

Student perceptions of privacy principles for learning analytics

Dirk Ifenthaler^{1,2} · Clara Schumacher¹

Published online: 17 August 2016

© Association for Educational Communications and Technology 2016

Abstract The purpose of this study was to examine student perceptions of privacy principles related to learning analytics. Privacy issues for learning analytics include how personal data are collected and stored as well as how they are analyzed and presented to different stakeholders. A total of 330 university students participated in an exploratory study confronting them with learning analytics systems and associated issues of control over data and sharing of information. Findings indicate that students expect learning analytics systems to include elaborate adaptive and personalized dashboards. Further, students are rather conservative in sharing data for learning analytics systems. On the basis of the relationship between the acceptance and use of learning analytics systems and privacy principles, we conclude that all stakeholders need to be equally involved when learning analytics systems are implemented at higher education institutions. Further empirical research is needed to elucidate the conditions under which students are willing to share relevant data for learning analytics systems.

Keywords Learning analytics · Privacy · Control over data · Transparency · Higher education

Introduction

Masses of administrative, system, academic, and personal data on educational settings and higher education institutions are becoming more and more available. These vast amounts of educational information provide new opportunities to improve administrative decision-making as well as to facilitate learning and instruction. Learning analytics uses this static

✉ Dirk Ifenthaler
dirk@ifenthaler.info

¹ University of Mannheim, Mannheim, Baden-Württemberg, Germany

² Deakin University, Melbourne, VIC, Australia

and dynamic information about learners and learning environments—assessing, eliciting and analyzing them—for real-time modeling, prediction, and optimization of learning processes, learning environments, and educational decision-making (Ifenthaler 2015b).

The adoption of learning analytics in the higher education sector is evolving fast. However, ethical and privacy issues have been identified as a major concern (Slade and Prinsloo 2013). Ethics refers to a system of fundamental principles and universal values of right conduct. A critical reception of ethics in educational technology is available (Moore and Ellsworth 2014; Yeaman et al. 1994).

Learning analytics should be aligned with organizational principles and values, include all stakeholders, collect, use, and analyze data transparently and free of bias, and be beneficial for all involved stakeholders. Privacy is closely linked with these ethical considerations. The most general definition of privacy is freedom from interference or intrusion (Warren and Brandeis 1890). A legal definition of the concept of privacy is a person's right to control access to his or her personal information (Gonzalez 2015). More precisely, privacy is a combination of control and limitations, which implies that individuals can influence the flow of their personal information and prevent others from accessing it (Heath 2014).

Currently, most research on privacy issues in learning analytics refers to guidelines from other disciplines, such as Internet security or medical environments (Pardo and Siemens 2014). However, due to the contextual characteristics of privacy in learning analytics, it is not recommendable to adopt guidelines from other contexts (Nissenbaum 2004). This exploratory study addresses this research gap by investigating students' perceptions of privacy principles related to learning analytics.

Privacy principles for learning analytics

Learning analytics

Learning analytics emphasizes insights and responses to real-time learning processes on the basis of educational information from digital learning environments, administrative systems, and social platforms. Vast amounts of static and dynamic educational information are used for real-time modeling, prediction, and optimization of learning processes, learning environments, and educational decision-making (Ifenthaler 2015b). More precisely, learning analytics frameworks use information about (1) learners' individual characteristics (e.g., prior knowledge, academic performance), (2) activities in the learning environment (e.g., user pathways, download activity), (3) curricular benchmarks (e.g., learning outcomes, historical course information), and (4) interactions with peers and teachers (e.g., social network activity) (Greller and Drachsler 2012; Ifenthaler 2015b).

Serious concerns and challenges are associated with the application of learning analytics (Pardo and Siemens 2014). For example, not all educational information are relevant and equivalent. Therefore, the validity of data and its analysis are critical for generating useful summative, real-time, and predictive insights (Macfadyen and Dawson 2012). Furthermore, limited access to educational data may generate disadvantages for stakeholders. For example, invalid forecasts may lead to inefficient decisions and unforeseen problems (Ifenthaler and Widanapathirana 2014). Another critical concern is related to the use of educational information for learning analytics and involves how personal data are collected and stored as well as how they are analyzed and presented to different stakeholders (Slade

and Prinsloo 2013). However, there is little empirical evidence regarding privacy issues, such as perceptions of students and teachers towards sharing and ownership of data in learning analytics.

Privacy in the digital world

Within the digital world, many individuals are willing to share personal information without being aware of who has access to the data, how and in what context the data will be used, or how to control ownership of the data (Solove 2004). Accordingly, data are generated and provided automatically by online systems, which limits the control and ownership of personal information in the digital world (Slade and Prinsloo 2013).

There are several reasons why learners would like to keep their information private:

First, there are competitive reasons, for example, if a learner performs poorly, a fellow student shall not know about it. Second, there are personal reasons, for example a learner might not want to share information about him-/herself (Aïmeur et al. 2008). There are also country-specific differences who owns the personal data. In the United States the collected data belongs to the collectors. In Europe the personal data belongs to the individual (e.g., the learner) (Weippl and Min Tjoa 2005).

