The privacy paradox and its implications for learning analytics

Author 1
Institute
City, Country
email

Author 2
Institute
City, Country
email

Author 3
Institute
City, Country
email

ABSTRACT

Learning analytics promises to support adaptive learning in higher education. However, the associated issues around privacy protec-tion, especially their implications for students as data subjects, has been a hurdle to wide-scale adoption. In light of this, we set out to understand student expectations of privacy issues related to learn-ing analytics and to identify gaps between what students desire and what they expect to happen or choose to do in reality when it comes to privacy protection. To this end, an investigation was carried out in a UK higher education institution using a survey (N=674) and six focus groups (26 students). The study highlight a number of key implications for learning analytics research and practice: (1) purpose, access, and anonymity are key benchmarks of ethics and privacy integrity; (2) transparency and communication are key levers for learning analytics adoption; and (3) information asymmetry can impede active participation of students in learning analytics.

CCS CONCEPTS

• Applied computing → Computer-assisted instruction; • Human-centered computing → Empirical studies in HCI .

KEYWORDS

Learning analytics, privacy, expectations, privacy paradox, higher education

ACM Reference Format:
Author 1, Author 2, and Author 3. . The privacy paradox and its implications for learning analytics. In ,. ACM, New York, NY, USA, 11 pages.

1 INTRODUCTION

In the context of higher education institutions (HEIs), there is an increasing demand to measure, demonstrate and improve perfor-mance. As a result, learning analytics emerges as a new solution to addressing issues around retention, progression, and enhancement of student success [17]. In contrast to educational data mining and academics analytics, learning analytics focuses on solving educa-tional challenges [17], leveraging human decisions [36], and sup-porting learning [18]. However, concerns of surveillance, privacy breach, and datafication have been raised regarding the constant collection and analysis of data about learners in higher education [39, 42, 47]. A number of scholars have made an attempt to ad-dress challenges of privacy protection in learning analytics with frameworks that promote transparency and strengthen students' control over personal data [10, 15, 20, 28, 30, 37, 46]. In Europe, the European General Data Protection Regulation 2016/679 (GDPR)

[43] has also placed pressure on higher education institutions to up-date existing practices to ensure that consent is first sought before any collection of personal data takes place. This is a contentious issue for the learning analytics community where experts advocate for consent to be obtained only prior to actioning interventions [10, 34], whereas students expect institutions to seek consent be-fore any handling of data [38]. With the fiduciary duty to provide high-quality educational service, how to collect and use student data effectively and ethically is a pressing question for HEIs. Im-portantly, the beliefs and expectations that students hold regarding how HEIs should use their data are important factors in the overall learning experience.

Although the awareness of including students in the discourse of ethics and privacy issues in the context of learning analytics has risen in recent years [20, 31, 33, 39, 42, 45], existing studies focusing on involving this group of stakeholders are only a handful. In order to inform the development of institutional policy and strat-egy for learning analytics in a UK HEI before learning analytics was formally introduced to students, we set out to explore student awareness and perceptions of the usage of their personal and edu-cational data in educational settings. Our investigation was guided by two research questions:

(1) What uses of personal and educational data do students consider to be legitimate and appropriate in a higher education context?
(2) Are there gaps between student perceptions or expectations of privacy and the actions that they have taken to protect their data?

Different from existing studies, this paper draws upon theories of privacy paradox and contextual integrity [2, 22, 24] to understand elements in risk-benefit assessments that students may undertake when facing decisions of sharing data to enable learning analytics related activities and services. This study employed mixed methods using both survey and focus group methods. In the scope of this paper, we focus our discussion on findings that derive from the focus groups, and compare them to findings from the responses to privacy-related items in the survey when discussing the first research question.

The study emphasises a user-centred approach to learning an-alytics. It contributes insights into the understanding of student perceptions of privacy and expectations for the use of their data for learning analytics, summarised below:

The privacy paradox, LAK'20, March 23–27, 2020, Frankfurt, Germany

- Privacy paradox exists in educational contexts where data enables certain support
- Students are protective of personal data but take contradictory actions
- Students generally take a cure approach to data protection
- Trust and imbalanced power-relationships exacerbate information asymmetry

The article concludes with recommendations useful for format-ting institutional strategies and polices for learning analytics.

2 THEORETICAL BACKGROUND

In this section, we first present common challenges regarding pri-vacy in the context of learning analytics, followed by a discussion about the privacy paradox to understand how students might per-ceive learning analytics in terms of its privacy implications and the extent to which their beliefs in privacy were reflected through actions. We then discuss expectation literature to understand the judgements that students may make concerning the use of learning analytics and to better manage the satisfaction of user experience. These theoretical frameworks informed the design and analysis of this study.

2.1 Privacy challenges in the context of learning analytics

Learning analytics benefits from a wide array of data including engagement data (information generated in physical and virtual learn-ing environments, such as attendance, log-ins and card swipes), aca-demic data (grades and educational history), and background data (age, gender, ethnicity, nationality, and economic background)[14]. The collection and use of student data inevitably challenges learning analytics with accusations of surveillance and potential breaches of privacy [15, 28]. A number of problems have emerged in discourses around these issues. First, the range and types of data collected have led to concerns about intrusions into spaces generally deemed private or personal [28]. As technology evolves with the capacity to record and analyze interactions in a learning environment, it is now possible to explore learning scenarios using analytics techniques which collect large datasets from learners in a constant manner. Moreover, learning analytics can be viewed as optimally useful for individual students if the pattern matching stage (following a pat-tern identifying stage)[10] can lead to personalized interventions to close a learning feedback loop. To do so, data must retain a cer-tain degree of individual linkages so as to deliver such customized interventions [32].

Second, the ownership of data is difficult to define after data has been collected and computed, which produces new sets of in-formation and data [3, 15, 27]. This complexity lies in multiple contributions from data subjects, data infrastructure owners, and data analysts, who all may have the right to claim some share of the data. Third, conflicts of autonomy exist in the very nature of learning analytics, which upholds the value of student agency in learning while decreasing the autonomy of students in the process of a massive and constant data collection process [32], that is arguably performed in an 'anti-democratic' manner [47]. Finally, the asymmetrical power relationship between students and their higher education institution makes informed consent problematic. Institutions may be transparent about their data practices, but the complexity of algorithms still makes analytics products a 'black box' for many [32]. Moreover, due to information asymmetries be-tween data collectors and data subjects, data subjects tend to have limited knowledge about who can access their data, what they do with the data, and what the consequences of privacy intrusions may be [1, 15]. As a result, control over one's own data becomes less definitive.

