitations with the questionnaire, both in regard to general and specific limitations have also been discussed. The results of the questionnaire are presented next, in Chapter 6.

6 Results

This chapter presents the results of secondary analysis and questionnaire. The results are presented as descriptive statistics. First the results of the secondary analysis is briefly described in Section 6.1, followed by a longer explanation of results from the questionnaire, in Section 6.2. The results produced by question 6 of the questionnaire is the most comprehensive in this chapter. This is described in Section 6.2.6, but first the secondary analysis.

6.1 Secondary analysis results

The results of the secondary analysis are for the most part presented in pie charts. The charts are presented in Figures 17 and 18, with the latter presenting to pie-charts. The charts displaying results from the secondary analysis has a dimly colored background in order to separate them from the charts displaying results from the questionnaire implemented for this research which has no background color.

6.1.1 Respondents

The questionnaire administered by Opinion for the NDPA recieved 1501 responses. The age of the respondents were distributed as follows:

- Below 30: 24%
- 30-39: 16%
- 40-49: 17%
- 50-59: 16%
- 60 and above: 28%

The respondents to this questionnaire were a cross-section of the Norwegian population, not only students as is the case for the questionnaire implemented for this research.

6.1.2 Privacy importance

Figure 17 displays the degree to which Norwegians are concerned with privacy. 83% of the respondents are to a **large degree** or to **some degree** concerned with privacy, while 16% of respondents are less concerned with privacy.

6.1.3 Feeling of lack of control and powerlessness

Figure 18(A) displays the extent to which Norwegians feel limited control over how their personal information is stored and used on the internet. 66% feels to a **large degree** or to **some degree** little control.

Figure 18(B) shows the extent to which respondents feel powerless when it comes to having control over their personal information on the internet. 61% feels to a **large degree** or to **some degree** powerless.

Based on these results, most of Norwegians are concerned with privacy, feels little control over personal information stored about them online, and, feels powerless when it comes to having control over their personal information on the internet.

6.2 Questionnaire results

The results of the questionnaire implemented for this research, are presented in this section. Each question from the questionnaire has a corresponding figure. The results are grouped together under headings reflecting their contents. Question 1 is not included as this question was only intended to confirm that respondents were students. First, the respondents are described before each question is addressed in turn.

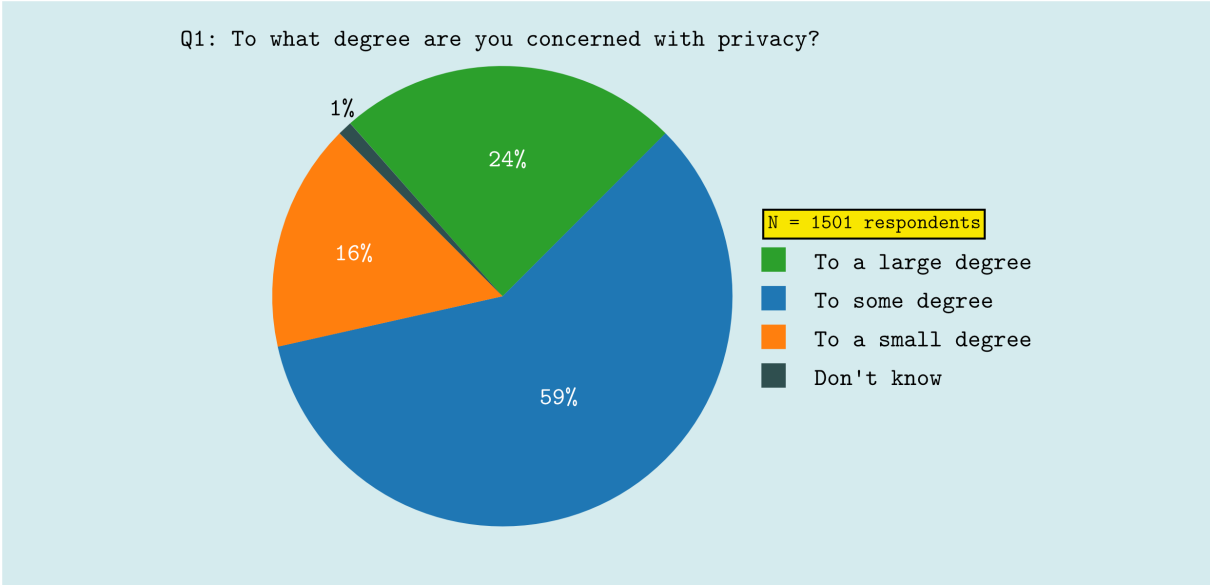


Figure 17: Norwegian's degree of concern with privacy.

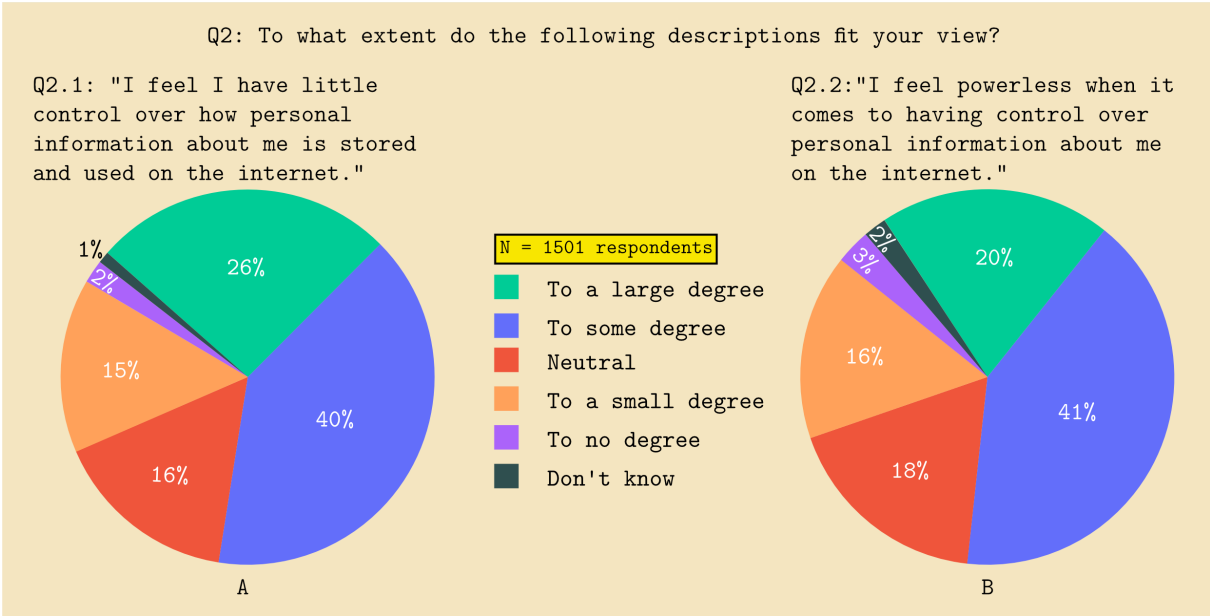


Figure 18: Norwegian's feelings of limited control and powerlessness online.

6.2.1 Respondents

The questionnaire received 568 responses. 165 of these were only partially completed and were discarded, resulting in 403 complete responses, meaning that they answered all questions. This results in a completion rate of 71%, while the response rate is 2.2%. 12.7% of respondents used more than 12 minutes to complete the questionnaire, 23.9% used, less than 5 minutes, and the rest (63.4%) used between 5 and 12 minutes. Of the complete responses, 8–9 were corrupt. The reason for the corruption is unknown, but it happened most likely in connection with the exportation of the results from SurveyXact.

This means that the data foundation for each of the questions is between 395 and 394 complete responses. Because of this slight variation in respondents, all figures have been labeled with an N= to indicate how many responses make up the data foundation for the different figures. All respondents stated that they were students at a higher education institution (Q1). When describing results about the fully completed responses, the respondents are referred to as students.

Figure 19 displays a heatmap of all (partial and complete) questionnaire responses. A larger version of this figure is presented in Appendix C (Figure C.1). Each row in this figure represents one participant's response. Each column represents one question or sub-question. The columns have been labeled to indicate what question they correspond to. The different nuances in grey color indicate how the respondents answered the questions and is of no importance in this Figure.

The pink color indicates where respondents have answered using the alternative *“don't know / no opinion”*. This is highlighted to display if respondents have provided passive responses, for example, answering *“don't know / no opinion”* on all of the questions. If this occurred, it would be displayed as long pink vertical lines in the figure. Only 6 respondents have answered a high number of *“don't know / no opinion”* for many of the sub-items of question 6. The green arrows indicate where the identified corrupt values are, two are still not identified. The red line indicates the start of question 6.

Some answered only *yes* to being a student at a higher education institution before quitting, while the majority of participants that partially completed the questionnaire, quit after question 5. Some only answered question 2 and 3, before quitting. Three respondents quit after completing question 6.

Of the partially completed responses 98 have provided an answer to question 3. The question examine how important they think privacy is. The arithmetic mean of this answer is 4.23 where 4 is equivalent to having answered *important* to the question.

As it is not possible to determine with certainty that the respondents (n=394) constitute a representative sample, it will be treated as a non-representative sample. Because of this, the results will not be treated as generalizable to a population outside of the pool of respondents. Still important insights into student perceptions of privacy in learning analytics can be gained, as is shown in the following results as well as in Chapter 7.

6.2.2 Privacy understanding and importance

Figure 20 displays the results of question 2, which measures how students understand privacy. Half of the students (50.9%) understand privacy as *“The right to decide when, how and to what degree information about me is shared with others”* (d). 21.8% of students understand privacy as *“The right to safeguard my personal integrity; safeguard my opportunity for private life, self-determination (autonomy) and self-expression”* (e). 3 students 0.8% provided other (g) definitions of privacy.

Figure 21 displays the results of question 3, which measures how important students find privacy. 91.7% of students find privacy important (very important / important), and 8 students (2.03%) finds privacy less important. None of the students answered that privacy is unimportant. Thus all students had an opinion on the importance of privacy.

Heatmap of questionnaire responses

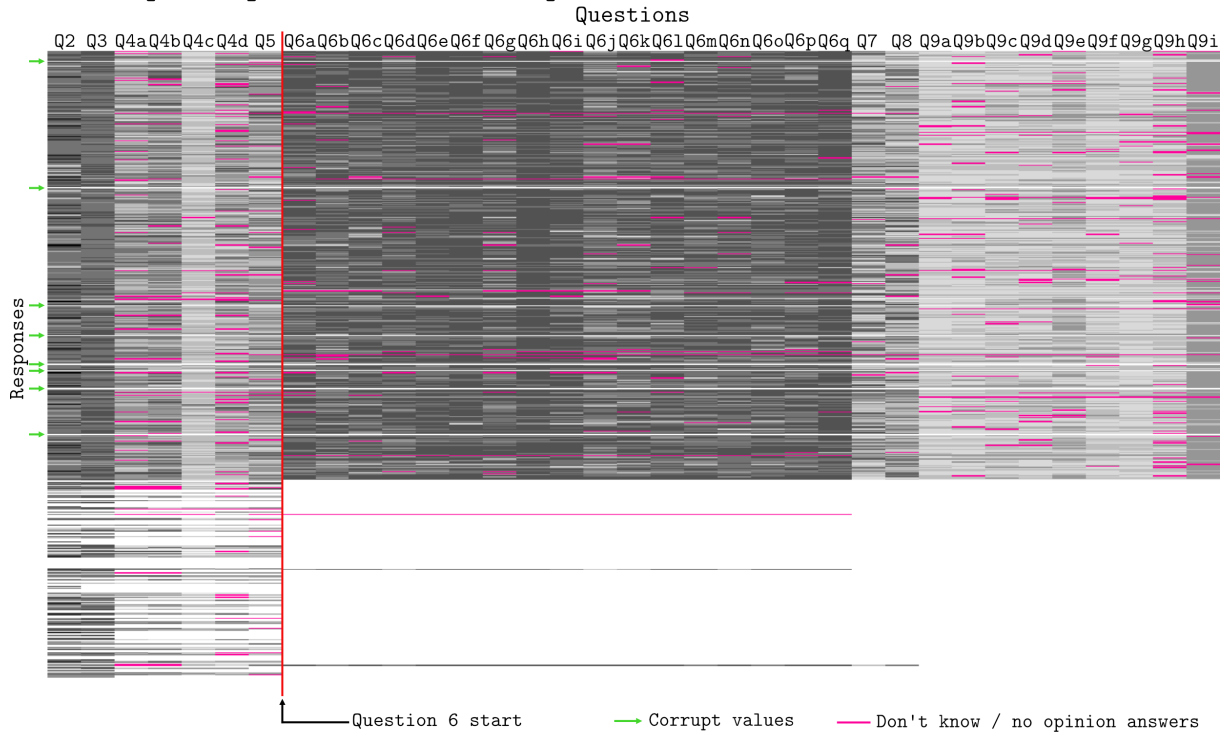


Figure 19: Heatmap of questionnaire responses.