Table 1 provides an overview of privacy theories in the digital age (Heath 2014). The first two concepts (1, 2) emphasize requirements for reaching privacy in a certain situation and focus on protection and normative or descriptive privacy. Early privacy theories (3) are based on control or limitation: Control refers to the influence of individuals on the flow of their personal data, whereas limitation means the possibility to prevent others from accessing personal data (Elgesem 1999). Contemporary privacy theories (4) incorporate

Table 1 Overview of privacy concepts (adopted from Heath, 2014)

(1) Individual has privacy in a particular situation if they are offered three protections	Protection from interference	Protection from information access by others	Protection from intrusion
Comparable with the concept of Warren and Brandeis (1890), which points out the rights of an individual to be left alone and free from intrusion and interference			
(2) Two broad classifications of privacy situations	Normative privacy Zones of privacy	Descriptive privacy	
	Individuals are protected by cultural norms: formal laws and informal policies	Privacy can be expected by natural means, such as physical barriers	
(3) Early theories of privacy	Control theory	Limitation theory	
	Allowing individuals control over their personal information	Limitations on the persons who could gain access to personal information	
(4) More recently proposed information privacy theories aimed at achieving necessary protections by building on earlier theories and normative and descriptive privacy	Floridi ontological theory of information privacy	Nissenbaum contextual integrity theory of informational privacy	Moor and Tavani hybrid RALC theory of privacy

Contemporary privacy theories are more holistic and go beyond the early theories of privacy; they were developed to apply them to diverse contexts

these earlier theories as well as normative and descriptive privacy concepts but go beyond them in being more holistic and applicable to different contexts.

Privacy in learning analytics

Higher education institutions have always used a variety of data about students, such as socio-demographic information, grades on higher education entrance qualifications, or pass and fail rates, to inform their academic decision-making as well as for resource allocation (Long and Siemens 2011). Such data can help to successfully predict dropout rates of first-year students and to enable the implementation of strategies for supporting learning and instruction as well as retaining students (Tinto 2005).

Advanced digital technologies and learning analytics systems enable higher education institutions to collect dynamic real-time data from all student activity, offering huge potential for personalized and adaptive learning experiences and support (Berland et al. 2014). Consequently, higher education institutions need to address privacy issues linked to learning analytics: They need to define who has access to which data, where and how long the data will be stored, and which procedures and algorithms to implement for further use of the available data.

As mentioned above, learning analytics systems merge data from various contexts, for example from student administration, learning environments, or social interaction (Heath 2014; Ifenthaler 2015a). Privacy perceptions of students may vary between these contexts. Students might not want to disclose data they disclosed in the administrative context, such as their postal code or previous academic performance, in the learning environment context or vice versa. This phenomenon tallies with assumptions of the contextual integrity theory, according to which information disclosed in a specific context is not transferable to a different context without associating it with its original context to prevent a shift of meaning and intrusion in privacy (Nissenbaum 2011).

Slade and Prinsloo (2013) as well as (Pardo and Siemens 2014) have established several conceptual principles for privacy in learning analytics. They highlight the active role of students in their learning process, the temporary character of data, the incompleteness of data on which learning analytics are executed, transparency regarding data use, and the purpose, analysis, access, control, and ownership of data.

Both emphasize the importance of transparency regarding privacy issues. Transparent information includes: (1) who has access to data, (2) how data are collected, (3) which analyses are conducted, (4) with what purpose (e.g., assessment or support for the individual learning process, etc.), (5) how long data and outcomes will be stored, (6) which form of de-identification or aggregation are applied. More controversial is the students' control over data, their right of access to it, and the possibility of adjusting collected data (Pardo and Siemens 2014).

Slade and Prinsloo (2013) further argue that learning analytics should focus on what is morally necessary and appropriate to support learning, while Pardo and Siemens (2014) emphasize that the learning analytics system itself and its analyses need to be reliable. In considering students as agents who collaborate on the learning analytics system to benefit their own learning, Slade and Prinsloo (2013) highlight the student-centered view of learning analytics. They furthermore mention the temporary character of data, as behavior varies over time. Additionally, learning success is a complex and multidimensional phenomenon which cannot be grasped in its entirety by learning analytics systems. Rather, they can only consider learning which occurs within the learning environment or through reactive data collected from sources outside of the learning analytics system. Nevertheless,

in order to gain deeper insight into learning processes and their optimization, higher education institutions cannot afford to not use educational information and neglect the potential of learning analytics (Slade and Prinsloo 2013).

Research questions and hypotheses

Empirical research is currently addressing the usability, effectiveness, and validity of learning analytics systems (Ali et al. 2012; Gašević et al. 2015; Ifenthaler and Widanapathirana 2014). However, privacy issues in learning analytics are in an early stage of research (Pardo and Siemens 2014).