The privacy challenges identified above need to be considered carefully when higher education moves towards a more system-atic and large-scale adoption of learning analytics. Moreover, to avoid unintentional harm to students, institutions must take into account students' awareness and attitudes regarding the use of their personal data and understand situations where students may associate learning analytics with risks that are higher than benefits, or the other way around. In order to understand the relationship between individuals' privacy beliefs and privacy related behavior or decisions, we will explore elements that constitute the concept of privacy and the phenomenon of the privacy paradox in the next section.

2.2 Privacy paradox in user behavior

Privacy is a state in which an individual is free from being dis-turbed or observed by others [40]. Specifically, information privacy concerns data about or generated by individuals [4]. Discussions around the protection of privacy in the information dimension tend to revolve around the ownership of personal data and implications thereof [9]. The concept of ownership suggests that individuals have the right to use, dispense, and dispose of such data. However, it is simultaneously true that individuals often take few precautions to protect their data, or share their data in ways contrary to their protective attitudes [2, 22]. These incongruent privacy beliefs and behaviors constitute a prominent social phenomenon–the privacy paradox, defined from the economic perspective as follows [25](p. 107):

> Although consumers seem to be concerned about their privacy as reflected in their intentions to disclose (e.g., measured via "willingness to provide information"), anecdotal evidence suggests their behaviors diverge from their intentions to disclose personal details.

Barth and Jong [2] reviewed 32 studies (35 theories) regarding privacy decision-making and concluded that a user's willingness to disclose private information is a result of a risk-benefit evaluation or not having it at all. In circumstances where little or no risk-benefit evaluation is taken, there is usually insufficient consideration of risks, and thus decisions are made solely based on significantly prevalent benefits. According to Barth and Jong, little or no risk assessment occurs due to a number of factors, including trust in a service provider, the acceptance of an imbalanced power state (the 'all-or-no service principle'), lack of concern about data protection, or knowledge deficiency. By contrast, when a rational calculation of risks and benefits is involved, the decision to disclose personal information will be made where the expected benefits for doing so outweigh the perceived risks. The cost-benefit evaluation, however, is a subjective one, based on each individual's own belief system. For example, a user may consider the effort and loss of time in reading

a lengthy and complex policy to outweigh the perceived risk of disclosing personal information, thereby consciously deciding that the benefits of using a service outweigh privacy risks, without actually understanding the consequences of sharing personal data. This is described as rational ignorance whereby individuals avoid assessing their privacy risks because the perceived effort and loss of time in finding out the situation outweigh the perceived risk of disclosing personal information [1].

The risk-benefit evaluation of privacy disclosure is highly con-textual. Nissenbaum [24] posits contextual integrity as a benchmark of privacy. According to Nissenbaum, there are two types of in-formation norms: norms of appropriateness, and norms of flow or distribution. The former considers the appropriateness of revealing information about individuals in a particular context, while the lat-ter regulates the distribution of information (how data flows from one party to another or others). The two norms together define contextual integrity, which is achieved when neither of the norms is violated. Nissenbaum suggested a number of parameters to evaluate contextual integrity, including the nature of information, the rela-tionship between the information and the given context, the roles (agents of information) involved in the context, the relationships between the roles, the rules of information flow, and how changes made in a context might affect its social values. These parameters can be very useful in understanding a person's perception and ex-perience of privacy. For example, online users tend to give lower monetary value to their browsing history than their offline personal information [22]. This phenomenon can be analysed by looking at the nature of information and its relationship with the context. One study revealed that middle-aged women tend to find it embar-rassing to reveal their date of birth to a younger male customer service employee [9]. This phenomenon can be understood by ana-lyzing the roles of the agents in the network and their relationships. Norberg and others [25] further highlight a trust relationship as a key factor of a disclosure behaviour, while perceived risks are a factor in behavioural intention, but not necessarily action. That is to say, in a higher education context, a student might disclose their data to a trustful party (educator or institution) despite perceiving potential risks associated with sharing their data. Similarly, Coll

[9] observed in an ethnographic study that people tended to define privacy subjectively in relational terms (i.e., personal relationships in a context) or in association with individuals' freedom of choice (i.e., the self-determination principle–every person should serve as a proactive actor of his or her own privacy).

The abovementioned studies have suggested that a decision to reveal personal information can be rather inconsistent among indi-viduals across different contexts, and discrepancies commonly exist in an individual's perceptions and actions towards the protection of information privacy under the influence of various factors in a given context. In fact, individuals tend to desire for more power over their personal data and more agency in the use of their data, while constraints and social factors in reality lead them to a seemingly contradictory behavior or a more realistic expectation. The phenom-enon of privacy paradox can translate from a consumer behavior to the context of learning analytics where students are expected to weigh their concerns over privacy against the expected benefits of the learning analytics system [20] when deciding whether and who to share their data with, and what actions to take to protect privacy. The discrepancies between expectations and behaviors can be further traced down to the cognition of expectations in different types.

2.3 Expectations as beliefs

Expectations are considered to be beliefs about the future [26] and they form the basis of judging whether we are satisfied with an experience [7]. In terms of technology adoption, research has pre-dominately focused on service-related user experiences [12], but the importance of gauging and managing user expectations are being recognised [6]. As shown in this literature, the failure to meet service user expectations within the pre-implementation stages of a new technology may result in limited adoption due to the dissat-isfaction that arises [6, 13]. Thus, a proactive approach should be undertaken with a view to understand what service users expect from any future implementation to improve the rate of adoption. In the context of learning analytics services, expectations are not constrained to the features offered [33], but also encompass the data practices of the university [20, 31]. These privacy expectations refer to how the university collects and analyses student data, specifi-cally encompassing student expectations towards the provision of consent and the security of the data itself. In line with the privacy paradox, the measurement of privacy expectations provides an in-sight into the beliefs that students hold towards the data handling procedures, which may be weighed up against the eventual benefits of the service.