Q2: Which of the following statements is closest to how you understand privacy?

N = 395 respondents

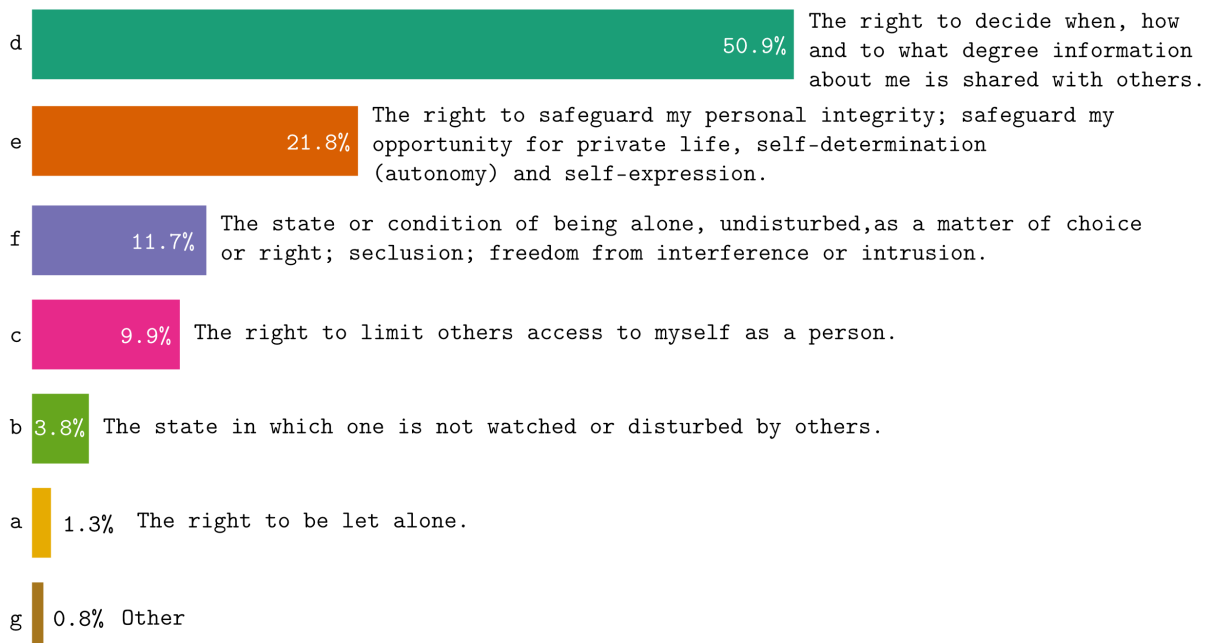


Figure 20: Results Q2: Definition students finds most descriptive of their privacy-view.

Q3: How important is privacy to you?

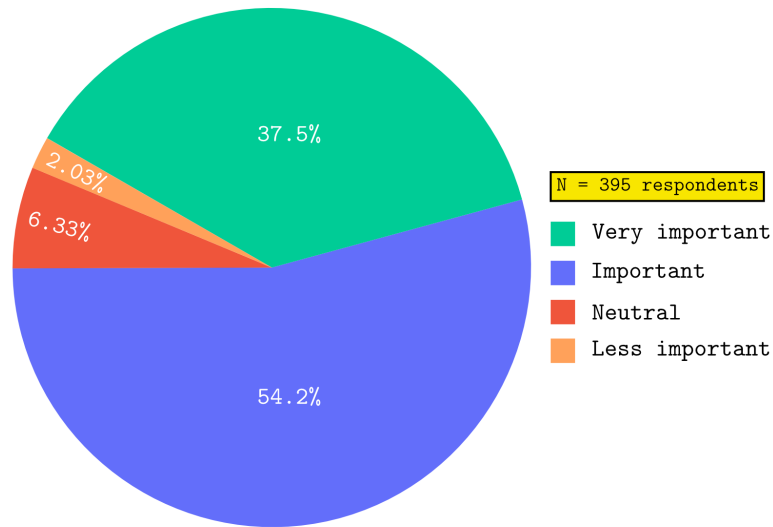


Figure 21: Results Q3: How important students find privacy.

6.2.3 Feeling of control and powerlessness

Figure 22 shows the results of sub-questions a and b of question 4. Figure 22(A) displays how little control students feel that they have over personal information about themselves, handled by their higher education institution. 37.7% of students feel little control (strongly agree / agree), while 27.4% do not feel little control (disagree / strongly disagree) and 28.1% are neutral to the question.

Figure 22(B) displays how powerless students feel when it comes to having control over their personal information in relation to their higher education institution. 23.6% of students feel powerless (strongly agree / agree). 41.3% do not feel powerless (disagree / strongly disagree), and 30.4% are neutral to the question.

6.2.4 Privacy self-management and expert-management

Figure 23 shows the results of sub-questions c and d of question 4. Figure 23(C) displays how students position themselves to privacy self-management. 87.6% of the students want to be able to control and ensure their own privacy (strongly agree / agree). 5 students (1.3%) do not want to control and ensure their own privacy (disagree / strongly disagree).

Figure 23(D) displays how many students want privacy expert-management. 42.6% of students want privacy experts or larger bodies (like an independent administrative body) to control and ensure their privacy (strongly agree / agree). 15.4% do not want this (disagree / strongly disagree), and 31.6% are neutral to the question. 10.4% do not know or do not have an opinion of the statement.

Q4(a/b): To what extent do you agree with the following statements? N = 395 respondents

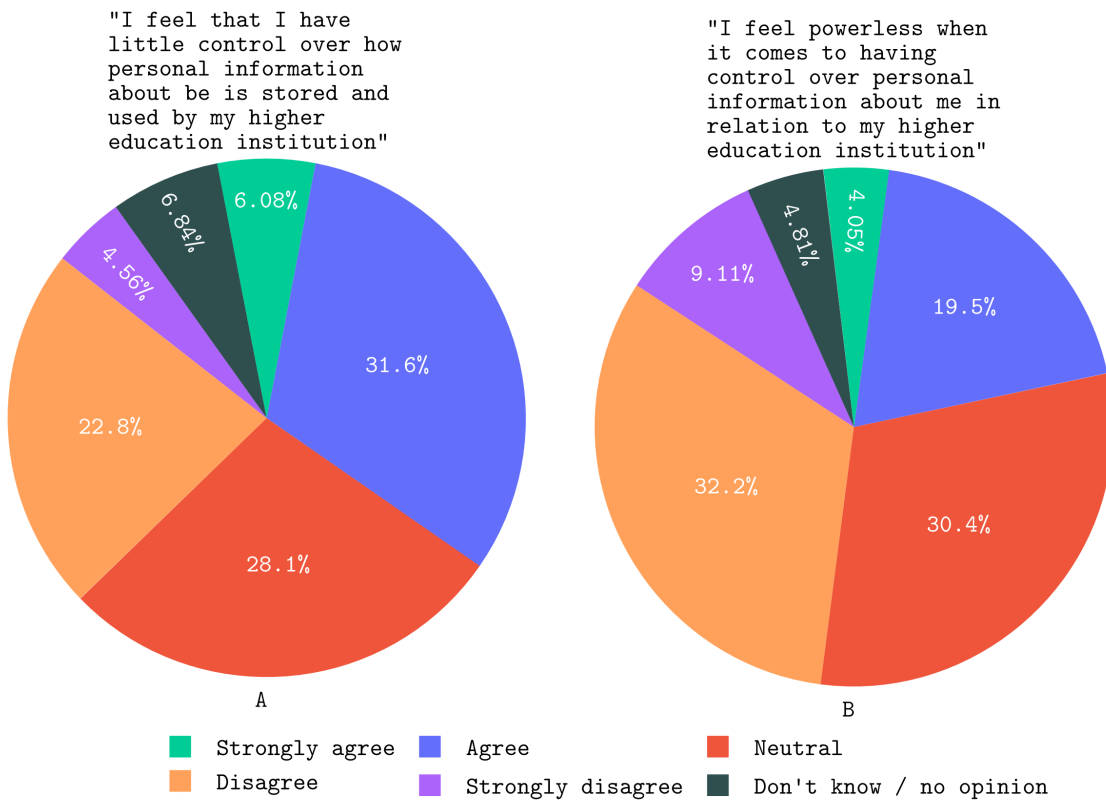


Figure 22: Results Q4 (a/b): Students' feeling of lack of control over their personal information in relation to their higher education institution (A). Students' feeling of powerlessness over their personal information stored at their higher education institution (B).

Q4(c/d): To what extent do you agree with the following statements? **N = 395 respondents**

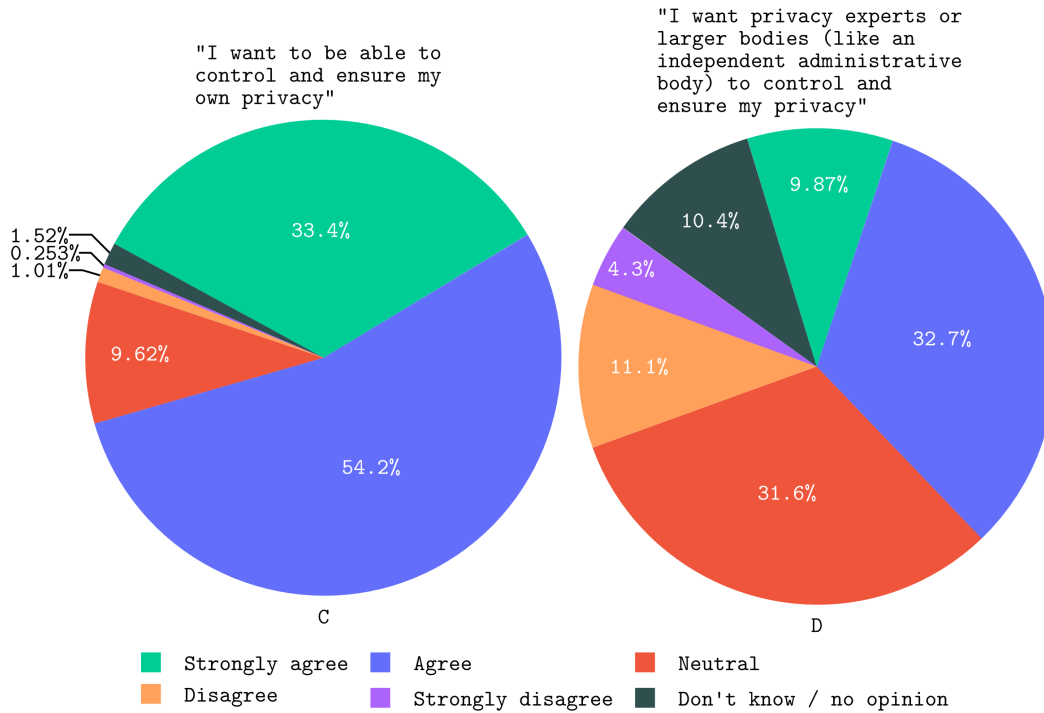


Figure 23: Results Q4 (c/d): Students' desire for privacy self-management (C). Students' desire for privacy expert-management (D).

6.2.5 Learning analytics desirability

Figure 24 displays what services/benefits students would accept/deny their personal information to be used for by their higher education institution. The figure is sorted by how many of the students would accept the service/benefit. The top three most accepted services/benefits for students are: (1) improved learning outcome (a), (2) improving courses (g), (3) improved grades (b). 59.9%, 56.3% and 55.8% of students (respectfully) would accept that their personal information is used for these purposes.