Given the limited availability of institution-wide learning analytics systems, the first research question aimed to determine the degree to which students themselves report that they want to use a particular learning analytics system to support their learning. Learning analytics have potential to facilitated learning processes in many ways and provide personalized and adaptive scaffolds to learners when they need it (Berland et al. 2014; Ifenthaler and Widanapathirana 2014). We assumed that students would prefer a learning analytics system offering a broad variety of support for learning (Hypothesis 1). The second research question focused on empirically documenting the degree to which students would be willing to share data on learning analytics systems. Given the lack of empirical evidence of privacy and data sharing in educational contexts (Prinsloo and Slade 2015), we assumed that students would be conservative (i.e., unwilling to disclose sensitive personal information) about sharing data in learning analytics systems (Hypothesis 2). The third research question aimed to determine whether the extent to which students share data is related to their preferences for using learning analytics systems. Acceptance, understanding, and expected benefits of such systems are drivers toward the use of learning analytics systems (Heath 2014). We assumed that greater levels of sharing of data (Hypothesis 3a) and control over data (Hypothesis 3b) would be associated with the intended preference for using a learning analytics system.

Method

Participants and design

The study was designed as one-group quasi-experimental design online laboratory study implemented on the university's server and conducted in June 2015. Participants received one credit hour for participating in the study.

The initial dataset consisted of 333 responses. After the removal of incomplete responses, the final dataset included $N = 330$ valid responses (223 female, 107 male). The average age of the participants was 22.75 years ($SD = 3.77$). The majority of the participants were enrolled in a German university's bachelor's program (80 %), with 20 % of the participants being enrolled in a Master's program at the same university. The students' average course load in the current semester was five courses ($SD = 1.70$). Participants reported that 33 % of their Internet use was for learning, 33 % was for social networking, 26 % for entertainment, and 8 % for work.

Learning analytics systems

Three different examples of learning analytics systems were presented to the participants. None of the presented examples are commercial products. Rather, the three learning analytics systems have been developed and implemented by large international universities.

The first example was based on the Course Signals project and included simple visual aids such as completion of assignments and participation in discussions (Pistilli and Arnold 2010). The design principle of the first example can be grounded in the facilitation of self-regulated learning being a balancing act between necessary *external support* and desired *internal regulation* (Koedinger and Alevan 2007). For example, early warning signs from students are detected (e.g., no activity in the course room) and the learning analytics systems provides interventions to learners such as emails composed by the course instructor. Information to the learner are displayed using three visual signals: red, orange, green.

The second example included a dashboard showing general information about the student, average activities over time (e.g., submissions, learning time, logins, interactivity), and average performance comparisons across major field of study and university. In addition to the design principles of the first example, *reflective thinking* is a key design principle for the second example (Ertmer and Newby 1996). For example, learners can monitor their performance over time and receive information about predictive performance toward the end of the course. Most information to the learner are displayed as graphs and charts.

The third example provided detailed insights into learning and performance, including personalized content and activity recommendations (e.g., reading materials), self-assessments, predictive course mastery, suggestions for social interaction, and performance comparisons. The third example expands the design principles of the first and second example with multiple elements such as *prompting* (Bannert 2009), *metacognitive awareness* (Azevedo et al. 2010), or *adaptive and personalized feedback* (Kinshuk 2012). For example, chapters relevant to reach a learning outcome are recommended to the learner. In addition, learners can take a self-test based on the assigned reading material. Information to the learner are displayed in various forms such as texts, graphs, icons, and personalized prompts.

Instruments

All instruments have been developed by the research group involved in this project. In order to establish validity, an international group of experts have been involved to further develop and revise the instruments. Overall, all instruments demonstrated acceptable reliability and validity.

Control over data scale

The control over data scale (COD) focuses on access, control, and use of data in learning analytics systems. It includes four subscales: 1. Privacy of data (PLA; 5 items; Cronbach's $\alpha = 0.78$), 2. Transparency of data (TAD; 8 items; Cronbach's $\alpha = 0.72$), 3. Access of data (AOD; 11 items; Cronbach's $\alpha = 0.83$), and 4. Terms of agreement (TOA; 6 items; Cronbach's $\alpha = 0.73$). All items were answered on a five-point Likert scale (1 = not at all

important; 2 = not important; 3 = neither important nor unimportant; 4 = important; 5 = very important). See Table 2 for example items.

Sharing of data questionnaire

The sharing of data questionnaire (SOD) focuses on specific personal information participants are willing to share on learning analytics systems, such as their date of birth, educational history (self and parents), online behavior, academic performance, library usage, etc. The 28 items are answered on a Thurstone scale (1 = Agree, 0 = Do not agree; Cronbach's $\alpha = 0.74$). See Table 2 for example items.

Acceptance and use of learning analytics system

Participants rated each of the three examples of learning analytics systems regarding acceptance and use (ALA; 10 items; Cronbach's $\alpha = 0.89$). Example items included "The learning analytics system facilitates my learning", "It is too complicated to use the learning analytics system".

Demographic information

Demographic information included age, gender, university entrance qualifications, major field of study, study year, current course load, and Internet use including access, usage, and preferences.