For the purposes of exploring student expectations of learning analytics services, our work seeks to understand their ideal and predicted expectations [11, 41]. Ideal expectations provide an upper reference point to enable institutions to explore what service stu-dents would like, while predicted expectations provide a realistic benchmark to understand the minimum standards that students ex-pect based on their understanding of constraints in reality. Further, the comparison between the two types of expectations may accen-tuate areas to improve upon so as to ensure student engagement and satisfaction with a learning analytics service. Moreover, for the purpose of aligning an adoption strategy with the key stakeholder's interests, the measure of expectation is important as students are likely to have limited or no experience with the proposed services. This work examines conditions that will support the implemen-tation and use of learning analytics in a way that serves student interests and builds up students with the critical skills required for digital citizenship in the 21th century. In the following section, we explain the methods adopted for this investigation, including a survey that measured the ideal and predicted expectations of stu-dents concerning learning analytics services, and focus groups that explored student perceptions of information privacy and factors that might affect their cost-benefit assessments regarding sharing personal data.

3 METHODOLOGY

This work adopts a mixed-method approach to understand learn-ing analytics related privacy issues and human behavior through students' own perception and beliefs. A 12-item questionnaire (the SELAQ)[44] was used to measure expectations of learning analytics

services from a large sample of students at a UK university. In addi-tion, six focus groups were conducted to increase the richness of data by taking advantage of group dynamics that allow participants to inspire one another and probe ideas among themselves [23]. It was expected that the shared experiences at the university would increase participant willingness to talk about their personal views and experiences regarding privacy. Both research activities have been approved by an ethics committee in this UK university.

3.1 Participants

3.1.1 Questionnaire sample. A total of 674 responses to the ques-tionnaire were collected (Female = 429, 63.65%; 10.11% response rate) from a UK higher education institution between March and April 2017. Respondents were aged between 18 and 72, with a mean age of 24.50 (SD = 7.94). The majority of the sample were under-graduate students (n = 396, 58.80%), followed by PhD students (n = 216, 32%), and then master's students (n = 62, 9.20%). A total of 31.20% (n = 210) of students were studying a subject from the Arts and Humanities, 24% (n = 162) were taking a subject within the faculty of Science, 19.30% (n = 130) stated they were taking a Social Science subject, 15.30% (n = 103) of students were from the faculty of Medicine and Health Sciences, and 10.20% (n = 69) were from Engineering departments. The sample comprised 475 domestic students (70.50%) and 199 international students (29.50%).

3.1.2 Focus group sample. To enable in-depth discussions, six focus groups were conducted, each comprising four to five participants who were recruited from the same UK higher education institution where the questionnaire was distributed. Participants were selected widely from the institution to include a diversity of student bodies. Four undergraduate focus groups were formed to represent the three university colleges, labeled as UG1, UG2, UG3, and UG4 in this paper. UG1 and UG2 were drawn from the Arts, Humanities, and Social Sciences College, which had the largest student body when compared to the other colleges. UG3 was from the Science and Engineering College, and UG4 from the Medicine and Veteri-nary Medicine College. In addition, two focus groups were drawn from mixed disciplines to include postgraduate students and online-distance learning students (who were generally part-time students) respectively, the former labeled as PG, and the latter as ODL in this paper. In total, 26 students (7 males, 19 females) participated in the study in February 2017 and only one participant from the ODL group had prior experience with learning analytics.

3.2 Procedure

3.2.1 Questionnaire procedure. Expectations towards learning an-alytics services were measured using a validated questionnaire– The Student Expectations of Learning Analytics Questionnaire (SELAQ)[44]. Five of these 12 items refer to Ethical and Privacy Expectations, covering themes of data security, consent, and third-party data usage (Table 1). Responses to each of the survey items are made on two seven-point Likert scales (1 = Strongly Disagree, 7 = Strongly Agree) that reflect two levels of expectations: ideal (Ideally, I would like this to happen) and predicted (In reality, I expect this to happen). Put differently, an ideal expectation refers to what a student desires in terms of service provision or related privacy protection, whilst a predicted expectation is a belief of what a student realistically expects to happen when a service is implemented. By differentiating between these two levels of ex-pectation, researchers can obtain both upper and lower reference points with regards to knowing what students expect of learning analytics services. Before completing the questionnaire, students were presented with a written description of learning analytics. All the participants needed to give their consent before starting the survey, and an option to opt into a prize draw was offered as an incentive.

3.2.2 Focus group procedure. All participants received a short video explaining the concept of learning analytics before their participation in focus groups. The focus group interviews were semi-structured, each lasting for an hour. As the institution's adop-tion of learning analytics was in a rather early stage, the focus groups were intended to understand student awareness and atti-tudes regarding existing data practices, which the interviewer drew upon to guide participants to consider the potential uses of their data for learning analytics in an attempt to understand student expectations and concerns regarding such uses. To this end, ten questions were asked, each intended to explore a theme related to privacy or learning analytics services (http://bit.ly/fg_q_lak). All participants signed a consent form to participate in the study and agreed to have their conversations recorded. Each participant received ten pounds in gratitude for their time.

3.3 Data analysis

3.3.1 Questionnaire: descriptive statistics. Descriptive statistics were used to analyze the results of the questionnaire, using both response frequencies and percentages. For the focus of this paper, only the responses to the five Ethical and Privacy Expectations items are reported.

3.3.2 Focus groups: thematic analysis. The focus group interviews were transcribed and then analyzed using a thematic coding method [19]. The coding scheme (http://bit.ly/fg_code_lak) was developed inductively, which involved one researcher reading the transcripts repetitively to identify recurring themes and types of issues raised. The qualitative analysis tool–Nvivo–was employed to assist in this process, which resulted in 64 codes categorised into three main themes and fourteen sub-themes: 1) educational services (chal-lenges, communication, data types, goals & benefits, intervention, and stakeholders); 2) ethics and privacy (access, anonymity, consent, opting in or out, and transparency); and 3) perceptions (attitudes, awareness, and concerns). For the focus of this paper, we report the coding results of the last two themes: 'ethics and privacy' and 'perceptions'. In the following sections, the participants of the focus groups are denoted as S (student) with numbers (1 to 5) to differen-tiate individuals in the same group. Some of the participants were second language speakers of English. The selected excerpts are faithful to the original responses, with the minor exception that some redundant words, such as 'like', were edited out whenever these words were not considered to contribute significant meaning to the study.