Most students (79.9%) would deny their personal information to be sold for their higher education institution's economic gain (i). 31.2% would deny the service of "Getting an overview of how well they are doing in their courses compared to other students"(e).

Figure 25 displays how desirable students find learning analytics. 38.2% of students want learning analytics (very desirable / desirable). 24.8% do not want learning analytics (undesirable / very undesirable). 32.9% are neutral to the question.

The arithmetic mean for the answers related to services and benefits of question 9, are:

- Accept: 42.8%
- Maybe accept: 31.0%
- Deny: 20.1%
- Don't know / no opinion: 6.1%

Q9: What services would you accept/deny your personal information to be used for by your higher education institution?

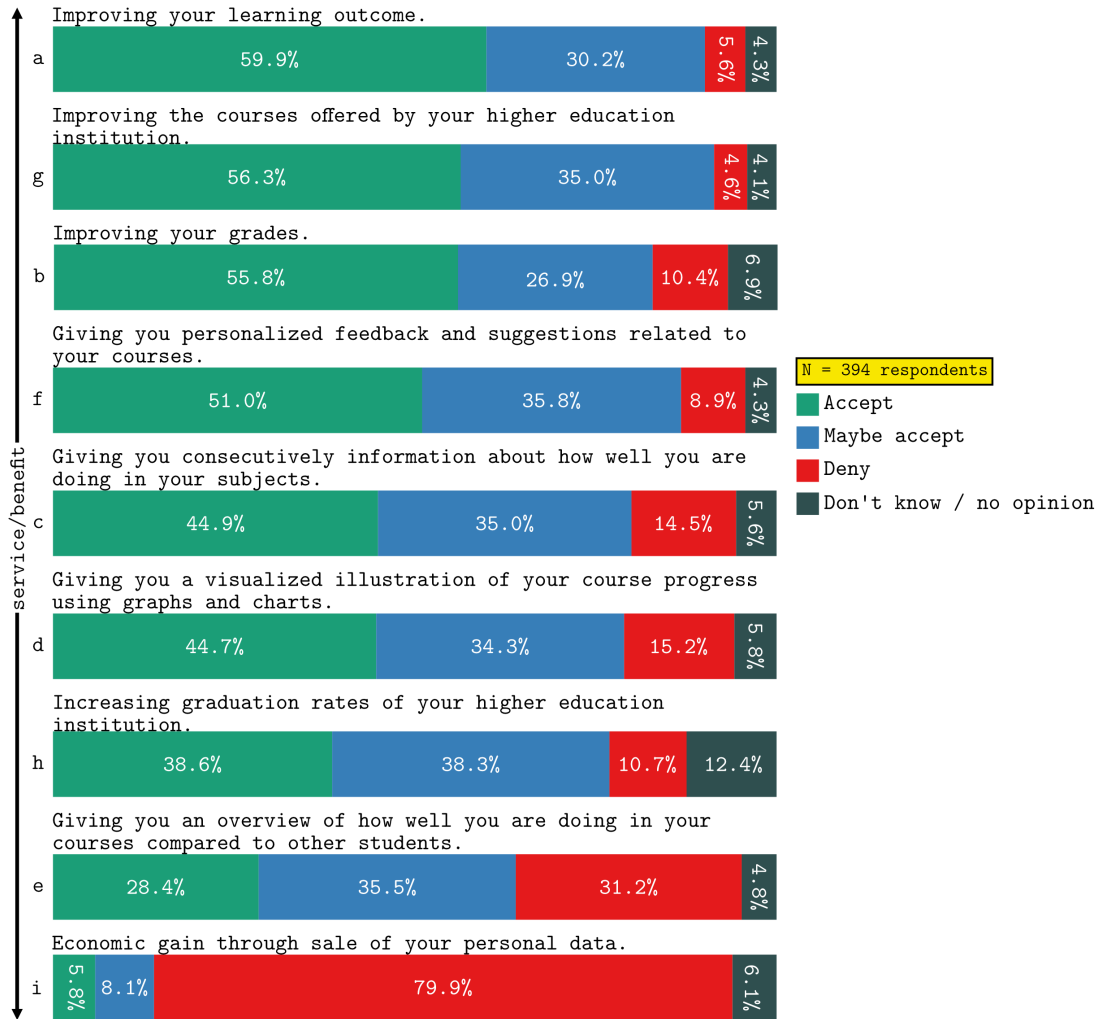


Figure 24: Results Q9: What information students would trade for learning analytics services/benefits

Q5: How desirable is it for you to have your personal information collected, processed and analyzed by your higher education institution in order for them to provide services that has the potential to improve your learning outcome?

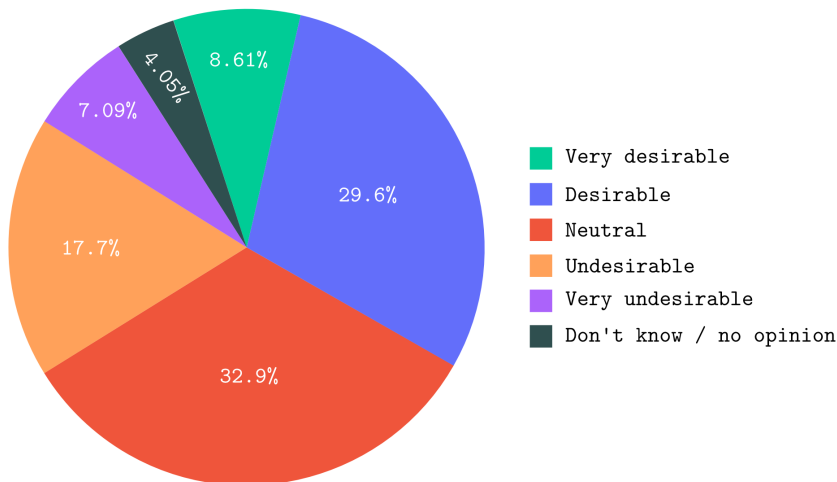


Figure 25: Results Q5: Learning analytics desirability among students.

6.2.6 Importance of privacy principles

Figure 26 displays how important students find 17 different learning analytics related privacy principles. The principles are sorted by importance (combination of very important / important answers). The figure is read clockwise.

Data security is the most important privacy principle for students. 95% of students answered that Data security is very important or important. 79.2% of students answered that Data security is very important, this is the highest **very important**-value for any of the principles. Data security was in the questionnaire explained as: *“That your personal information is being protected from unauthorized access, the risk of being intercepted, decoded or tapped”*.

After data security, the topmost important principles (very important / important) are: Consent (94%, Awareness (93%), Trust (92%), and Data sharing (88%). These were in the questionnaire, explained in the following way:

- Consent – *“That you are asked to agree to a request, before your personal information is potentially collected and analyzed by your higher education institution”*.
- Awareness – *“To be informed. To know what personal information about you is collected, processed and analyzed by your higher education institution”*.
- Trust – *“That you have firm belief in, can rely on, and have confidence in your higher education institution to protect your privacy”*.

The arithmetic mean importance value for each principle is 84.4% (very important / important). The arithmetic mean for all the answer alternatives are:

- Very important: 53.8%
- Important: 30.6%
- Neutral: 9.9%
- Less important: 2.2%
- Unimportant: 0.8%
- Don't know / no opinion: 2.7%

Data preservation and Limited access are the least important privacy principles. 68% and 74% of students (respectfully) answered that these are important (very important / important). These privacy principles also have the highest amount of **neutral**-responses: 20% and 17%, respectfully. Data preservation also have the highest amount of **unimportant**-responses. 8 students (2%) answered that Data preservation is unimportant. These privacy principles are explained in the following way in the questionnaire:

- Data preservation – *“That your personal information stored in your higher education institution is deleted periodically”*.
- Limited access – *“That as few as possible are able to view, process and analyze your personal information”*.

Accountability and Data ownership share the most **don't know / no opinion** responses, rated at 4.6% (18 students).

Q6: Imagine that your higher education institution is collecting, processing and analysing your personal information in order to improve your learning outcome. How important are the following principles to you?

N(a-e) = 395 respondents
N(f-q) = 394 respondents

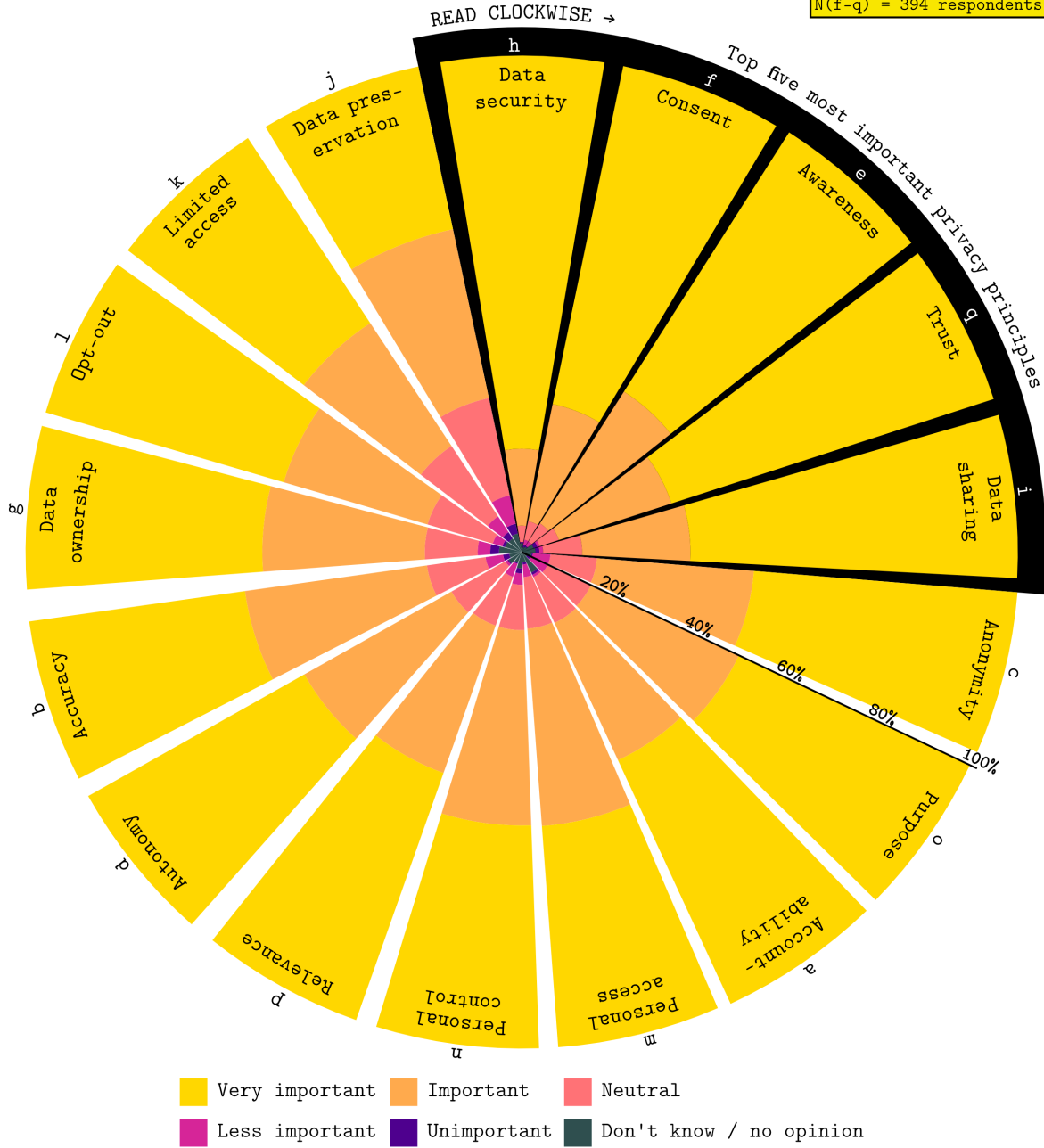


Figure 26: Results Q6: Importance of privacy principles for students (A)

6.2.7 Terms & Conditions and the feeling of pressure to consent

Figure 27 displays how likely it would be for the students to read a Terms and Conditions agreement related to learning analytics, if prompted by their higher education institution. 34% would be very likely or likely to read such an agreement, while 47.2% would be unlikely or very unlikely to do so.