Procedure

Over a period of 2 weeks in June 2015, students were invited to participate in the laboratory study. All instruments mentioned above were implemented as an online platform which also included the three examples of the learning analytics systems. First, participants

Table 2 Example items of instruments

Control over data scale (COD): four subscales; 30 items, Cronbach's $\alpha = 0.72$ – 0.83	<p>I agree that the learning analytics system shares/sells my personal data (PLA, item 1)</p> <p>Instructors are allowed to access your personal and user generated data at any time in order to optimize learning (TOA, item 6)</p> <p>Please indicate who you would grant access to your personal data within a Learning analytics system: Tutors of my course/researchers university-internal/examination office (AOD, item 1/4/6)</p> <p>Please indicate what is relevant for you to trust the visualizations of a Learning Analytics system: Valuation standards the analyses are based on need to be transparent (TAD, item 4)</p>
Sharing of data questionnaire (SOD): 28 items; Cronbach's $\alpha = 0.74$	<p>Please indicate whether you would agree to disclose the following data for a Learning Analytics system...</p> <p>Records of my online user paths (item 9)</p> <p>Competence test results (item 14)</p> <p>Prior academic achievement (item 23)</p>

received a general introduction regarding learning analytics and use of personal data in digital university systems. Then, participants were confronted with three different learning analytics systems. After a short time of familiarizing them with each of the learning analytics systems, we asked them to rate and compare the three different learning analytics systems for acceptance and expected use for learning (ALA; 30 min). Further, participants completed the control over data scale (COD; 30 items; 20 min) and the sharing of data questionnaire (SOD; 28 items; 20 min). Finally, participants reported their demographic information (14 items; 7 min).

Data analysis

All data stored on the online laboratory platform were anonymised, exported, and analysed using SPSS V.22. Initial data checks showed that the distributions of ratings and scores satisfied the assumptions underlying the analysis procedures. All effects were assessed at the 0.05 level. As effect size measures, we used Cohen's d (small effect: $d < 0.50$, medium effect $0.50 \leq d \leq 0.80$, strong effect $d > 0.80$) and partial η^2 (small effect: $\eta^2 < 0.06$, medium effect $0.06 \leq \eta^2 \leq 0.13$, strong effect $\eta^2 > 0.13$).

Results

Hypothesis 1: use of learning analytics systems

Student's evaluation of their use of the three learning analytics systems (ALA) differed significantly, $F(2, 987) = 9.21, p < 0.001, \eta^2 = 0.018$. Tukey-HSD post hoc comparisons between the three learning analytics systems indicated that example 3 ($M = 3.41, 95\% \text{ CI } [3.32, 3.51]$) was rated significantly higher than example 2 ($M = 3.16, 95\% \text{ CI } [3.07, 3.25]$), $p < 0.05$. Additionally, example 3 ($M = 3.41, 95\% \text{ CI } [3.32, 3.51]$) was rated significantly higher than example 1 ($M = 3.20, 95\% \text{ CI } [3.11, 3.28]$), $p < 0.05$.

Accordingly, we accept Hypothesis 1. Students prefer learning analytics systems offering a broad variety of support for learning (example 3) over other systems (examples 1 and 2).

Hypothesis 2: sharing of data in learning analytics systems

Figure 1 shows the students' overall willingness to share data. Clearly, the majority of students agree to share their course enrollment data (84 %), learning strategies test results (78 %), and motivation test results (75 %) for learning analytics purposes. In contrast, students are not willing to share their medical data (92 %), income (91 %), externally produced data, for example from social media (90 %), and marital status (87 %). Hence, students are willing to share university-relevant data, including test scores or course enrollment. However, students do not agree to share personal information and data trails of their online behavior (user path, online times, downloads, etc.).

We conducted an ANOVA to test for differences in sharing of data (SOD) for the three examples of learning analytics systems. Sharing of data (SOD) differed across the three examples of learning analytics systems, $F(2, 989) = 8.20, p < 0.001, \eta^2 = 0.016$. Tukey-HSD post hoc comparisons between sharing of data (SOD) for the three learning analytics systems indicated that participants were willing to share significantly more data for the