Table 1: SELAQ–Ethics and Privacy Expectation Items

Ethics and Privacy Expectation Items	Abbreviations
The university will ask for my consent to collect, use, and analyse any of my educational data (e.g., grades, attendance, and virtual learning environment accesses).	Use Edu Data
The university will ask for my consent before my educational data is outsourced for analysis by third party companies.	Third Parties
The university will ensure that all my educational data will be kept securely.	Keep Secure
The university will ask for my consent before using any identifiable data about myself (e.g., ethnicity, age, and gender).	Identifiable Data
The university will request further consent if my educational data is being used for a purpose different to what was originally stated.	Alternative Purpose

4 RESULTS

In this section, we present students' attitudes towards supplying personal and educational data for the institution to enable educa-tional services. The findings are presented in two sections to answer the two research question respectively.

In Section 4.1, we compare findings of the survey to focus groups. We first highlight three emerging themes–purpose, anonymity, and access–that determine how students perceive the legitimacy of data practices. We then present student perceptions of autonomy over personal and educational data and the conflicts between retaining their data and receiving educational support. In Section 4.2, we contrast the abovementioned expectations with what students ac-tually do in reality to protect their data. Based on the results, we discuss the implication of these observations for the deployment of learning analytics in higher education (Section 5).

4.1 Legitimate uses of student data for learning analytics

4.1.1 Survey findings. The SELAQ [44] contains five items that explore student expectations towards the data practices of the university (Table 1). Figure 1 shows the response frequencies for the seven Likert scale categories across the Ethical and Privacy Expectations items; the figure generally shows these expectations to be both strong and positive across each scale.

In terms of purpose, the majority of students strongly agreed that the university should obtain consent when collected data is used for an alternative purpose. This is shown on both the ideal (n = 450, 66.77%) and predicted (n = 282, 41.84%) expectation scales. In terms of anonymity, the majority of students also strongly agreed that the university should seek student consent before using identifiable data. This is shown on both the ideal (n = 407, 60.39%) and predicted (n = 293, 43.47%) expectation scales. Similarly, ensuring that all data is kept secure was similarly met with the majority of students strongly agreeing on both the ideal (n = 489, 72.55%) and predicted (n = 347, 51.48%) expectation scales. When it comes to seeking consent before using educational data or outsourcing data to third parties, the latter received a strong positive response to the ideal expectation (strongly agree that they ideally wanted this to happen (n = 492, 73%) and strongly agree that this would happen in reality (n = 313, 46.44%)), whereas responses to the former declined to 51.48% on the ideal expectation scale (n = 374) and 31.16% on the predicted expectation scale (n = 210).

These descriptive statistics showed that the majority of students held strong expectations regarding the university's data handling practices. For all items, over 51.48% of students ideally wanted the described data practices to be undertaken at the university. In other words, they desired for the university to obtain consent before using data for alternative purpose, before utilizing any identifiable data, before outsourcing data to third party companies, and before collecting and analyzing educational data; in addition to ensuring that all data is kept secure. It is clear from the aforementioned points that the students held high expectations when it came to the security of their data and data sharing with external parties. By contrast, they seemed more relaxed about sharing educational data. Importantly, while the majority of the students in the sample expressed a strong desire for the university to abide by the data practices in an ideal situation, they generally expected the university to adhere to this in reality too, despite the slight decline in holding high expectations. A possible interpretation of the gap between ideal and predicted expectations of these items is that student experience with these practices or understanding of associated challenges led to lower expectations of what would occur in reality. The data obtained using the SELAQ [44] allow us to explore what students generally expected of the university with regards to their data practices. The next section uses the data obtained from the analysis of the focus groups to provide a deeper understanding of students' expectations and experiences of data practices at the university.

4.1.2 Focus group findings. In general, the participants welcomed the university collecting and using their data for three purposes: first, to comply with legal requirements, such as visas; second, to improve educational services, such as learning support, teaching delivery, career development, educational resources management, and the support of student well-being; and third, to improve the overall performance of the university, such as league rankings, equality, and the recruitment of future students.

There was a consensus among the participants that student data should be used to benefit students. They generally agreed that using student data to improve the learning environment was legitimate so long as the process does not identify individuals, particularly through the use of demographic or other sensitive data. Similarly,

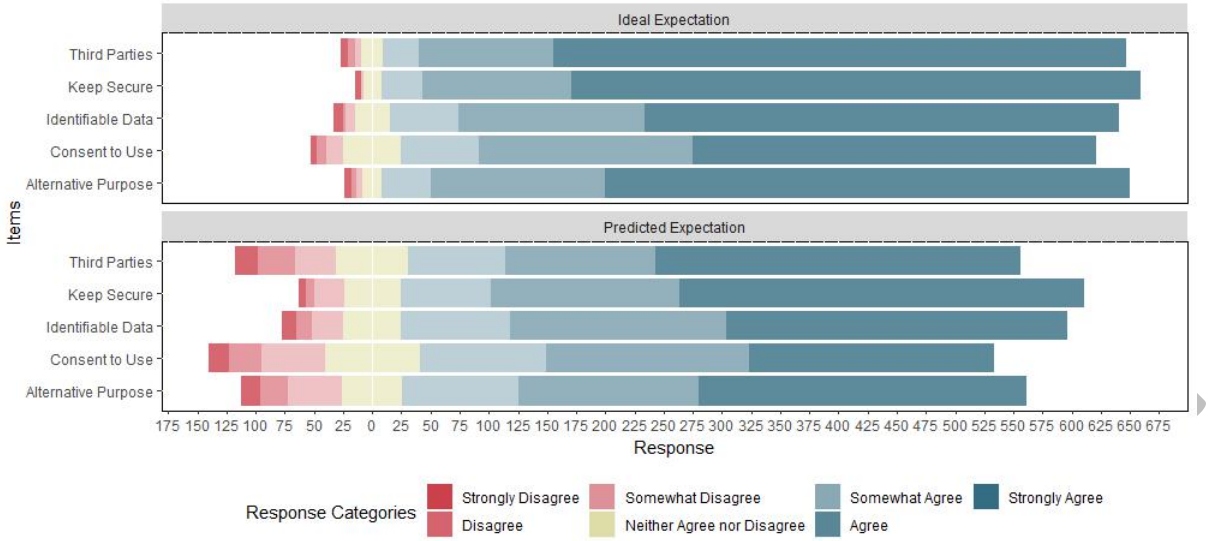

Figure 1: Diverging Stacked Bar Plot showing the Category Response Frequencies for the Ethical and Privacy Expectation Items

when it came to academic data, the students believed in the value of using such data to improve academic offerings, although they indicated that anonymity principles should be strictly applied in incidents including grading exams or assignments, sharing data with external parties, publishing exam results, and presenting peer comparison data. Some participants also emphasized that for any data that is collected compulsorily, anonymity principles should be applied to respect the data subjects.