Figure 28 displays how many students would feel pressured to give their consent if prompted by their higher education institution. 50.3% would feel pressured to give consent (strongly agree / agree). 28.5% would not feel pressured to give consent (disagree / strongly disagree).

Q7: If you were asked to give consent to have your personal information collected, processed and analyzed, would it be likely that you would read a Terms and Conditions declaration before accepting/rejecting?

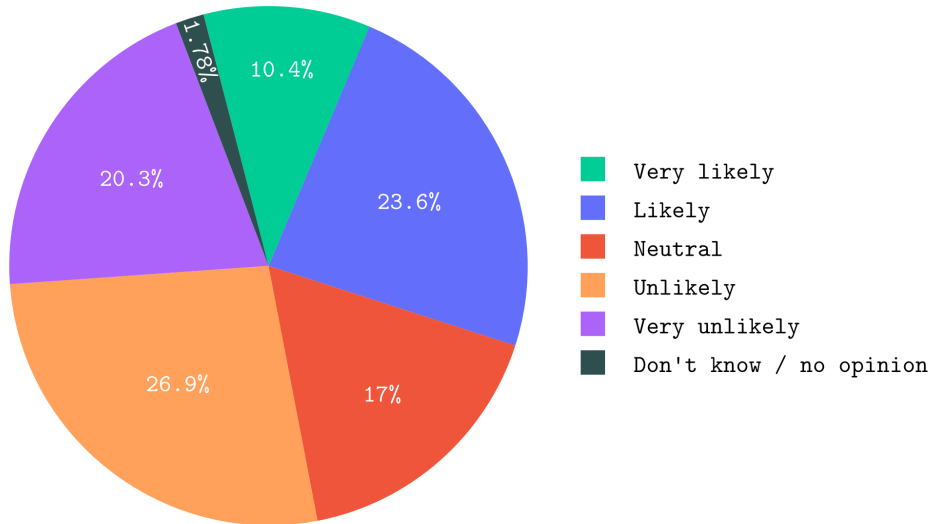


Figure 27: Results Q7: Likelihood of reading Terms and Conditions agreements.

Q8: If you were asked to give consent to have your personal information collected, processed and analyzed by your higher education institution - would you feel pressured to give consent?

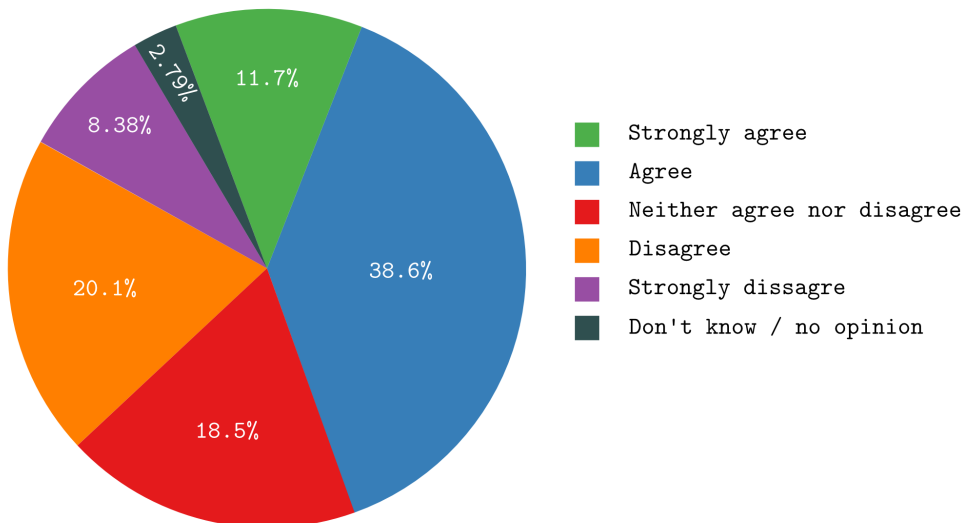


Figure 28: Results Q8: Students' feeling of pressure to consent

7 Discussion

This chapter is structured around answering the research questions, with one section dedicated to each research question. In each of the sections, the research questions will first be stated and then answered. The main discussion follows after Section 7.2. Here the findings from the results will be interpreted and connected to the literature review findings, and implications will be explored. The chapter starts by answering the first research question.

7.1 Privacy principles in learning analytics

(RQ1) What privacy principles are most relevant for learning analytics?

The most relevant privacy principles for learning analytics, based on the findings in the literature review, as well as the questionnaire, are: consent, students control, and data security. Consent is considered especially relevant as it is found across multiple analyses. From a student perspective, data security and consent are found to be important privacy principles, in multiple analyses and are therefore considered highly relevant for students.

Following is an explanation on how this answer was researched. Throughout this research, many different privacy principles have been explored. Different results relating to which privacy principles are the most important for learning analytics, have been found through quantification of the privacy debate, in in the literature review (Chapter 3), as well as in the questionnaire results. Table 11 displays a comparison of the different findings. It compares findings from the two quantifications in the literature review (A and B) and the findings of the questionnaire (C). The table displays the top five most important principles found by the different analysis. Column B has three principles on shared 4th place (anonymity, restricted access, and access) therefore all of these are included.

Table 11: Privacy principles important for learning analytics. (A, B): literature review findings, (C): questionnaire findings

Ranking	A (from Table 4)	B (from Figure 13)	C (from Figure 26)
1	Consent	Student control (Control)	Data security
2	Transparency	Consent	Consent
3	Data ownership	Data security	Awareness
	Student control	Anonymity	Trust
	Access	Restricted access	Data sharing
		Access	

This table represents the answer to the first research question. The research question can be answered in multiple ways, dependent on what perspective is used. Column A in the Table 11 represents the principles found important for the general privacy debate in learning analytics. Here other stakeholders than students are also represented. This is apparent in the principles transparency and data ownership, as these are important principles for learning analytics implementors. Consent is the top ranking privacy principle in the learning analytics discussion, based on the findings in the literature review.

Column B and C in Table 11 represents the student perspective. For students, student control, data security, consent, and awareness are the highest ranking privacy principles, thus these are considered the most important to students.

Most interesting are the principles that are found important in more than one analysis. Consent is ranked high in all the analyses. Data security is prominent in analyses of the student perspective (B and C). Student control is high ranking for students based on the literature review (B), and is also found important in the general privacy debate in learning analytics (A). The same can be said for Access.

The principles that are found to be highly important to students, based on the questionnaire results (C) will be discussed more in detail, later.

7.2 Privacy perception

(RQ2) How do Norwegian students perceive privacy in general, and in relation to learning analytics?

Based on the findings from the questionnaire, students perceive privacy to be highly important, more important than the population in general. Most students have an information-privacy centered understanding of privacy. Seven out of ten student privacy perceptions fit inside the established conceptualization of *personvern* and *personopplysningsvern*. Privacy self-management is an important part of most student perceptions of privacy, students want to control and ensure their own privacy. Many students also want privacy expert-management. Student views on privacy management are reflected in the current privacy management structure in Norway; a shared responsibility of privacy management between self-management and expert-management.

In relation to learning analytics, students find Data security, Consent, Awareness, Trust, and Data sharing to be the most important privacy principles. This reflects the dominant student privacy perception, which is centered around information privacy. Consent is the second most important privacy principle for students, but despite this, a high number of students would feel pressured to give their consent to their higher education institution if prompted. A high number of students are also not likely to read Terms and Conditions agreements, if prompted.

Following is the interpretation and discussions that lead to the research question answer. Privacy is considered highly important for students. As displayed in the results section, nine out of ten students think that privacy is either very important or important. This is almost 10% higher than the Norwegian population in general, where approximately eight out of ten, are to a large degree, or to some degree concerned with privacy. Many of the students that partially completed the questionnaire also find privacy to be important.

As privacy is highly important for students, their privacy should be given attention. Stating that privacy is important for students is easy, but knowing how to accommodate for this, in practical learning analytics, is more challenging. A more nuanced view of the importance of privacy can be uncovered by looking at the different privacy principles students find important. Still, this does not say anything about how privacy should be implemented in practice. But, it can provide insights into what privacy topics should be prioritised, in order to satisfy the needs of students. This has implications for the learning analytics research field, as it needs to figure out how to accommodate these privacy priorities.

An interesting observation is that when asked about the importance of privacy, nine out of ten students, express that it is either very important or important, but when asked the same on a per-principle basis (i.e., 17 different privacy principles), closer to eight in ten express that privacy is either very important or important. This is based on the arithmetic mean of all the principles, combined. This finding can indicate that students initially overstate how important privacy is to them. One possible explanation for this, could be that students have not introspected their opinion on privacy, and when faced with privacy broken down into smaller pieces, they discover their true opinions of privacy, which is reflected in the results. That these questions give dissimilar results can indicate that the use of privacy principles as indicators of privacy, provides a more nuanced perspective on privacy, compared to just asking students how important they think privacy is. Another explanation is that a more nuanced scale will provide more nuanced results, this is independent of what is measured.

Most students understand privacy as “*The right to decide when, how and to what degree information about you is shared with others*”. This is interesting as this definition is more proximate to information-privacy, than general privacy. An explanation for this can be the digital society we live in today. Information-privacy is possibly the most relevant form of privacy in today’s society, as our lives becomes increasingly intertwined and dependent on online technology generating and aggregating large amount of information about us. Privacy in the traditional form, as the right to be let alone, or the Norwegian term *privatliv* may be less relevant, as it does not capture the zeitgeist of today. The fact that half of the students think that this definition lies closest to how they understand privacy, clearly indicates that students want to have control over their information flow (based on how the definition is formulated). This is similar to results found by other learning analytics researchers (e.g. Slade, Prinsloo, and Khalil, 2019). This needs also to be seen in relation to privacy self-management.

Privacy self-management is desired by almost nine out of ten students. Privacy self-management is not described by its term in the questionnaire but students are asked to answer how much they agree with the following statement: *‘I want to be able to control and ensure my own privacy’*. It is clear from this answer that privacy self-management is desirable for students. This has implications for learning analytics, as opportunities for privacy self-management should be respected. This is not a big challenge, as this is partly covered by the GDPR. The question of how much impact students should have over privacy decisions is still an open question.

Less than half of students, four out of ten, want privacy expert management, while a larger amount is neutral to the question, or do not know if they want expert management. The combination of privacy expert-management and privacy self-management is similar the privacy management model currently used in Norway. In addition to being under legislation requiring privacy enforcement, independent bodies, like the NDPA, exist to assist in safeguarding privacy. This constitutes privacy expert-management. Individuals are also provided a high degree of privacy self-management through rights granted by for example the GDPR. The rights grants individuals control over their own privacy to a large extent, at least in principle. Based on numbers from the questionnaire, students are generally satisfied with the current model of controlling and enforcing privacy.

Privacy self management, as described in the questionnaire, entails that students want to control their own privacy, but also want privacy to be partly controlled by experts. This could be interpreted as a need for privacy experts to lay the foundation of how privacy should be managed in addition to be given tools so they can control and ensure their own privacy. For learning analytics, this might implicate that learning analytics would provide privacy protection, based on expertise, in addition to providing tools for students that can be used to control and ensure their own privacy.

What is understood by the terms *ensure* and *control* can be up for discussion, but the terms, at least, entail impact over decisions. This does not necessarily mean that students want to enforce their own privacy in a strict manner, possibly limiting learning analytics functionality. It could also mean that they want to relax their privacy preferences. The results of the questionnaire indicate that students want this responsibility, as they want to control and ensure their own privacy. Although, it is possible that students do not fully understand what privacy self-management entails, as it also comes with a responsibility. This can be connected to the question asked in the literature review: *Is privacy management a challenge too big to handle alone?* This question is still unanswered as it is not possible to provide an answer based on the questionnaire findings. This topic needs to be researched further in order to find out if students take on too much responsibility with their wish of privacy self-management, and if they actually desire it when they are aware of the consequences.