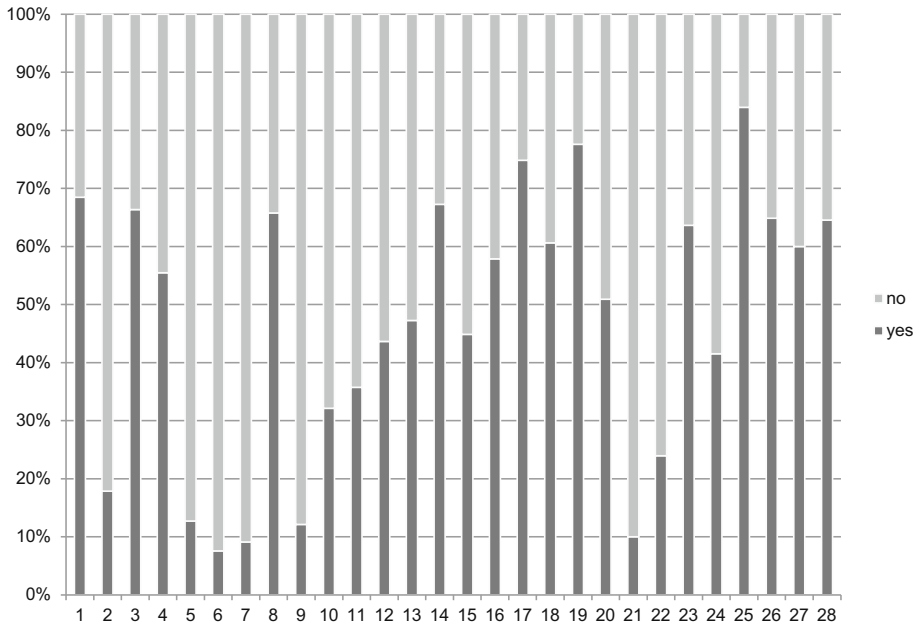


Fig. 1 Sharing of data in learning analytics systems (1 name, 2 address, 3 email, 4 date of birth, 5 marital status, 6 medical data, 7 income, 8 prior knowledge, 9 user path, 10 online times, 11 downloads, 12 course-specific discussion activity, 13 semantic analysis of posts, 14 test scores, 15 higher education entrance qualification grade, 16 school history, 17 motivation test results, 18 interest test results, 19 learning strategies test results, 20 intelligence quotient, 21 externally produced data, e.g., social media, 22 parents' educational level, 23 academic achievements, 24 occupation other than university studies, 25 course enrollment, 26 self-test scores, 27 general discussion activity, 28 library activity statistics)

learning analytics example 2 ($M = 9.46$, 95 % CI [8.75, 10.18]) than for the learning analytics example 1 ($M = 7.40$, 95 % CI [6.69, 8.11]), $p < 0.001$, and for the learning analytics example 3 ($M = 8.21$, 95 % CI [7.49, 8.93]), $p < 0.05$. However, the overall willingness to share personal data (SOD) for the three learning analytics systems was rather low ($M = 8.36$, $SD = 6.64$, $Min = 0$, $Max = 28$).

Accordingly, we accept Hypothesis 2. Students are conservative (i.e., unwilling to disclose sensitive personal information) in sharing data to support learning analytics systems (Hypothesis 2).

Hypothesis 3: relation between sharing of data and use of learning analytics systems

We examined the zero-order correlations between the acceptance and expected use of learning analytics systems (ALA) and the expected control over data (COD) as well as willingness to share data (SOD) (see Table 3).

Participants' control over data (COD) was positively related with the acceptance and expected use of the learning analytics example 1 (ALA_1 ; $r = 0.142$, $p < 0.01$), example 2 (ALA_2 ; $r = 0.216$, $p < 0.01$), and example 3 (ALA_3 ; $r = 0.439$, $p < 0.01$). Additionally, participants' control over data (COD) was positively related to their willingness to share data for example 1 (SOD_1 ; $r = 0.275$, $p < 0.01$), example 2 (SOD_2 ; $r = 0.322$, $p < 0.01$),

Table 3 Zero-order correlations and descriptive statistics for use of three learning analytics systems on sharing of data and control over data

Variable	1	2	3	4	5	6	7
(1) Use of learning analytics example 1 (ALA ₁)	–						
(2) Use of learning analytics example 2 (ALA ₂)	0.334**	–					
(3) Use of learning analytics example 3 (ALA ₃)	0.313**	0.402**	–				
(4) Control over data (COD)	0.142**	0.216**	0.439**	–			
(5) Sharing of data for learning analytics example 1 (SOD ₁)	0.166*	0.061	0.181**	0.275**	–		
(6) Sharing of data for learning analytics example 2 (SOD ₂)	0.007	0.350**	0.178**	0.322**	0.426**	–	
(7) Sharing of data for learning analytics example 3	0.117*	0.107	0.387**	0.338**	0.415**	0.427**	–
<i>M</i>	3.20	3.16	3.41	2.71	7.40	9.46	8.21
<i>SD</i>	0.76	0.83	0.86	0.39	6.53	6.60	6.66
<i>N</i>	330	330	330	330	330	330	330

* $p < 0.05$, ** $p < 0.01$

and example 3 (SOD₃; $r = 0.338$, $p < 0.01$). Finally, positive correlations were found between the acceptance and expected use of the learning analytics examples (ALA) and the willingness to share data for the three learning analytics examples (SOD) (see Table 3).

Next, we used three regression analyses to determine whether control over data (COD) and sharing of data (SOD) were significant predictors of acceptance and use of the three learning analytics systems (ALA). The results of the regression analyses are presented in Table 4, yielding a ΔR^2 of 0.032 (example 1), 0.129 (example 2), and 0.253 (example 3). Clearly, the importance of control over data (COD) and willingness to share data (SOD) positively predicted the acceptance and expected use of the learning analytics system.

Table 4 Regression analysis predicting use of learning analytics systems on sharing of data and control over data

	R^2	ΔR^2	<i>B</i>	SE <i>B</i>	β
Learning analytics example 1	0.038	0.032			
Sharing of data (SOD)			0.016	0.007	0.137*
Control over data (COD)			0.203	0.109	0.104
Learning analytics example 2	0.135	0.129			
Sharing of data (SOD)			0.039	0.007	0.313***
Control over data (COD)			0.243	0.115	0.115*
Learning analytics example 3	0.257	0.253			
Sharing of data (SOD)			0.035	0.007	0.269***
Control over data (COD)			0.766	0.111	0.348***

* $p < 0.05$, *** $p < 0.001$

We assume that greater levels of sharing of data (Hypothesis 3a) and control over data (Hypothesis 3b) are associated with the acceptance and expected use of a learning analytics system.