In incidents of sharing data with personal tutors (academic staff who look after the overall learning experience and welfare of stu-dents in UK universities), students were happy to reveal personal and academic data in exchange for identified benefits. For example, a participant indicated the value of sharing such data, which led to a discussion around consent among the participants (PG):

> S4: I haven't been in University for so long, so for me to get back to the school was challenging, and so for me with the personal tutor I don't mind sharing my data 'cause she will help me to develop myself further. S2: I think though that makes sense to give consent 'cause what if you're paired with a personal tutor that you didn't...?
> S4: Yeah, I meant the personal tutor acts right so I mean it's in my....
> S2: It's your trust to give her.
> S4: ....Right, also I have to give them permission.
> S2: Oh okay, yeah. That makes sense to me.
> S5: I agree with you but I think the personal tutor they shouldn't compare the grades or share information to other students....
> S4: No.
> S5: ....that's good, it's just to be beneficial for my de-velopment or academic purpose. I think that's fine for me.

The conversation above gave a glimpse into the students' risk-benefit assessments towards revealing personal data to tutors. While personal learning gains were perceived as benefits, they came with risks such as data traveling beyond students' control or used for a purpose not agreed to by students. Specifically, the students above highlighted their autonomy in giving consent based on the pres-ence of trust in the service provider (i.e., the personal tutor), which revealed their conceptualization of privacy based on ownership [9] and contextual integrity–the nature of data (i.e., personal), roles in a context (i.e., students and personal tutors), and relationships among the roles (i.e., mentorships and peer relationships) together determine the appropriateness of sharing data [24]. Moreover, it is clear from the above-conversation that the decision to share data was based on the presence of trust [25, 39].

Aligned with the findings of the survey, most students in the focus groups expressed extremely uncomfortable feelings towards sharing data with external parties. The noticeable distrust in exter-nal parties among participants were explained by concerns such as having little control over data that have travelled out of the university and the likelihood of becoming the targets of commer-cial advertisements. One participant pointed out that transparency would be key to her judgement on whether or not to share her data externally:

> I think once you start allowing some, it's quite hard to sort of regulate it. So without knowing more about what the data is being used for, I'd probably say no third parties at this point (S3, UG4).

The same group of participants then went on to complain about their experience of receiving advertisements from medical compa-nies, for example in the sale of tampons. Despite the lack of trust in external parties across the groups, most students agreed that sharing data externally for the purpose of facilitating educational services or employment would be a legitimate use. Nevertheless,

students' existing experience with spam e-mails has led to resis-tance in the first instance, unless the purpose and uses of data have been clearly conveyed. Overall, the participants believed that the use of data should not exceed the range of purposes that students previously consented to or ways that were not explicitly stated in the agreement.

Although students in general were happy to share their educa-tional data, as also shown in the survey results, the focus group participants' views towards the options to opt into or out of data collection were divided into 'overall freedom' and 'selective free-dom'. Those who held the view of 'overall freedom' believed that individuals are entitled to choose what they want to do with their own data. This attitude concurs with the self-determination princi-ple [9], which highlights data subjects' ownership over their data and decisions with it. For example, one participant (S3, UG4) said, "It's things about you. That's personal. People shouldn't just be able to do things with information about you without asking. It's just not okay." This participant's view illustrates a sense of exclusive ownership over one's own data. On the other hand, students who held the view of 'selective freedom' believed that students should only own partial freedom over their data so as to work together with the university to improve educational services. Specifically, they differentiated academic data from personal and sensitive data– the collection of the former should be compulsory while the latter should not. For example, one participant (S5, UG3) said:

> For examinations, I don't think you should opt out. I mean maybe you can remain anonymous within the pool, but the university needs to know how many people are passing their course, because for example, say there's more than 50% of the people that fail a course, it might not just be the students, but it may be some problems with the course. So I think that the university should have access at least anonymously to principal data like this.

Two other participants (S1 & S2) in the group held the same view and suggested that students should recognize that 'being at the university' means 'agreeing to share their academic data'. These students expressed a sense of inclusive ownership over their aca-demic data and acknowledged a partnership relationship with their university in improving the quality of educational services. To them, the benefits of sharing academic data outweighed the cost of giving up partial autonomy. This is also reflected in the survey results where expectations for the university to seek consent before utiliz-ing educational data were comparatively lower than the other data practices.

As the survey results showed, students in general expressed higher expectations towards what they would ideally like to happen than what they would expect to see in reality (Figure 1), the focus group participants also recognised the gap between an ideal state of having an overall control over their data and the challenge of it to happen in reality (PG):

> S3: I mean ideally in my mind, I think there should be the option to attend a university without giving every bit of information, maybe saying 'yes' for some, 'no' for others. But I guess that's a very difficult thing to do, how do you divide that up?
>
> S2: Yeah, and I think if there's an opt-out, I think a lot of people end up opting out because they just don't understand what it's used for....
>
> S3: That's true.
>
> ....
>
> S2: I wonder if you'd get a biased sample of people who did opt-out based on the fact that they were hav-ing difficulties, 'cause you're more likely to opt out if you're unhappy than if you are happy, just like pretty much most research. But I think that what if those are the people that the University needs to be focusing on the most and they're missing the opportunity to help those students.

The conversation above reveals conflicts between retaining pri-vacy and losing the opportunity to receive a specific service. As the participants implied, overall-freedom is ideal, but not practical under the assumption that the university has the responsibility to support every student to succeed in their studies. The realization of the gap between 'what I want' and 'what I expect to get' (or 'what I am willing to accept') is where the students assessed the benefits and costs of sharing personal data, which is likely to lead to the discrepancy between one's belief in privacy and decisions about it [2, 22].

4.2 Gaps between expectations and actions

As mentioned earlier, existing experiences can affect student expec-tations of the rightful uses of data and hence their willingness to share personal data (e.g., sharing data with third parties). However, it is also notable from the analysis of focus groups that the aware-ness of data protection is not equal to the awareness of existing data practices, nor does it necessarily correlate positively with the actions that students would take to protect their data. The gaps are observed in the confessions among students regarding their experience in understanding data process and giving consent.