That the dominant privacy perception among student is: *“The right to decide when, how and to what degree information about you is shared with others”* can also be seen in connection with Solove’s (Solove, 2009) privacy challenges of information aggregation, as described in his privacy taxonomy. Aggregation was also a recurring topic in the literature review, as it is one of the cornerstones of learning analytics.

This privacy understanding can indicate that students are more concerned about the information that they produce, than being concerned about surveillance, or being disturbed by others, as other privacy definitions entails (Figure B.3 definition a, b and c, Appendix , B).

An implication for learning analytics is thus to accommodate for students information-privacy centered, privacy understanding. As the nature of learning analytics entails collection of students information, this privacy understanding is probably the most relevant for learning analytics regardless of student perceptions. This dominant understanding of privacy is also possibly relevant for a broader context, outside of learning analytics, although the results cannot be used to directly support this implication.

Another observation related to privacy understanding, is that one in five students think that the personvern definition (definition: e) describes their understanding of privacy. This is interesting as it was expected, prior to the questionnaire, that this privacy understanding would be the most dominant, as personvern is the dominant term used in Norway when discussing *privacy*. Why this privacy understanding is subordinate can have multiple explanations. One interpretation is that it could mean that students

do not think that privacy and personvern are the same concept or it could indicate that the personvern definition is not very fitting for their understanding of privacy.

Judging by how important students find privacy, the first interpretation is unlikely. This is because it would be odd that privacy would be considered a whole different concept than personvern but still be considered highly important. This would be unnatural because of the high degree of overlap between privacy and personvern. Personvern and privacy are in an every-day setting generally considered to be equivalent. Perhaps personvern is considered to include more than privacy and it can also be considered more refined, except for this, the concepts are fairly similar. A total separation of the concepts would be unnatural. That personvern is not as fitting for their privacy understanding compared to that of information-privacy is a more plausible explanation. It is also possible to combine these two most dominant privacy understandings.

Personvern is, as explained earlier, closely connected to *personopplysningsvern*. *Personopplysningsvern* is similar to the concept of *information privacy*. This allows for the combination of the two most dominant privacy understandings, as the second most dominant privacy understanding corresponds to *personvern*, and the most dominant privacy understanding corresponds to *personopplysningsvern*. When combining these two, 71% of students are represented by the same understanding of privacy. This constitutes the large majority of students and entails that students' dominant understanding of privacy can be described using the *personvern* and *personopplysningsvern* definitions explained in Chapter 2.

Adding these understandings together does not mean that the majority of students understand privacy in the same way. What it means, is, by utilizing the personvern definition in addition to personopplysningsvern (which is the dominant position in Norway) the majority of student privacy perceptions are covered and accounted for in how privacy is handled. This is beneficial, as there is not a large mismatch between how privacy is defined in Norway (entailing how it is handled), and how students understand privacy. Still approximately 30% of the students are not included in the dominant perception and their privacy interest can possibly be overlooked because of this. Ultimately this has implications for learning analytics in Norway. This implicates that, if learning analytics in Norway, use the *personvern* and *personopplysningsvern* definitions as their foundation of their privacy management, it will result in privacy protection that covers the interests of most students.

This finding is also important for the applicability of the results of this research, as they can be directly viewed as results related to personvern, and not only *privacy*. This is important as personvern is the dominant concept in Norway, and results relating only to *privacy* could be considered second class results. The reason for bringing attention to this fact is because there exists a position that personvern and privacy are not the same concept. This finding however implies that the term *privacy* can be used as equivalent to personvern, when doing empirical research on Norwegian students. This can possibly have large implications for future research as English terms can be used in research on Norwegians, possibly without the loss of quality.

The implication that most students share the same understanding of privacy can also be considered contradictory to for example Solove's (Solove, 2009) conception that privacy needs a reformation. It is possible that privacy needs a reformation, but this does not seem applicable to personvern. Interestingly, as pointed out previously, Solove's conception of how privacy should be conceptualized is similar to how personvern is currently understood, with some distinct differences, for example how the value of privacy is assessed. This fusion of concepts is interesting as it could implicate the start of a universal understanding of privacy. Locally in Norway this would constitute little difference if personvern and personopplysningsvern becomes the dominant global perception of privacy, as it is already dominant locally. A global understanding of privacy could entail great benefits as the ambiguity of privacy will be reduced as cultural/national understandings disappears. A shift towards a more generic privacy perception is represented in the implementation of the GDPR. Although, this is an international phenomenon, not a global one. It also represents the *de jure* understanding and not the necessarily the *de facto* understanding of privacy, implying that although privacy is handled similarly between nations because of the GDPR, the perception of privacy can still vary vastly. How privacy is perceived fits into a global discussion of privacy, related to how privacy should be understood and how it should be handled. The way privacy is perceived entails how it is handled, and different privacy perceptions can entail great variations of how privacy is handled. This underscores the importance of exploring how privacy is understood.

The dominant privacy perception discovered in the questionnaire results, can be interesting for learning analytics implementors as well as anyone concerned with developing digital solutions for students, as it indicates where privacy ensuring efforts should be invested when dealing with Norwegian students.

A more nuanced view of how students understand privacy in relation to learning analytics is gained by exploring results relating to privacy principles, from the questionnaire. The following discussion use a selection of the privacy principle results to explore student perceptions of privacy for learning analytics. When asked about these principles, in the questionnaire, students were asked to envision that their higher education institution was collecting, processing, and analyzing their personal information. The students voiced their opinions of different privacy principles, based on this premise. The most important privacy principles are included in the discussion; Data security, Consent, Awareness, Trust, and Data sharing, but other less important principles are also highlighted, as the principles are highly interconnected. Discussing the results isolated by principle, would be less fruitful. This has a small consequence of the clarity of the discussion, as the principles are not discussed in a sorted order.

Data security is the most important privacy principle to students based on the questionnaire results. Data security is also one of the most important privacy principles to students based on findings in the literature review, for example represented by the findings of Whitelock-Wainwright et al. (Whitelock-Wainwright et al., 2020), who finds that students have high ideals when it comes to data security. In light of how most students understand privacy, a connection can be made. The information privacy centered understanding, of students, can explain the importance of data security. Data security cannot be seen in isolation as it was defined in the following way, in the questionnaire: *“That your personal information is being protected from unauthorized access, the risk of being intercepted, decoded or tapped”*. This explanation connects Data security to Data sharing and Limited access, and to some extent also to Trust.

Data sharing (fifth most important) is explained in the questionnaire, as: *“That your higher education institution does not share your personal information with others internally, with other companies or with other service providers.”* In the literature review it was discovered that the collection of student data was the most asked about “principle”, in research exploring student perceptions of privacy. This finding is related to both the collection by higher education institutions, as well as collection by third parties. Here Tsai et al. (Tsai, Whitelock-Wainwright, and Gašević, 2020) found that students are uncomfortable with sharing data with third parties. This finding is supported by results from this questionnaire, related to data sharing. Students do not want their information shared internally or externally. This has implications for learning analytics systems, as the information it collects cannot be shared with others, if students were to decide. If this was the case, it could entail interoperability challenges. A solution to this, can be that learning analytics anonymize the personal information they collect. This would remove the personal identification aspect of the personal information making it noncontroversial to share both internally and externally.

How this should be solved on a technical level is outside the scope of this discussion, but a lot of effort, both inside and outside, of the learning analytics field, is put in creating good anonymization techniques. The task of anonymization is not a learning analytics specific task, but a task relevant to all areas having to deal with personal information. Data sharing as well as Data security are also connected with Limited access.

Limited access is considered one of the least important privacy principles to students, according to the questionnaire results. Still 74% of students find it very important / important. This underscores how important even the least important privacy principles are to the students, based on the questionnaire results. Limited access was explained, in the questionnaire, as: *“That as few as possible are able to view, process, and analyze your personal information”*. It is interesting that students express that they do not want their information to be shared, but are less concerned with whom sees their personal information. This can be connected back to the information privacy centered privacy understanding of students. As mentioned, students want to control their flow of personal information (based on the Data security principle), and they do not want their personal information to be shared internally or externally (based on the Data sharing principle). But, they are less concerned with how many people view, process, and analyze their personal information (based on the Limited access principle). This further supports the information privacy centered understanding of students, as they apparently, are less concerned with how many sees

their information (surveillance) compared to having the information shared internally or externally.

The factor of asking about external sharing in the same question as internal sharing, can be an explanatory factor for this differences in priority. It is plausible that students find external sharing worse than internal sharing, but the questionnaire results cannot be used to state this, as the question is not nuanced enough to give an answer to this.

Consent is the second most important principle for students, based on the questionnaire findings. In the questionnaire, consent was explained as: *“That you are asked to agree to a request, before your personal information is potentially collected, processed and analyzed by your higher education institution”*.

As displayed in Section 7.1, consent is a privacy principle that dominates privacy in learning analytics, from multiple perspectives. It is a principle important in the general privacy debate in learning analytics, but also when the topic is student perceptions of privacy in learning analytics. The nature of consent entails that it is important for multiple stakeholders. Why consent is important depends on the stakeholder asked, thus, the perspective of different stakeholders are briefly discussed.

For students, consent is related to Awareness and Personal control. Students want to know what they agree too, and what their personal information is used for. This is apparent when looking at how important students find Awareness. Awareness is explained in the questionnaire as: *“To be informed. To know what personal information about you is collected, processed and analyzed by your higher education institution”*. This is the third most important privacy principle to students, based on the questionnaire results. Students also regard Awareness as more important than Personal control. This is interesting as Personal control was found in the literature review (called student control) to be the most important privacy principle to students. Although, most students want privacy self-management, as discussed previously indicating that they want to control and ensure their own privacy.

For other stakeholders than students, such as the learning analytics providers, consent can be important for legal reasons (among other reasons). It can be vital for learning analytics providers to get consent in order to be legally allowed to use learning analytics. This is for example required by the GDPR.

As consent is such a large gatekeeper for learning analytics, it is important that the processes around it is handled properly. It is the interest of all stakeholders, that this process handled in a good way. It is then concerning that half of the students, based on the questionnaire results, reported that they would feel pressured to give their consent if prompted by their higher education institution.

This is concerning as it undermines the institution of consent, making it invalid. This is a complex challenge to solve and the challenge of consent validity is not limited to learning analytic, which the results of the questionnaire shows, as the question is about consent in general, not specified to learning analytics. It can be relevant everywhere consent is obtained. This challenge is perhaps not noticed in practice, as consent would seem valid on paper even if students are pressured into giving it. But, it entails ethical challenges that need to be addressed.

Basing learning analytics on invalid consent is not optimal, but the challenge might be too big to be solved by learning analytics alone. It seems like a general tendency that this is the reality for all “digital” consent, and that it is pushed out of sight as it is a hard reality to deal with. The way consent works is flawed but realistically, the chances of improving it, or completely fixing it, are slim. Thus, the ethical problems of consent needs to be addressed on a higher level, than just inside one research field. It would also be naive to demand that learning analytics should fix the problem of consent validity, alone, but being aware of the ethical challenges of consent can promote new developments in the direction of valid consent. Perhaps learning analytics will use their methods to develop new, better ways of obtaining consent that ensures its validity, compared to solutions used today, represented by Terms and Conditions agreements.

Based on the questionnaire results, less than half of students, about three out of ten, report that they would read a Terms and Conditions agreement if provided this by their higher education institution. This relatively low number is a challenge of Consent as well as of Awareness. Other learning analytics researchers also find that few read Terms and Conditions agreements (Khalil, Prinsloo, and Slade, 2018;

Prinsloo and Slade, 2015), as highlighted in the literature review. Terms and Conditions agreements are customarily used to obtain consent, in addition to serving as a means of information regarding how personal information is handled and what services the information is used to support, prior to allowing the use of a service.

It poses a challenge when as few as three out of ten students would read such an agreement, if prompted. If also consent is asked in connection with such an agreement, it poses additional challenges. This could entail that students do not read the agreements, but accept it, as most students would feel pressured to give their consent. Similarly to consent, this is not a learning analytics specific challenge, but a general problem for all use of Terms and Conditions agreements.