Discussion

At a time of growing interest in learning analytics systems used by higher education institutions, it is important to understand the implications of privacy principles to ensure that implemented systems are able to facilitate learning, instruction, and academic decision-making and do not impair students' perceptions of privacy. To a large extent, students are the producers of data used in learning analytics systems but are passive recipients of information provided on dashboards as well (Prinsloo and Slade 2014).

Participants indicated that learning analytics systems should offer a broad variety of support for learning. Such support may include self-assessments, dynamic content recommendations, visual signals, suggestions for social interaction, personalized learning activities, predictive course mastery, and adaptive scaffolds. The first example of a learning analytics system was based on the Course Signals project, including simple visual aids such as completion of assignments and participation in discussions (Arnold and Pistilli 2012), but the students rated this system significantly lower with regard to the expected support for learning in comparison with the more elaborate learning analytics systems. Likewise, they rated the second learning analytics example significant lower than the third example with regard to its expected support for learning, although it included a more advanced dashboard showing personalized student information, graphs with average activities over time (e.g., submissions, learning time, logins, interactivity), and graphs on average performance comparison across major field of study and university. Accordingly, students expected learning analytics systems including elaborate dashboards with adaptive and personalized information (example three of the learning analytics systems) to be more useful for their university studies.

Further, we found that participants are not willing to share all data. A majority of students are open to sharing data related to their university studies, but they hold back personal information and data trails of their online behavior. This fact may be critical with regard to the successful implementation of elaborated learning analytics systems, which require a vast array of data to produce the expected personalized and adaptive information. For example, participants are cautious about sharing data trails of their online behavior (e.g., user path, online times, download frequencies). However, such data includes much potential for understanding and optimizing learning processes (Gašević et al. 2016). We also found that participants were open to share more data if the learning analytics system (e.g., example 2 and 3) provided rich and meaningful information. The obvious contradiction between the requirements for diverse information (learning analytics perspective) and unwillingness to disclose too much information (student perspective) needs to be resolved before implementing learning analytics systems. Accordingly, transparency is an important driver for successfully implementing learning analytics systems at higher education institutions.

The relationship between the acceptance and use of learning analytics systems and privacy principles, that is, control over data and sharing of data, highlights the need to actively involve students and other stakeholders (e.g., teachers, instructional designers, administrators, etc.) when implementing learning analytics at higher education institutions.

With regard to control over data, ownership of learning analytics data is a critical issue and unclear from a legal perspective (Pardo and Siemens 2014). Guidelines on stewardship and protection of learning analytics data are important artifacts which can help to build and enhance trust in learning analytics systems (Clarke and Nelson 2013). With regard to sharing data, strategies for obtaining consent need to be implemented (Kay et al. 2012). (1) Students may opt in at the start of a course with further opt-in consent while changes occur. (2) Students may opt in at the start of a course with further opt-out consent while changes occur. (3) Students may opt out at the start of a course with further opt-out consent while changes occur (D. Kay et al. 2012). Accordingly, questions such as who should receive access to which data, where and how long the data will be stored, which analyses and deductions are conducted, and whether the students are aware of the data being collected from them need to be discussed in future research. Future research may also draw from other existing frameworks and adopt already existing recommendations (Clarke and Nelson 2013; Kay and Siriwardena 2001; Moore and Ellsworth 2014).

Implications

Our assumption of the contextual integrity theory, which posits that individuals may not be willing to share data from one context in another context, is supported by the results presented above (Nissenbaum 2011). Students do not want to share data they disclose on social media platforms on a learning analytics system. This may lead to limitations in the capabilities of learning analytics systems, because learning analytics systems are reliant on rich information about learners, learning environments, and curricular components (Ifenthaler and Widanapathirana 2014). Hence, algorithms implemented in learning analytics systems may be inaccurate, biased, or useless when they lack important information (Dringus 2012).

To enhance the acceptance of learning analytics systems, it is relevant to involve all stakeholders as early as possible. Students need to be considered in particular, as they take on two roles in the learning analytics system: (1) as producers of learning analytics data and (2) as recipients of the analyses derived from them (Heath 2014; Slade and Prinsloo 2013).

Transparency can be considered as one of the most crucial factors regarding the acceptance of learning analytics systems: This involves disclosing information about the collected data, its purpose, the underlying algorithms, the people who receive access to the data, and the analyses derived from them, as well as the amount of time the data will be stored and its degree of de-identification (Pardo and Siemens 2014). A simple feature for achieving high transparency in learning analytics systems could be an extended information system including *question mark buttons* close to displayed analytics data providing information about the underlying algorithms, the raw data used, storage, access, etc. Furthermore, a FAQ or an alert system could be implemented. An information system of this kind could be suitable for tackling the transparency paradox (Nissenbaum 2011), making it possible to achieve a depth of information that meets the students' expectations.