4.2.1 Data collection. Respondents in the focus group expressed uncertainty when it came to knowing who may have access to data collected by the university, and how such data could be used to improve academic offerings and support. The students tended to assume that this kind of information was accessible and could be located, if they were motivated to look for it. For example, a student confessed that she had no interest in understanding how the university used her data:

> That's something I don't think I would ever focus on or look for, so I honestly don't know. It could be out there and I could maybe Google it, and it would be on a page somewhere if I wanted to find it. I don't really care if they use my data. I think it's probably beneficial if they do use everybody's data (S2, PG).

The aforementioned responses are indicative of two problems with the university's communication of data policies. First, the policy information was either not made explicit to students or not commu-nicated to students effectively. Second, students in general lacked interest in engaging with such information. By contrast, the survey responses discussed earlier showed that students generally have

high ideal and realistic expectations of the university ensuring all data is kept secure and obtaining consent for various data han-dling steps. This shows that both ideal and realistic expectations of protection over one's own data can sometimes appear as a con-tradiction with the intention to take action towards it. While the phenomenon of information asymmetry (one party has more or better information than the other) was already observed among the students due to the ineffective communication of data policies, the trust that students placed in the institution seemed to have contributed further to their passive engagement with information about the processing of their data. Again, this affirms the role of 'trust' in the behavior to disclose data despite the perceived risks or intention to protect one's data [25, 39].

4.2.2 Consent experience. The passive engagement with informa-tion about existing data practices was also observed among the students when it came to their experience in providing consent to the university. Although most students remembered that they explicitly gave consent at enrolment to let the university use their data, some students could only 'assume' that they did. One partici-pant (S1, UG4) recalled that she might have seen "small print things" and ticked a box, while another participant (S3, UG4) said, "I don't think I paid too much attention to it at the time". It seems that the priority to complete enrolment made students less interested in engaging with such information.

When students were asked about what they consented to at en-rolment, none of them could recall any details. Only one student (S2, UG3) mentioned that he read the whole data policy document, while another participant (S1, UG2) said, "Unless some issues arise we might check". This remark was mirrored by two other partici-pants (S3 &S4) in the group (UG2). These students showed that they took a cure approach rather than a prevention approach towards privacy intrusion. Again, this shows a trust relationship between the students and their institution, and the exacerbation of infor-mation asymmetry as a result. Other explanations for low consent awareness included policy length and difficulty to recall such events. The former refers to what is known as rational ignorance whereby users consider the effort and loss of time in reading a lengthy and complex policy to outweigh the perceived risk of disclosing per-sonal information [1]. Respondents in four of the six focus groups (G1, G2, G4, and PG) pointed out a preference that communications from the university (e.g., data policies) be presented in a succinct format, in contrast to "overwhelming people with an onslaught of information"(S4, UG1).

Beyond the issues of information recall and complexity, the perceived power imbalance between students and the university has drawn the attention to the authenticity of 'voluntary' and 'informed' consent. As exemplified in the words of two students:

> You have to agree to share this data otherwise you wouldn't enroll, so you are not probably thinking that much about consequences of every single piece of data that you provide to the university. It's just because it's a part of the process of application (S4, UG3).

> If I can't sign up for University without giving in-formation, I am being pressured quite a lot to hand over that information, 'cause if I don't then I can't get further education (S3, PG).

In this instance, not being able to recall the purposes of data collec-tion is attributed to the compulsory nature of providing consent in order to complete the enrolment process. What these students have described is a 'state of resignation'–the recognition of possessing little power to negotiate (i.e., agreeing to supply data or losing the educational opportunity)[2]. It can be expected, on this basis, that students are unlikely to have extensively read the details con-tained within the data policies; therefore, resulting in an inability to describe what they have consented for.

5 DISCUSSION

In this section, we reflect on the observation of a privacy paradox phenomenon and the implications thereof for the deployment of learning analytics in higher education.

5.1 A privacy paradox

The study shows that there were clear contradictions between expectations and experiences among the students when it came to sharing their data to be used at the university. The majority of the students held high expectations of the university obtaining consent to collect and analyse data, outsourcing data to third parties, and using identifiable data. In particular, ensuring that all data is kept secure and obtaining consent before data is outsourced to third parties received the highest expectations on both ideal and expected scales in the survey. While the responses from focus groups aligned with the expectations observed from those to the survey, focus groups further revealed a phenomenon of passive engagement with data policies, which can be explained in a number of ways.

First, the presentation of data policies was too lengthy to achieve effective communication. Second, the priority of completing en-rolment reduced the participants' motivation to engage with data policies before giving consent. Third, the participants' trust in the institution biased their perception of risks [2], leading them to take a 'cure' approach towards data protection. Finally, power imbalance led to a state of resignation [2] where the students accepted that consenting to supply personal data was a prerequisite to receiving higher education. Thus, while students appeared to hold protective attitudes towards their data, their described actions did not reflect such beliefs.

We argue that students' passive engagement with information about the use of their personal data exacerbates the phenomenon of information asymmetry, resulting in the lack of knowledge to critically judge the appropriateness of existing data usage or give informed consent. This can be problematic when data protection policies, such as General Data Protection Regulation [43], place growing emphasis on the responsibility of data subjects to make decisions for the use of their data. The paradox of being protective over one's data yet indifferent in taking appropriate actions to ensure its security highlights the need for institutions to examine the extent to which students are free to or encouraged to explore the implications of data disclosure. This is particularly important in the context of learning analytics where both data collection and interventions are focused on individuals. It is also crucial to the success of learning analytics, as students need to own their decisions if they are expected to respond to data and interventions as part of the feedback loop [8].

When it came to discussions surrounding data control, the fo-cus group findings are indicative of students either wanting com-plete control over their data or being open to sharing specific data, but only under certain conditions (e.g., data being insensitive and anonymized, and trust relationships being in place). From the sur-vey responses, we observed that the majority of students desired to have control over their data. With regards to whether students expected to have such control in reality, it appeared that students were more pessimistic in their view. An illustration of this is for the expectation that the university will obtain consent to collect and analyze educational data, to which 4.15% of students expressed some form of disagreement on the ideal expectation scale. On the predicted expectation scale, the percentage of disagreement rose to 14.80%. Similarly, 2.37% of students expressed disagreement that the university would obtain consent when using collected data for alternative purposes on the ideal expectation scale, compared to 12.90% on the predicted expectation scale. Obtaining consent to use identifiable data, however, received disagreement from 2.67% and 7.86% of respondents on the ideal and predicted expectation scale, respectively. In other words, responses to the ideal expectation scale are indicative of students desiring to have greater control over their data; whereas, predicted expectation scale responses seemingly show students to be less expectant of this occurring in reality. In the latter case, it may be that students do not believe the university to be capable of allowing students to have such control over their data or that students recognize the challenge or trade-off of doing so in reality. On the other hand, this is precisely where institutions should aim to dialogue with students to address the discrepancies between student expectation and experience.