Consent invalidity is not a learning analytics specific problem and need to be addressed on a higher level as this problem exists everywhere consent is obtained. Not doing anything about the problem, embracing the *status quo*, is the most comfortable position to take, but also the most unsatisfactory. It is not expected that learning analytics should solve the problem of consent validity and low Terms and Conditions reading rates, alone, but perhaps the research field can contribute to solutions with creative methods, close to the nature of learning analytics. This would improve learning analytics' own position as its integrity would be strengthened, as well as providing a valuable service for students.

Trust can be connected to multiple of the topics discussed above. It is for example possible to view the lack of interest in reading Terms and Conditions in connection with Trust. Trust is the fourth most important privacy principle to students, based on the results of the questionnaire. Nine out of ten students find it important. Trust is explained in the questionnaire as: *“That you have firm belief in, can rely on, and have confidence in your higher education institution to protect your privacy”*.

It is possible that students highly regard trust in order to outsource responsibility to their higher education institution. Trust can also be viewed as an obviously important privacy principle, not only for students but for anyone. If one is to share personal information, it is desirable that the ones you provide the information, can be trusted.

The challenge of trust is that it can be fragile. It takes time to build up, and can be eliminated by small mistakes. This highlights two important facets of trust. Trust between student and higher education institution needs to be established. This is perhaps intrinsic to the relation, but promoting trust can rarely be negative. Thus, it would be beneficial if learning analytics can assist in nourishing trust between student and higher education institution. This can possibly be done by promoting transparency of learning analytics (Drachler and Greller, 2016; Pardo and Siemens, 2014), as highlighted in the literature review. It can also be important that learning analytics are not implemented prematurely, as this can increase the risk for mistakes. This can for example be done by exploring the potential of learning analytics at a specific higher education institution, as this research does. The explanation of Trust provided to the students in the questionnaire weakens the comparison with other trust related findings. The reason is that Trust is defined in relation to ensuring privacy, but generally trust is considered as containing more than this.

7.3 Privacy priorities

(RQ3) What privacy priorities do Norwegian students have for learning analytics?

This question can be answered by providing a list of priorities based on the findings in the questionnaire, that have been discussed above. The following list represents what privacy priorities Norwegian students have for learning analytics:

- Privacy is highly important and should therefore be prioritized by learning analytics.
- Privacy in learning analytics should be centered around information privacy.
- Data security should be a high priority in learning analytics.
- Student information collected for learning analytics, should not be shared internally or externally.
- Learning analytics should require consent in a way that promotes consent validity.

- Terms and Conditions agreements are not an optimal solution for addressing awareness and consent. Terms and Condition agreements needs a reformation, or to be replaced by other solutions that to a larger degree promotes awareness and consent validity, in learning analytics.
- Opportunities for privacy self-management should be represented in learning analytics.
- Learning analytics should attempt to promote trust between student and higher education intuition.

The list is not exhaustive but outlines the most highly regarded privacy-values learning analytics should prioritize in order to ensure the privacy interests of the students. The discussion of privacy importance has been centered around the most important privacy principles, but it is important underscore that, all of the 17 principles were found important by the majority of students, thus, all of these should, most optimally, be considered by learning analytics implementors when designing learning analytics systems and services, as these principles could have implications that are not covered in this discussion.

None of the recorded priorities are impossible for learning analytics to implement or respect, proving promising opportunities for learning analytics in Norway. Although not impossible, some challenges can be difficult to address. The greatest challenge for learning analytics, discovered in this research, is to handle consent in a way that promotes consent validity. This also needs to be seen in connection with the use of Terms and Conditions agreements.

7.4 Learning analytics desirability

(RQ4) What learning analytics services and benefits are acceptable to Norwegian students?

Most students would accept the implementation of learning analytics. Students desire the benefits of learning analytics, more than the services it can provide. The most desirable learning analytics benefits for students, are: improved learning outcomes, improved courses, and improved grades. Most students do not want their personal information to be sold, nor to have their course progress compared to other students.

Following is the interpretation and discussion that lead to the answer above. When asked if they desire learning analytics, four out of ten students responds positively. When given information about specific benefits/services of learning analytics, slightly more students seem positive to learning analytics. It is possible that students do not really understand what learning analytics is or what it entails. This is not considered a challenge as it is believed that they understand the advantages/disadvantages of the different services/benefits, if they do not understand the concept fully, they at least understand the outcomes it can produce.

Improved learning outcome, improved courses, and improved grades are the most highly regarded learning analytics features according to students, based on results of the questionnaire. These findings are similar to findings by other learning analytics researchers, as documented in the literature review (e.g., Ifenthaler and Schumacher, 2016; Slade, Prinsloo, and Khalil, 2019; Tsai, Whitelock-Wainwright, and Gašević, 2020).

Based on results from the questionnaire, students desire the *benefits* of learning analytics higher than the *services*. They want improved learning outcomes, improved courses, and improved grades. Less desirable is personalized feedback, consecutive feedback, learning dashboards, and comparison to other students. Still some of these are considered desirable as between three and five, out of ten, would want such services.

There is no pattern in the results indicating that students only prioritize learning analytics features that only benefit themselves. For example, the second most highly desired learning analytics benefit is improving courses. This will not have a personal benefit for the student as their information would improve the course in the future, benefiting future students. Although, it is possible that this question could be interpreted as: students would receive improved courses instantly by implementing learning analytics, and not that courses are improved for the class coming after them. Nevertheless improving courses benefits not only students, but also their higher education institution.

These findings implicate that learning analytics needs to provide the benefits that are desirable for students, in order to be considered worth the sharing of personal information. Most important are improved grades, improved courses, and improved learning outcome. This is obviously more nuanced than this statement indicates, it would for example be naive to suggest that learning analytics will, with no efforts from the students' side, increase their grades or improve their learning outcomes. Another observation that can be drawn from these findings, is that students are interested in the potential of learning analytics, and they are interested in learning analytics that support them on the way to higher grades, better courses or improved learning outcomes.

It is also important to underscore that learning analytics offer more than what is promoted to the students in the questionnaire, as the list of services/benefits is not exhaustive. Thus, learning analytics could still be attractive for students even without providing any of the mentioned services or benefits. However, for this to be stated confidently, new examinations of student opinions need to be made, targeted at exploring other services and benefits.

7.5 Summary

In this chapter, the research questions have been answered, and discussed. The discussion has explored what principles are the most important in learning analytics, both in regards to the general privacy debate in learning analytics, as well as in relation to the student perspective. Student perceptions of privacy has also been discussed, with a focus on information privacy. What services and benefits students want from learning analytics was the last topic discussed.

8 Conclusion

This research has explored student perceptions of privacy in relation to learning analytics. This was done within the frames of quantitative research, as a questionnaire administered to gather empirical data. The questionnaire was developed with inspiration from secondary analysis as well as a literature review, its main feature was that it utilized privacy principles as indicators to measure privacy perceptions. The questionnaire was conducted at the University of Bergen, obtaining 394 student responses. The responses were analysed using descriptive statistics. A stacked polar bar chart (Figure 26) displaying the results of student's opinions on 17 different principles were the main result from the descriptive statistics. In the Discussion, Chapter 7, four research questions were answered:

1. What privacy principles are most relevant for learning analytics?
2. How do Norwegian students perceive privacy in general, and in relation to learning analytics?
3. What privacy priorities do Norwegian students have for learning analytics?
4. What learning analytics services and benefits are acceptable to Norwegian students?

These research questions were answered in the following way:

The most relevant privacy principles for learning analytics, based on the findings in the literature review, as well as the questionnaire, based on ranking, is: consent, students control, and data security. Consent is considered especially relevant as it is found across multiple analyses. From a student perspective, data security and consent are found to be high ranking privacy principles, in multiple analyses and are therefore considered highly relevant for students.

Based on the findings from the questionnaire, the following can be said about student perceptions. In general, students perceive privacy to be highly important, more important than the population in general. Most students have an information-privacy centered understanding of privacy. Seven out of ten student privacy perceptions fit inside the established conceptualization of *personvern* and *personopplysningsvern*. Privacy self-management is an important part of most student perceptions of privacy, students want to control and ensure their own privacy. Many students also wants privacy expert-management. Students' view on privacy management is reflected in the current privacy management structure in Norway; a shared responsibility of privacy management between self-management and expert-management.

The following statements are important privacy priorities for learning analytics, based on student opinions, from the questionnaire results: Privacy is highly important and should therefore be prioritized by learning analytics; Privacy in learning analytics should be centered around information privacy; Data security should be a high priority in learning analytics; Student information collected for learning analytics, should not be shared internally or externally; Learning analytics should require consent in a way that promotes consent validity; Terms and Conditions agreements are not an optimal solution for addressing awareness and consent, therefore, Terms and Condition agreements needs a reformation, or to be replaced by other solutions that to a larger degree promotes awareness and consent validity, in learning analytics; Opportunities for privacy self-management should be represented in learning analytics; and, Learning analytics should attempt to promote trust in the relation between student and higher education intuition.

Most students would accept the implementation of learning analytics at their higher education institution. Students desire benefits of learning analytics, more than the services it can provide. The most desirable learning analytics benefits for students, are: improved learning outcomes, improved courses, and improved grades. Students do not want their personal information to be sold, nor to have their course progress compared to other students.

This implicates that learning analytics is desirable among Norwegian students, but it needs to respect their privacy. This is possible to do by implementing the *personvern* and *personopplysningsvern* concepts as well as attempting to improve consent validity.

8.1 Contributions

The research has contributed insights into student perceptions of privacy for learning analytics at a Norwegian higher education institution, these insights are helpful for stakeholders wanting to implement learning analytics in the Norwegian higher education sector.

Another contribution is the mapping of learning analytics desirability among Norwegian students. This is also helpful information for stakeholders wanting to implement learning analytics in the Norwegian higher education sector.

The literature review produced, is another contribution. Furthermore the privacy debate in learning analytics was quantified, highlighting important privacy principles for the research field in general, and important privacy principles for students, in relation to learning analytics. Using these findings together with other privacy theory to tailor questions to ask students, is also a contribution, as it represents a different take on exploring privacy.

8.2 Limitations

The greatest limitation in this research is related to generalizability. The results from this research is not generalizable because of multiple factors. The most central reasons are the total anonymity of the questionnaire, and the lack of a representative sample. As a consequence, the results cannot be generalized to apply to a population outside of the students that answered the questionnaire. Another limitation is the scale of the implemented questionnaire, as it was only implemented at one higher education institution (University of Bergen), where it got a low response rate.

Other limitations are also related to methodology. The use of secondary analysis proved less useful than expected. The means of comparison between responses from the questionnaire was not possible because of the changes made to the questions. The consequence of this was that only one of the secondary analysis results were relevant for the research.

Questionnaire limitations are another relevant point. The questionnaire specific limitations have been described earlier in Chapter 4, here the use of an English questionnaire in a Norwegian speaking country is considered one of the biggest limitations, probably causing the response rate to drop.

The analysis is considered limited as it consists only of descriptive statistics. A more advanced analysis could have yielded interesting findings related to privacy. For example by looking at relationships between the different privacy principles, asked about in the questionnaire.

Another limitation in connection with the scope of the research, is that the research only focused on one stakeholder, the students. More stakeholders could have been included to get a fuller view of privacy in learning analytics.

The legal perspective of privacy is given little attention in this research, this can be considered a limitation as privacy and Law are tightly connected topics.

Privacy has gotten a lot of attention in this research. A limitation with this, is that privacy comes of as the only important value in existence. Privacy is not a *prima facie* duty, even though it is considered important. Other interests should also be focused.

8.3 Future work

Future work should address more learning analytics stakeholders in the Norwegian higher education sector, in order to get a better idea of the stakeholders' interests. When a better picture of all stakeholders' interests are known, it will be possible to balance privacy against other interests as discussed in Chapter 2.

Future work should also perform this questionnaire after the pandemic, to indicate if the results of this research was affected by the pandemic. This would be interesting to research, as the forced switch to digital schooling may have caused privacy perceptions to change.

Future work should also address the challenge of consent validity and attempt to develop consent solutions for learning analytics that promotes consent validity.

More research should also be put in to exploring privacy self-management. Results from this research indicate that students want privacy self-management, but it is uncertain if they understand what consequences it entails.