Higher education institutions may use a privacy calculus model to inform stakeholders about the complex decisions required for learning analytics systems (Dinev and Hart 2006). Figure 2 shows the deliberation process for disclosing information for learning analytics systems. Students assess their concern over privacy on the basis of the specific information required for the learning analytics system (e.g., name, learning history, learning path, assessment results, etc.). This decision can be influenced by risk-minimizing factors (e.g., trust in the learning analytics systems and/or institution, control over data

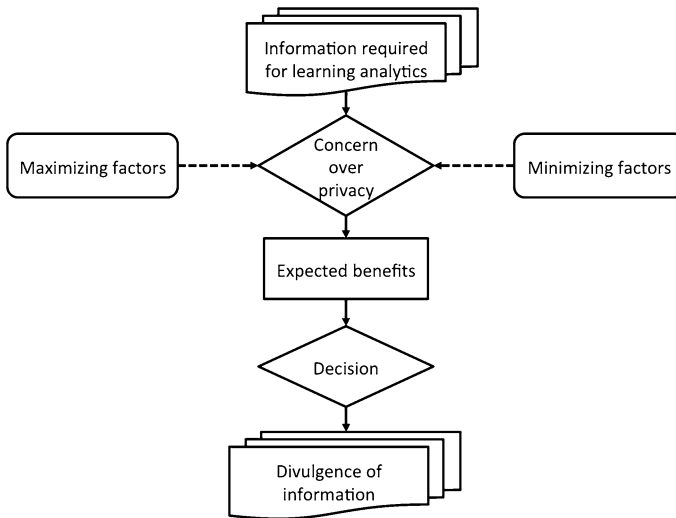


Fig. 2 Deliberation process for sharing information for learning analytics systems

through self-administration) and risk-maximizing factors (e.g., non-transparency, negative reputation of the learning analytics system and/or institution). Concerns over privacy are then weighed against the expected benefits of the learning analytics system. The probability that the students will disclose required information is higher if they expect the benefits to be greater than the risk. Hence, the decision to divulge information on learning analytics systems is a cost–benefit analysis based on available information to the student.

Limitations and future work

This study is limited in several aspects that must be addressed in future research. First, the results rely on self-report data assuming honest and accurate responses for all administered instruments and collected data and thus possibly include a response bias, meaning that participants tend to respond in a certain way regardless of the phenomenon in question. Second, the sample was from a single institution. Future studies might include a multi-institutional perspective and international comparisons, as ethics and privacy are dependent on different cultural perspectives (Pitta et al. 1999). Third, the participants had no prior experience in using learning analytics systems at their institution and their prior knowledge about learning analytics was very limited. Moreover, they were only exposed to the three examples of learning analytics systems for a very limited time and in a linear order. Therefore, the findings of the presented research are limited toward external validity. Future studies might focus on students with experience in using learning analytics systems as well as experimentally varying the order of presentation of different learning analytics systems.

Additionally, future studies should expand the perspective of ethics and privacy to include different stakeholders, such as teachers, tutors, learning designers, departmental chairs and deans of schools, university management, and governing authorities. Also, learning analytics research needs to address issues where students do not want to share data, however, systems would require these data to produce reliable and valid results. Finally, future studies should be conducted as mixed-methods designs, including

qualitative components which may help to better understand the drivers of privacy issues in learning analytics.

Conclusion

Learning analytics provides the pedagogical and technological background for producing real-time interventions at all times during the learning process. The availability of personalized, dynamic, and timely feedback could support the learner's self-regulated learning and increases their motivation and success. However, such automated systems may also hinder the development of competencies such as critical thinking and autonomous learning (Ifenthaler 2015b).

Procedures regulating access to and usage of educational data need to be conducted before learning analytics are implemented (Willis and Strunk 2015). The storage and processing of anonymized personal data is only a small step towards a more comprehensive educational data governance structure for learning analytics (Prinsloo and Slade 2015). Students are more than shattered bits of information provided and produced while interacting with learning analytics systems implemented by higher education institutions (Solove 2004). Learning analytics may reveal personal information and insights into an individual learning history, but they are not accredited and far from being unbiased, comprehensive, and valid.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Aïmeur, E., Hage, H., & Onana, F. S. M. (2008). *Anonymous credentials for privacy-preserving e-learning*. Paper presented at the International MCETECH Conference.
- Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, 58(1), 470–489.
- Arnold, K. E., & Pistilli, M. D. (2012). *Course signals at Purdue: using learning analytics to increase student success*. Paper presented at the 2nd International Conference on Learning Analytics and Knowledge, Vancouver.
- Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated learning with MetaTutor: advancing the science of learning with metacognitive tools. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning* (pp. 225–247). New York: Springer.
- Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie*, 23(2), 139–145.
- Berland, M., Baker, R. S. J. D., & Bilkstein, P. (2014). Educational data mining and learning analytics: applications to constructionist research. *Technology, Knowledge and Learning*, 19(1–2), 205–220. doi:10.1007/s10758-014-9223-7.
- Clarke, J., & Nelson, K. (2013). Perspectives on learning analytics: issues and challenges. Observations from Shane Dawson and Phil Long. *The International Journal of the First Year in Higher Education*, 4(1), 1–8. doi:10.5204/intjfyhe.v4i1.166.
- Dinev, T., & Hart, P. (2006). An extended privacy calculus mode for E-Commerce transactions. *Information Systems Research*, 17, 61–80.
- Dringus, L. P. (2012). Learning analytics considered harmful. *Journal of Asynchronous Learning Networks*, 16(3), 87–100.
- Elgesem, D. (1999). The structure of rights in directive 95/46/EC on the protection of individuals with regard to the processing of personal data and the free movement of such data. *Ethics and Information Technology*, 1(4), 283–293.