5.2 Implications for learning analytics

In general, the participants expected themselves to be the main ben-eficiaries when their data (both personal and academic) is collected, used, and shared. In instances where data is expected to be shared, the participants expressed the highest trust in personal tutors and least in external parties due to the fear of losing control of their personal data and becoming marketing targets. Nissenbaum's [24] contextual integrity is useful to understand the way participants benchmark legitimate data usage: first, norms of appropriateness ensure the right purposes of using student data; second, norms of flow/ distribution draw a boundary of access to student data. However, apart from contextual integrity, anonymity is considered key to risk management, and thus part of the parameters used by the participants to judge legitimacy of data usage is anonymity. Therefore, when it comes to the formation of a strategy or policy for learning analytics, an ethics and privacy integrity framework comprising three basic elements–purpose, access, and anonymity– will be essential to ensure that students feel comfortable when sharing their data for learning analytics.

In addition, transparency and effective communication are key levers to scale up student willingness to use learning analytics to enhance learning. On the one hand, institutions need to make the benefits of learning analytics visible to students so that the 'gains' of sharing personal data are clear and of relevance to individuals. On the other hand, institutions need to be particularly transparent about the purposes of data collection, boundaries of access, and principles of anonymity, as mentioned previously. It is not good enough for students to 'assume' that their data is safe in the hands of the institution or that the responsibility to safeguard the use of student data is solely on the institution [10]. Instead, institutions should be cautious about the exacerbation of information asym-metry due to a trust relationship [20], resulting in compromising students' rational calculations of the risks and benefits of supplying personal data [2] for learning analytics. As making data-informed decisions becomes a mainstream in the data society [16, 35], it is critical that learning analytics is not simply used in a reactive manner to prompt actions, but to train students to be discerning when giving away their data, and to develop critical awareness to 'know through data'[21]. This means that communications of policies related to learning analytics should not be operated under the assumption that transparency equals to understanding [42]. Institutions should not only examine students' understanding of consenting to use their data for learning analytics, but also give students opportunities to review their decisions at different times as their experience at the university increases, and encourage them to clarify questions about the processing of their data.

6 CONCLUSION

Enhancing education with data technologies is becoming a global phenomenon [5, 16], which highlights the importance of installing digital literacy among students to enable active and informed par-ticipation in education [29]. Crucial to digital literacy is the critical ability to make informed decisions as to sharing personal data to be processed with data technologies, such as learning analytics. Our study showed that while students held protective attitudes towards personal data and high expectations of how the university should process their data, their actions to protect personal data did not reflect such awareness. Various factors were identified that could impact the risk assessment of offering data to be used for learning analytics, including perceived benefits and risks, a power relation-ship, trust relationship, and information asymmetry. In light of this, we urge decision makers in higher education to ensure that the development of learning analytics strategy and policy priori-tise the clarification of the purpose to use learning analytics, the parties that have access to data, and the procedure of anonymi-saiton. Moreover, for an institution to be truly transparent with their data practices, an effective communication of data polices and open dialogue with students are crucial to clarify any 'mist' and to address assumptions that could potentially cloud a decision to engage with learning analytics. That is to say, despite the fact that students in general demonstrate trust in their institution, as also observed in a recent study [39], institutions need to address the problem of information asymmetry in order for students to take up an active role and ownership over the use of learning analytics. Some practical steps might be holding workshops or meetings with students, and embedding relevant training on digital literacy into academic development programmes so as to ensure that students are not only aware of the importance of data protection, but know what actions to take to protect their data or disclose data for learn-ing analytics with informed consent. To conclude, we highlight the following key implications for learning analytics research and practice: (1) purpose, access, and anonymity are key benchmarks of

ethics and privacy integrity; (2) transparency and communication are key levers for learning analytics adoption; and (3) information asymmetry can impede active participation of students in learning analytics.

7 LIMITATIONS AND FUTURE STUDIES

While the observations presented in this paper were obtained from a study based in a UK higher education institution, the paradox is likely to exist in other educational and national contexts. This study did not intend to investigate or prove the difference between student attitudes towards privacy and their actual behaviors in choosing to supply or retain personal data. Instead, the study aimed to gain insights from students' perceptions of privacy, in order to understand how these might affect their willingness to engage with learning analytics, and therefore make suggestions for institutional strategies and policies. Therefore, future studies could scale the investigation and compare it with perceptions and expectations of students who have had significant experience with learning analytics.

REFERENCES

[1] Alessandro Acquisti and Jens Grossklags. 2007. What can behavioral economics teach us about privacy. Digital privacy: theory, technologies and practices 18 (2007), 363–377.

[2] Susanne Barth and Menno DT De Jong. 2017. The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. Telematics and Informatics 34, 7 (2017), 1038–1058.

[3] Scott Beattie, Carolyn Woodley, and Kay Souter. 2014. Creepy analytics and learner data rights. Rhetoric and reality: Critical perspectives on educational technology. Proceedings ascilite (2014), 421–425.

[4] France Bélanger and Robert E Crossler. 2011. Privacy in the digital age: a review of information privacy research in information systems. MIS quarterly 35, 4 (2011), 1017–1042.

[5] Malcolm Brown, Joanne Dehoney, and Nancy Millichap. 2015. The next genera-tion digital learning environment. A Report on Research. ELI Paper. Louisville, CO: Educause April (2015).

[6] Susan A Brown, Viswanath Venkatesh, and Sandeep Goyal. 2014. Expectation confirmation in information systems research: A test of six competing models. Mis Quarterly 38, 3 (2014).

[7] Wendy Christiaens, Mieke Verhaeghe, and Piet Bracke. 2008. Childbirth expecta-tions and experiences in Belgian and Dutch models of maternity care. Journal of reproductive and Infant Psychology 26, 4 (2008), 309–322.

[8] Doug Clow. 2012. The Learning Analytics Cycle: Closing the Loop Effectively. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12). ACM, 134–138. https://doi.org/10.1145/2330601.2330636

[9] Sami Coll. 2014. Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. Information, Communication & Society 17, 10 (2014), 1250–1263.