To research the topic of student privacy in learning analytics has been an interesting task, it has been both challenging and rewarding to research a topic (student perceptions of privacy in learning analytics) so foreign to most people. The challenge of researching privacy lies in the ambiguity and elusiveness of the concept. Privacy is something that is hard to define, and hard to express why is important, but it *feels* important. Learning analytics has great potential to increase student learning outcomes in an ever more digital world. The challenge is to find the right balance between data driven technology and privacy.

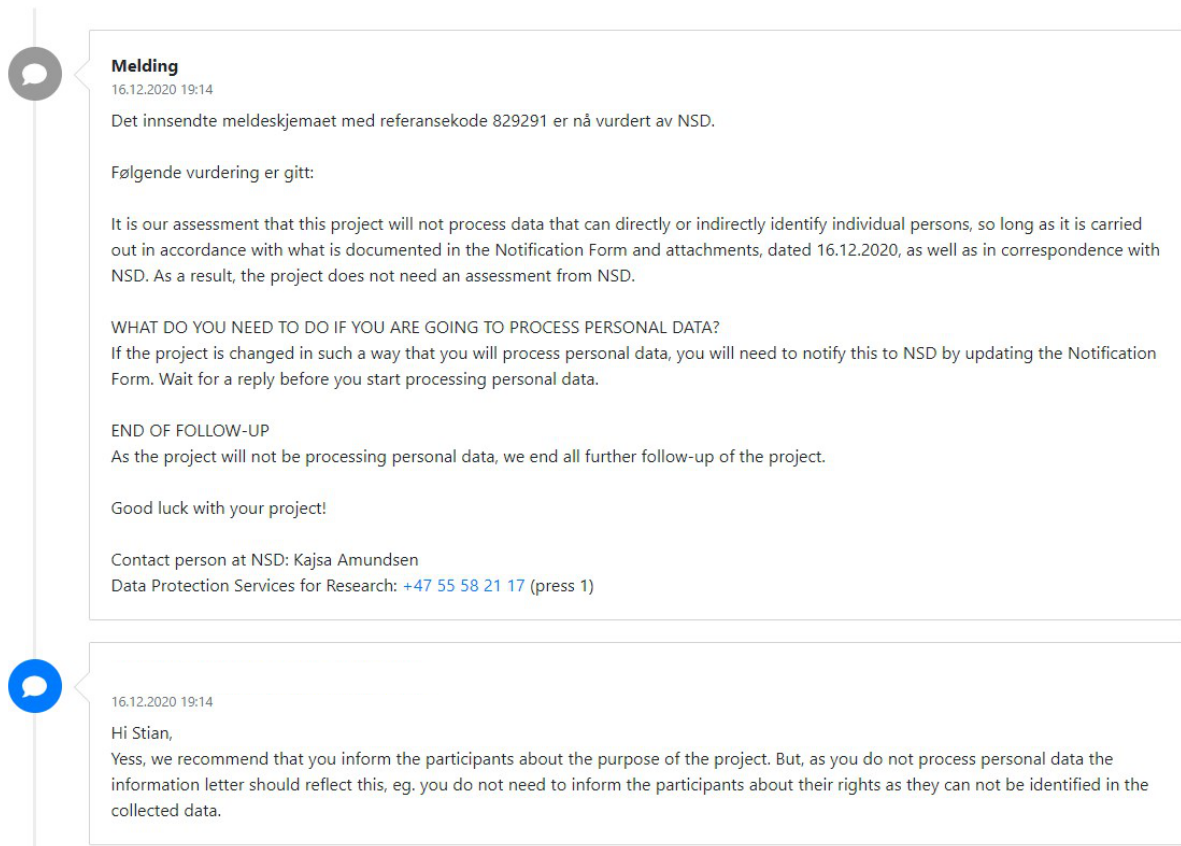
References

- 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(4), 22–29. [%3CGo%20to%20ISI%3E://WOS:000403339400004](#)
- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at purdue: Using learning analytics to increase student success. *Proceedings of the 2nd international conference on learning analytics and knowledge*, 267–270.
- Arnold, K. E., & Sclater, N. (2017). Student perceptions of their privacy in leaning analytics applications. *Proceedings of the seventh international learning analytics & knowledge conference*, 66–69.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>
- Bell, J., & Waters, S. (2018). *Doing your research project*.
- Bellini, C., De Santis, A., Sannicandro, K., & Minerva, T. (2019). Data management in learning analytics: Terms and perspectives.
- Blaxter, L., Hughes, C., & Tight, M. (2010). *How to research, 4th edition*. Open University Press. <https://doi.org/doi:10.1036/9780335238699>
- Botnevik, S., Belinskiy, A., Asotic, E., Platou, H., Søvik, T., & Slavkovik, M. (2020). Addressing the ethical principles of the norwegian national strategy for ai in a kindergarten allocation system. *Norsk IKT-konferanse for forskning og utdanning*, (1).
- Botnevik, S., Khalil, M., & Wasson, B. (2020). Student awareness and privacy perception of learning analytics in higher education. *European Conference on Technology Enhanced Learning*, 374–379.
- Brown, M., & Klein, C. (2020). Whose data? which rights? whose power? a policy discourse analysis of student privacy policy documents. *The Journal of Higher Education*, 91(7), 1149–1178.
- Bryman, A. (2015). *Social research methods*. Oxford University Press.
- Datatilsynet. (2019). Personopplysninger — Datatilsynet. Retrieved March 13, 2021, from <https://www.datatilsynet.no/rettigheter-og-plikter/personopplysninger/>
- Denscombe, M. (2017). *The good research guide : For small-scale social research projects*.
- Drachler, 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*, 89–98.
- Falcão, T. P., Mello, R. F., Rodrigues, R. L., Diniz, J. R. B., Tsai, Y.-S., & Gašević, D. (2020). Perceptions and expectations about learning analytics from a brazilian higher education institution. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 240–249.
- Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5-6), 304–317.
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68–84.
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let’s not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71.
- GDPR. (2016). The general data protection regulation (eu) 2016/679. <https://doi.org/10.1177/0743915619858924>.
- Greller, W., & Drachler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3), 42–57.
- Grønmo, S. (2016). *Samfunnsvitenskapelige metoder*. Fagbokforlaget.
- Haythornthwaite, C. (2017). An information policy perspective on learning analytics. *Proceedings of the seventh international learning analytics & knowledge conference*, 253–256.
- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Etr&D-Educational Technology Research and Development*, 64(5), 923–938. <https://doi.org/10.1007/s11423-016-9477-y>
- Khalil, M., Prinsloo, P., & Slade, S. (2018). The unbearable lightness of consent: Mapping mooc providers’ response to consent. *Proceedings of the fifth annual ACM conference on learning at scale*, 1–11.
- Macfadyen, L. P., & Dawson, S. (2010). Mining lms data to develop an “early warning system” for educators: A proof of concept. *Computers & education*, 54(2), 588–599.
- Misiejuk, K., & Wasson, B. (2017). State of the field report on learning analytics.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *Annals of internal medicine*, 151(4), 264–269.

- NOU:1997:19. (1997). *Et bedre personvern: Forslag til lov om behandling av personopplysninger* (Report). <https://www.regjeringen.no/no/dokumenter/nou-1997-19/id140970/?ch=1>
- NOU:2009:1. (2009). *Individ og integritet: Personvern i det digitale samfunnet* (Report). <https://www.regjeringen.no/no/dokumenter/nou-2009-1/id542049/?ch=1>
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.
- Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. *Proceedings of the fifth international conference on learning analytics and knowledge*, 83–92.
- Regjeringen.no. (2019). Hva er personvern? Retrieved March 6, 2021, from %5Curl%7Bhttps://www.regjeringen.no/no/tema/statlig-forvaltning/personvern/hva-er-personvern/id448290/%7D
- 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. <https://doi.org/10.3339/fpsyg.2016.01959>
- Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The information society*, 32(2), 143–159.
- Sci Team. (2009). Science of science (sci2) tool. *Indiana University and SciTech Strategies*, 379.
- Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Journal of educational technology & society*, 15(3), 3–26.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529.
- 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*, 235–244.
- SoLAR. (2021). What is learning analytics? Retrieved March 7, 2021, from <https://www.solaresearch.org/about/what-is-learning-analytics/>
- Solove, D. J. (2009). *Understanding privacy*. Harvard University Press.
- Sun, K., Brooks, C., Mhaidli, A. H., Schaub, F., & Watel, S. (2018). Taking student data for granted? a multi-stakeholder privacy analysis of a learning analytics system. *EDM’18 Workshop on Policy and Educational Data Mining*.
- Technopedia. (2021). What is Data Exhaust? - Definition from Techopedia. Retrieved March 7, 2021, from <https://www.techopedia.com/definition/30319/data-exhaust>
- Tor Guttu et al. (Ed.). (2021). Naob – det norske akademis ordbok. Retrieved March 7, 2021, from <https://naob.no/ordbok/privatliv>
- 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*, 230–239.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500–1509.
- Vu, P., Adkins, M., & Henderson, S. (2020). Aware, but don’t really care: Students’ perspective on privacy and data collection in online courses. *Journal of Open Flexible and Distance Learning*, 23(2), 42–51.
- West, D., Luzecky, 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*, 36(2), 60–70.
- Whitelock-Wainwright, A., Gasevic, D., Tejeiro, R., Tsai, Y. S., & Bennett, K. (2019). The student expectations of learning analytics questionnaire. *Journal of Computer Assisted Learning*, 35(5), 633–666. <https://doi.org/10.1111/jcal.12366>
- Whitelock-Wainwright, A., Gasevic, D., Tsai, Y. S., Drachsler, H., Scheffel, M., Mu?oz-Merino, P. J., Tammets, K., & Kloos, C. D. (2020). Assessing the validity of a learning analytics expectation instrument: A multinational study. *Journal of Computer Assisted Learning*, 36(2), 209–240. <https://doi.org/10.1111/jcal.12401>
- Zinoviev, D. (2018). *Complex network analysis in python: Recognize-construct-visualize-analyze-interpret*. Pragmatic Bookshelf.

9 Appendices

A Communication with NSD



The figure displays two chat messages from NSD. The first message, titled 'Melding', is dated 16.12.2020 19:14 and contains the following text: 'Det innsendte meldeskjemaet med referansekode 829291 er nå vurdert av NSD. Følgende vurdering er gitt: It is our assessment that this project will not process data that can directly or indirectly identify individual persons, so long as it is carried out in accordance with what is documented in the Notification Form and attachments, dated 16.12.2020, as well as in correspondence with NSD. As a result, the project does not need an assessment from NSD. WHAT DO YOU NEED TO DO IF YOU ARE GOING TO PROCESS PERSONAL DATA? If the project is changed in such a way that you will process personal data, you will need to notify this to NSD by updating the Notification Form. Wait for a reply before you start processing personal data. END OF FOLLOW-UP As the project will not be processing personal data, we end all further follow-up of the project. Good luck with your project! Contact person at NSD: Kajsa Amundsen Data Protection Services for Research: +47 55 58 21 17 (press 1)'. The second message, also dated 16.12.2020 19:14, says: 'Hi Stian, Yes, we recommend that you inform the participants about the purpose of the project. But, as you do not process personal data the information letter should reflect this, eg. you do not need to inform the participants about their rights as they can not be identified in the collected data.'

Melding
16.12.2020 19:14

Det innsendte meldeskjemaet med referansekode 829291 er nå vurdert av NSD.

Følgende vurdering er gitt:

It is our assessment that this project will not process data that can directly or indirectly identify individual persons, so long as it is carried out in accordance with what is documented in the Notification Form and attachments, dated 16.12.2020, as well as in correspondence with NSD. As a result, the project does not need an assessment from NSD.

WHAT DO YOU NEED TO DO IF YOU ARE GOING TO PROCESS PERSONAL DATA?
If the project is changed in such a way that you will process personal data, you will need to notify this to NSD by updating the Notification Form. Wait for a reply before you start processing personal data.

END OF FOLLOW-UP
As the project will not be processing personal data, we end all further follow-up of the project.

Good luck with your project!