- Ertmer, P. A., & Newby, T. J. (1996). The expert learner: strategic, self-regulated, and reflective. *Instructional Science*, 24(1), 1–24.
- Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: the effects of instructional conditions in predicting academic success. *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. doi:10.1007/s11528-014-0822-x.
- Gonzalez, D. (2015). *Managing online risk: Apps, mobile, and social media security*. Oxford: Elsevier.
- Greller, W., & Drachler, H. (2012). Translating learning into numbers: a generic framework for learning analytics. *Educational Technology & Society*, 15(3), 42–57.
- Heath, J. (2014). Contemporary privacy theory contributions to learning analytics. *Journal of Learning Analytics*, 1(1), 140–149.
- Ifenthaler, D. (2015a). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Thousand Oaks: Sage.
- Ifenthaler, D. (2015b). Model-based approaches. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 512–525). Thousand Oaks: Sage.
- Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: two case studies using support vector machines. *Technology, Knowledge and Learning*, 19(1–2), 221–240. doi:10.1007/s10758-014-9226-4.
- Kay, D., Korn, N., & Oppenheim, C. (2012). Legal, risk and ethical aspects of analytics. *Cetis Analytics Series*, 1(6), 1–30.
- Kay, R., & Siriwardena, A. K. (2001). The process of informed consent for urgent abdominal surgery. *Journal of Medical Ethics*, 27(3), 157–161. doi:10.1136/jme.27.3.157.
- Kinshuk. (2012). Guest editorial: Personalized learning. *Educational Technology Research and Development*, 60(4), 561–562. doi:10.1007/s11423-012-9248-3.
- Koedinger, K. R., & Alevan, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264.
- Long, P. D., & Siemens, G. (2011). Penetrating the fog: analytics in learning and education. *EDUCAUSE Review*, 46(5), 31–40.
- Macfadyen, L., & Dawson, S. (2012). Numbers are not enough Why e-Learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15(3), 149–163.
- Moore, S. L., & Ellsworth, J. B. (2014). Ethics of educational technology. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 113–127). New York: Springer.
- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119–157.
- Nissenbaum, H. (2011). A contextual approach to privacy online. *Deedalus*, 140(4), 32–48.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*. doi:10.1111/bjet.12152.
- Pistilli, M. D., & Arnold, K. E. (2010). Purdue Signals: mining real-time academic data to enhance student success. *About campus: Enriching the student learning experience*, 15(3), 22–24.
- Pitta, D. A., Fung, H.-G., & Isberg, S. (1999). Ethical issues across cultures: managing the differing perspectives of China and the USA. *Journal of Consumer Marketing*, 16(3), 240–256.
- Prinsloo, P., & Slade, S. (2014). Student data privacy and institutional accountability in an age of surveillance. In M. E. Menon, D. G. Terkla, & P. Gibbs (Eds.), *Using data to improve higher education. Research, policy and practice* (pp. 197–214). Rotterdam: Sense Publishers.
- Prinsloo, P., & Slade, S. (2015). *Student privacy self-management: Implications for learning analytics*. Paper presented at the Fifth International Conference on Learning Analytics and Knowledge, New York.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. doi:10.1177/0002764213479366.
- Solove, D. J. (2004). *The digital person: Technology and privacy in the information age*. New York: New York University.
- Tinto, V. (2005). Reflections on student retention and persistence: moving to a theory of institutional action on behalf of student success. *Studies in Learning, Evaluation, Innovation and Developmental Psychology*, 2(3), 89–97.
- Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, 4(5), 193–220.
- Weippl, E. R., & Min Tjoa, A. (2005). Privacy in e-learning: anonymity, pseudonyms and authenticated usage. *Interactive Technology and Smart Education*, 2(4), 247–256.
- Willis, I. J. E., & Strunk, V. A. (2015). Ethical responsibilities of preserving academicians in an age of mechanized learning: Balancing the demands of educating at capacity and preserving human

- interactivity. In J. White & R. Searle (Eds.), *Rethinking machine ethics in the age of ubiquitous technology* (pp. 166–195). Hershey: IGI Global.
- Yeaman, A. R. J., Koetting, J. R., & Nichols, R. G. (1994). Critical theory, cultural analysis, and the ethics of educational technology as social responsibility. *Educational Technology*, 34(2), 5–13.

Dirk Ifenthaler is Professor for Learning, Design and Technology at University of Mannheim, Germany, Adjunct Professor at Deakin University, Australia, and Affiliate Research Scholar at the University of Oklahoma, USA. His research focuses on the intersection of cognitive psychology, educational technology, learning science, and data analytics. Dirk Ifenthaler's research outcomes include numerous co-authored books, book series, book chapters, journal articles, and international conference papers, as well as successful grant funding in Australia, Germany, and USA—see Dirk's website for a full list of scholarly outcomes at www.ifenthaler.info. He is the Editor-in-Chief of the Springer journal *Technology, Knowledge and Learning* (www.springer.com/10758).

Clara Schumacher is research assistant for Learning, Design and Technology at the University of Mannheim, Germany. Her research interest focuses on informal learning, educational technology, and learning analytics.