[10] Andrew Cormack. 2016. Downstream consent: A better legal framework for Big Data. Journal of Information Rights, Policy and Practice 1, 1 (2016).

[11] Daniel David, Guy H Montgomery, Rosana Stan, Terry DiLorenzo, and Joel Erblich. 2004. Discrimination between hopes and expectancies for nonvolitional outcomes: psychological phenomenon or artifact? Personality and individual differences 36, 8 (2004), 1945–1952.

[12] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user accep-tance of information technology. MIS quarterly (1989), 319–340.

[13] Fred D Davis and Viswanath Venkatesh. 2004. Toward preprototype user ac-ceptance testing of new information systems: implications for software project management. IEEE Transactions on Engineering management 51, 1 (2004), 31–46.

[14] Shane Dawson, Dragan Gašević, George Siemens, and Srecko Joksimovic. 2014. Current state and future trends: A citation network analysis of the learning analytics field. In Proceedings of the fourth international conference on learning analytics and knowledge. ACM, 231–240.

[15] Hendrik Drachsler and Wolfgang Greller. 2016. Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. In Proceedings of the sixth international conference on learning analytics & knowledge. ACM, 89–98.

[16] EDUCAUSE. 2018. NMC Horizon Report Preview: 2018 Higher Education Edition.

[17] Rebecca Ferguson. 2012. Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning 4, 5/6 (2012), 304–317.

[18] Dragan Gašević, Shane Dawson, and George Siemens. 2015. Let's not forget: Learning analytics are about learning. TechTrends 59, 1 (2015), 64–71.

[19] Carol Grbich. 2012. Qualitative data analysis: An introduction. Sage.

[20] Dirk Ifenthaler and Clara Schumacher. 2016. Student perceptions of privacy prin-ciples for learning analytics. Educational Technology Research and Development 64, 5 (2016), 923–938.

[21] Jeremy Knox. 2017. Data power in education: Exploring critical awareness with the "Learning Analytics Report Card". Television & New Media 18, 8 (2017), 734–752.

[22] Spyros Kokolakis. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & security 64 (2017), 122–134.

[23] Pranee Liamputtong. 2011. Focus group methodology: Principle and practice. Sage Publications.

[24] Helen Nissenbaum. 2004. Privacy as contextual integrity. Wash. L. Rev. 79 (2004), 119.

[25] Patricia A Norberg, Daniel R Horne, and David A Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of consumer affairs 41, 1 (2007), 100–126.

[26] Jerry C Olson and Philip Dover. 1976. Effects of expectation creation and dis-confirmation on belief elements of cognitive structure. ACR North American Advances (1976).

[27] Abelardo Pardo, Jelena Jovanovic, Shane Dawson, Dragan Gašević, and Negin Mirriahi. 2019. Using learning analytics to scale the provision of personalised feedback. British Journal of Educational Technology 50, 1 (2019), 128–138.

[28] Abelardo Pardo and George Siemens. 2014. Ethical and privacy principles for learning analytics. British Journal of Educational Technology 45, 3 (2014), 438–450.

[29] Carlo Perrotta. 2013. Assessment, technology and democratic education in the age of data. Learning, media and technology 38, 1 (2013), 116–122.

[30] Paul Prinsloo and Sharon Slade. 2016. Student vulnerability, agency, and learning analytics: An exploration. Journal of Learning Analytics 3, 1 (2016), 159–182.

[31] Lynne D Roberts, Joel A Howell, Kristen Seaman, and David C Gibson. 2016. Student attitudes toward learning analytics in higher education:"The fitbit version of the learning world". Frontiers in psychology 7 (2016), 1959.

[32] Alan Rubel and Kyle ML Jones. 2016. Student privacy in learning analytics: An information ethics perspective. The Information Society 32, 2 (2016), 143–159.

[33] Clara Schumacher and Dirk Ifenthaler. 2018. Features students really expect from learning analytics. Computers in Human Behavior 78 (2018), 397–407.

[34] Niall Sclater. 2016. Developing a Code of Practice for Learning Analytics. Journal of Learning Analytics 3, 1 (2016), 16–42.

[35] George Siemens. 2013. Learning analytics: The emergence of a discipline. Ameri-can Behavioral Scientist 57, 10 (2013), 1380–1400.

[36] George Siemens and Ryan SJ d Baker. 2012. Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, 252–254.

[37] Sharon Slade and Paul Prinsloo. 2013. Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist 57, 10 (2013), 1510–1529.

[38] Sharon Slade and Paul Prinsloo. 2015. Student perspectives on the use of their data: between intrusion, surveillance and care. European Journal of Open, Distance and E-learning 18, 1 (2015).

[39] Sharon Slade, Paul Prinsloo, and Mohammad Khalil. 2019. Learning analytics at the intersections of student trust, disclosure and benefit. InProceedings of the 9th International Conference on Learning Analytics & Knowledge. ACM, 235–244.

[40] Angus Stevenson. 2015. Oxford English Dictionary (3rd ed.). 10.1093/acref/9780199571123.001.0001

[41] Andrew GH Thompson and Rosa Sunol. 1995. Expectations as determinants of patient satisfaction: concepts, theory and evidence. International journal for quality in health care 7, 2 (1995), 127–141.

[42] Yi-Shan Tsai, Carlo Perrotta, and Dragan Gašević. In press. Empowering Learners with Personalised Learning Approaches? Agency, Equity and Transparency in the Context of Learning Analytics. Assessment and Evaluation in Higher Education (In press).

[43] The European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council. Official Journal of the European Union59 (2016). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[44] Alexander Whitelock-Wainwright, Dragan Gašević, Ricardo Tejeiro, Yi-Shan Tsai, and Kate Bennett. 2019. The Student Expectations of Learning Analytics Questionnaire. Journal of Computer Assisted Learning (2019).

[45] Alexander Whitelock-Wainwright, Dragan Gašević, Yi-Shan Tsai, Hendrik Drach-sler, Maren Scheffel, Pedro Muñoz-Merino, Kairit Tammets, and Carlos Del-gado Kloos. In press. Assessing the validity of a learning analytics expectation instrument: A multinational study. Journal of Computer Assisted Learning (In press).

[46] James E Willis, Sharon Slade, and Paul Prinsloo. 2016. Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. Educational Technology Research and Development 64, 5 (2016), 881–901.

[47] Shoshana Zuboff. 2015. Big other: surveillance capitalism and the prospects of an information civilization. Journal of Information Technology 30, 1 (2015), 75–89.