Contact person at NSD: Kajsa Amundsen
Data Protection Services for Research: [+47 55 58 21 17](tel:+4755582117) (press 1)

16.12.2020 19:14

Hi Stian,
Yes, we recommend that you inform the participants about the purpose of the project. But, as you do not process personal data the information letter should reflect this, eg. you do not need to inform the participants about their rights as they can not be identified in the collected data.

Figure A.1: Communication with NSD

B Questionnaire

The image shows a screenshot of a questionnaire cover letter. At the top left is the logo of Universitetet i Bergen. The title is 'DigiTrans: The Digital Transformation of UiB' with a subtitle 'Exploring Student's Perceptions of Privacy for Learning Analytics'. The text explains the project's goal, the importance of learning analytics, and the purpose of the questionnaire. It details the voluntary nature of participation, privacy guarantees, and contact information for the research team. At the bottom, there are 'PREVIOUS' and 'NEXT' buttons, a progress bar at 8%, and the names of the masters student (Stian Botnevik) and supervisor (Dr. Mohammad Khalil).

UNIVERSITETET I BERGEN

DigiTrans: The Digital Transformation of UiB

Exploring Student's Perceptions of Privacy for Learning Analytics

As part of Theme D in the [DigiTrans](#) project we are using learning analytics to investigate student and instructor use of **digital learning tools** at UiB.

Learning analytics in the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.

The field of learning analytics is relatively new, and applications of learning analytics are starting to appear in higher education institutions. If learning analytics is to have a bright future, it is important to get the opinions and insights of the stakeholders to be able to make good and stable applications. Students, as the biggest stakeholder in learning analytics applications, should be addressed and we invite you to participate in a questionnaire related to the use of student data.

This questionnaire is arranged as part of a master's research project where the main goal is to **explore student's perceptions of privacy related to learning analytics**. Privacy principles are central to data collection for learning analytics. Student perceptions about the use their data is very important, and we would like your input. We ask general questions about privacy and particular questions on the importance of certain privacy principles that are central in the learning analytics literature.

Purpose of the questionnaire
The aim of this questionnaire is to explore how students' views on privacy affects the outlooks for learning analytics in Norwegian higher education institutions.

Who is responsible for the research project?
Centre for the Science of Learning & Technology (SLATE), UiB is the institution responsible for the project.

Why are you being asked to participate?
You are asked to participate in this questionnaire because you are a student at the University of Bergen.

What does participation involve for you?
If you choose to participate you will fill out a questionnaire about privacy and privacy principles that will take approximately 12 minutes. The questionnaire contains 9 questions (37 small subquestions). Your answers will be recoded digitally.

Participation is voluntary
Participation in the questionnaire is voluntary. There will be no negative consequences for you if you choose not to participate.

Your personal privacy
We do not handle any personal information about you. The questionnaire is completely anonymous. The only information we collect is your answers in the questionnaire.

- We use SurveyXact (<https://www.surveymxact.no/>) to collect and process your answers.
- We do not have access to any personal information about you.
- The questions require no personal information to be provided.

As a participant in this questionnaire you will not be able to be recognized in a publication.

Where can I find out more?
If you have questions about the questionnaire, contact:

- Centre for the Science of Learning & Technology (SLATE) via *Mohammad Khalil*, Mohammad.Khalil@uib.no (supervisor) or *Stian Botnevik*, Stian.Botnevik@uib.no (master's student).
- For questions about DigiTrans, contact project leader Professor Barbara Wasson, Director, SLATE barbara.wasson@uib.no

Yours sincerely,

Masters student Stian Botnevik	Supervisor Dr. Mohammad Khalil
------------------------------------------	------------------------------------------

PREVIOUS NEXT

8%

A

Figure B.1: Questionnaire cover letter (A).

In this questionnaire you will be asked some questions about your opinions in relation to **"personal information" (personopplysninger)**. By this term we mean: **All types of information that can be connected to you as an individual – not including the categories of "sensitive personal information", as described below.** Examples of personal information are; name, address, phone number, email-address, birth date, pictures, video, audio recordings, dynamic ip-address, behavioral patterns, log of your interaction with your higher education institution's systems, electronic communication within your higher education institution's systems.

Personal information does **NOT** include **"sensitive personal information"** which contains the following categories:

- information on racial or ethnic origin
- information on political opinion
- information about religion
- information on philosophical belief
- information on union membership
- genetic information
- biometric information (when the purpose of treatment is to uniquely identify someone)
- health information
- sexual information
- information about sexual orientation

This explanation of what the term personal information contains, and not contains, is provided to help you understand what type of information are the subject of the following questions.

 16%

B

(1/9) Are you a student enrolled in a Norwegian higher education institution?

- Yes
 No

 25%

C

Figure B.2: Distinction between personal information and sensitive personal information (B) and question 1/9 (C).

(2/9) Which of the following statements is closest to how you understand *privacy*?

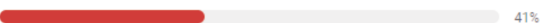
- a I understand privacy as the right to be let alone.
- b I understand privacy as the state in which one is not watched or disturbed by others.
- c I understand privacy as the right to limit others access to myself as a person.
- d I understand privacy as the right to decide when, how, and to what degree information about me is shared with others.
- e I understand privacy as the right to safeguard my personal integrity; safeguard my opportunity for private life, self-determination (autonomy) and self-expression.
- f I understand privacy as the state or condition of being alone, undisturbed, as a matter of choice or right; seclusion; freedom from interference or intrusion.
- g Other – please enter (English or Norwegian).



D

(3/9) How important is privacy to you?

- Very important
- Important
- Neutral
- Less important
- Unimportant
- Don't know / no opinion



E

Figure B.3: Question 2/9 (D) and 3/9 (E).

(4/9) To what extent do you agree with the following statements?

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Don't know / no opinion
a I feel that I have little control over how personal information about me is stored and used by my higher education institution.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b I feel powerless when it comes to having control over personal information about me in relation to my higher education institution.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c I want to be able to control and ensure my own privacy.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d I want privacy experts or larger bodies (like an independent administrative body) to control and ensure my privacy.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

PREVIOUS

NEXT

50%

F

(5/9) How desirable is it for you to have your personal information collected, processed and analyzed by your higher education institution in order for them to provide services that has the potential to improve your learning outcome?

- Very desirable
- Desirable
- Neutral
- Undesirable
- Very undesirable
- Don't know / no opinion

PREVIOUS

NEXT

58%

G

Figure B.4: Question 4/9 (F) and 5/9 (G).

(6/9) Imagine that your higher education institution is collecting, processing and analyzing your personal information in order to improve your learning outcome. How important are the following privacy principles to you?

	Very important	Important	Neutral	Less important	Unimportant	Don't know / no opinion
a	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
l	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
m	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
n	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
o	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
p	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
q	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

PREVIOUS NEXT

66%

H

Figure B.5: Question 6/9 (H). 6/9. Oxford English Dictionary has inspired the explanations of the different terms. <https://www.oed.com/>

(7/9) If you were asked to give consent to have your personal information collected, processed and analyzed, would it be likely that you would read a Terms and Conditions declaration before accepting/rejecting?

- Very likely
- Likely
- Neutral
- Unlikely
- Very unlikely
- Don't know / no opinion

PREVIOUS

NEXT



I

(8/9) If you were asked to give consent to have your personal information collected, processed and analyzed, by your higher education institution – would you feel pressured to give consent?

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree
- Don't know / no opinion

PREVIOUS

NEXT



J

Figure B.6: Question 7/9 (I) and (8/9) (I)

(9/9) What services would you accept/deny your personal information to be used for, by your higher education institution?

	Accept	Maybe accept	Deny	Don't know / no opinion
a Improving your learning outcome.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
b Improving your grades.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
c Giving you consecutively information about how well you are doing in your subjects.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
d Giving you a visualized illustration of your course progress using graphs and charts.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
e Giving you an overview of how well you are doing in your courses compared to other students.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
f Giving you personalized feedback and suggestions related to your courses.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
g Improving the courses offered by your higher education institution.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
h Increasing graduation rates of your higher education institution.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
i Economic gain through sale of your personal data.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

PREVIOUS

NEXT

91%

K

Thank you for your participation! If you have any questions regarding this questionnaire or the project, feel free to send me an email: Stian.Botnevik@uib.no

PREVIOUS

FINISH

100%

L

Figure B.7: Question 9/9 (K) and end note with contact information (L).

C Large figures

Heatmap of questionnaire responses

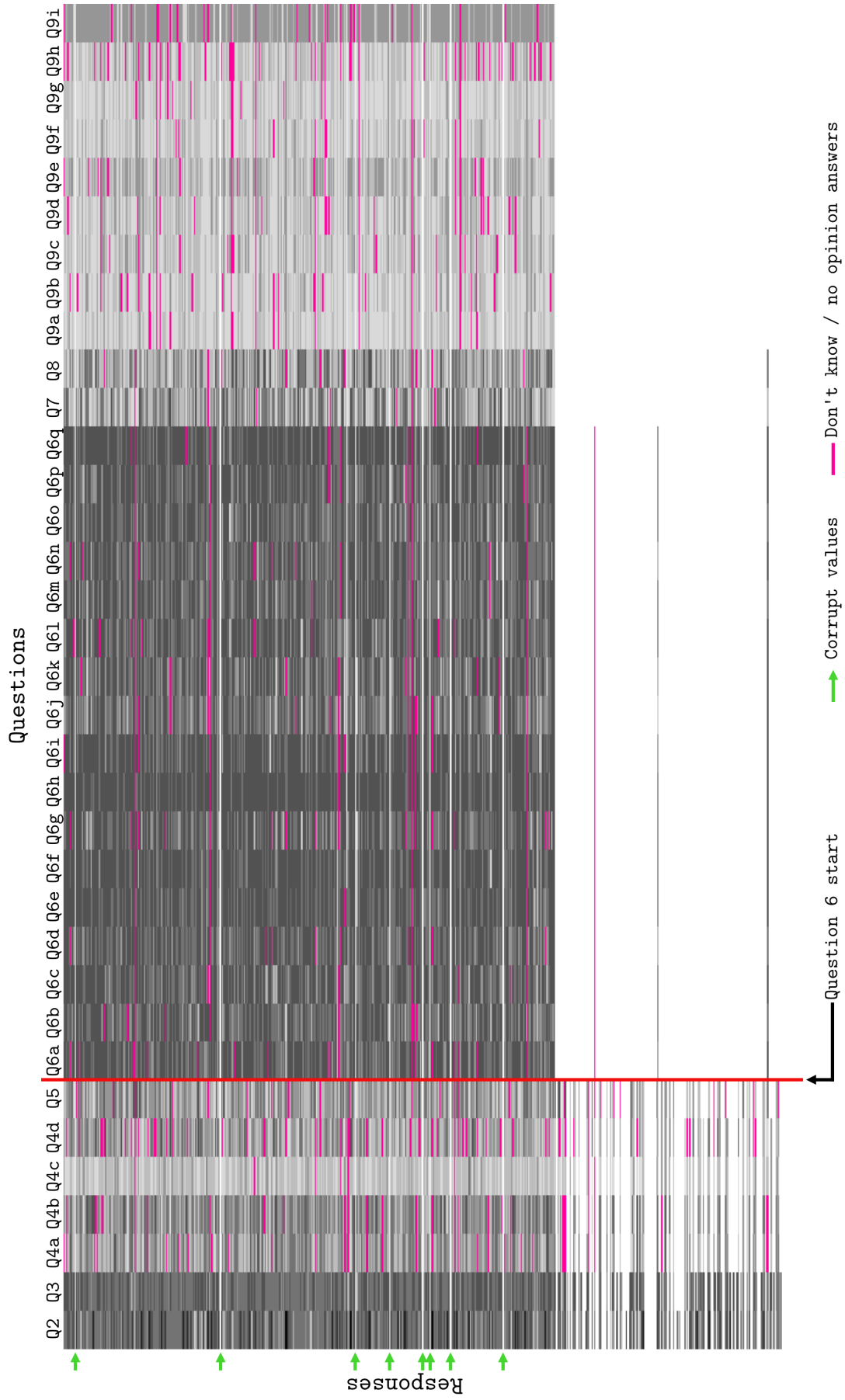


Figure C.1: Heatmap of participation, large.