DIVULGING PERSONAL INFORMATION WITHIN LEARNING ANALYTICS SYSTEMS

Dirk Ifenthaler[1] and Clara Schumacher[2]
[1]*University of Mannheim and Deakin University, L4, 1, 68131 Mannheim, Germany*
[2]*University of Mannheim, L4, 1, 68131 Mannheim, Germany*

ABSTRACT

The purpose of this study was to investigate if students are prepared to release any personal data in order to inform learning analytics systems. Besides the well-documented benefits of learning analytics, serious concerns and challenges are associated with the application of these data driven systems. Most notably, empirical evidence regarding privacy issues for learning analytics is next to nothing. A total of 330 university students participated in an exploratory study confronting them with learning analytics systems and associated issues of control of data and sharing of information. Findings indicate that sharing of data for educational purposes is correlated to study related constructs, usage of Internet, awareness of control over data, and expected benefits from a learning analytics system. Based on the relationship between the willingness to release personal data for learning analytics systems and various constructs closely related to individual characteristics of students it is concluded that students need to be equally involved when implementing learning analytics systems at higher education institutions.

KEYWORDS

Learning analytics; higher education; control over data; transparency

1. INTRODUCTION

Massive administrative, systems, academic, and personal data within 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 dynamic information about learners and learning environments, assessing, eliciting and analyzing it, for real-time modeling, prediction, and optimization of learning processes, learning environments, and educational decision-making (Ifenthaler, 2015).

Serious concerns and challenges are associated with the application of learning analytics (Pardo & Siemens, 2014). For instance, not all educational data is relevant and equivalent. Therefore, the validity of data and its analyses is critical for generating useful summative, real-time, and predictive insights (Macfadyen & Dawson, 2012). Furthermore, limited access to educational data may generate disadvantages for involved stakeholders. For example, invalid forecasts may lead to inefficient decisions and unforeseen problems (Ifenthaler & Widanapathirana, 2014). Moreover, ethical and privacy issues are associated with the use of educational data for learning analytics. That implies how personal data are collected and stored as well as how they are analyzed and presented to different stakeholders (Slade & Prinsloo, 2013).

Currently, most research towards privacy issues in learning analytics refers to guidelines from other disciplines such as Internet security or medical environments (Pardo & Siemens, 2014). However, due to the contextual characteristics of privacy an adoption from other contexts is not recommendable (Nissenbaum, 2004). More importantly, empirical evidence regarding privacy issues for learning analytics is scarce. Therefore, the purpose of this exploratory study was to investigate if students are willing to release any personal data for informing learning analytics systems.

ISBN: 978-989-8533-43-2 © 2015

2. RELEASE OF PERSONAL DATA

2.1 Privacy in the Digital World

The most general definition of privacy is freedom from interference or intrusion (Warren & 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 the possibility of individuals to influence the flow of their personal information and to hamper others to access their information (Heath, 2014).

Within the digital world, this view on privacy seems to be no longer valid. Many individuals are willing to share personal information without being aware of who has access to the provided data and how the data will be used as well as how to control ownership of the provided data (Solove, 2004). Accordingly, data are generated and provided automatically through online systems which limits the control and ownership of personal information in the digital world (Slade & Prinsloo, 2013). Only recently, this phenomenon has been adopted by higher education institutions through the implementation of learning analytics.

2.2 Privacy Principles for Learning Analytics

Higher education institutions have always used a variety of data about students, such as socio-demographic information, higher education entrance qualification grades, or pass and fail rates, for academic decision-making as well as resource allocation (Long & Siemens, 2011; Prinsloo & Slade, 2014). Such data can help to successfully predict dropout rates of first-year students and implement strategies to support learning and instruction as well as to retain students (Tinto, 2005).

Accordingly, advanced digital technologies and learning analytics systems enable higher education institutions to collect dynamic real-time data from all student activity within the higher education institutions' systems which offers huge potential for personalized and adaptive learning experiences and support (Berland, Baker, & Bilkstein, 2014). Consequently, higher education institutions are required to address privacy issues linked to learning analytics: They need to define who gets access to which data, where and how long will the data be stored, and which procedures and algorithms are implemented to further use the available data.

Slade and Prinsloo (2013) as well as Pardo and Siemens (2014) established several principles for privacy and ethics 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, as well as purpose, analyzes, access, control, and ownership of the data. However, empirical evidence toward student perceptions of privacy principles related to learning analytics is lacking.

2.3 Purpose of the Study

Empirical research is currently addressing the validity and effectiveness of learning analytics systems for learning, instruction, and educational decision-making (Ali, Hatala, Gašević, & Jovanović, 2012; Gašević, Dawson, & Siemens, 2015; Ifenthaler & Widanapathirana, 2014). In contrast, ethics and privacy issues in learning analytics are in an early stage of research (Pardo & Siemens, 2014).

In this regard, the purpose of this exploratory study was to investigate if students are willing to release any personal data for informing learning analytics systems and if other constructs such as study interest and use of Internet are related.

It is argued that first year students have to adjust to different learning and teaching requirements, manage workloads and course loads, as well as matching the universities' expectations and personal interest (Bowles, Fisher, McPhail, Rosenstreich, & Dobson, 2014). Learning analytics may provide scaffolds to overcome the before mentioned hurdles especially in the first year of university studies. Specifically, we assume that divulging personal information within learning analytics systems is related to study related constructs such as year of study (Hypothesis 1a), course load (Hypothesis 1b), and study interest (Hypothesis 1c).

Another factor which may influence the use and acceptance of learning analytics are the students' technology competencies (Kennedy, Judd, Churchward, Gray, & Krause, 2008). It is increasingly recognized that a majority of students possess a core set of technology-based competencies, however, no empirical evidence exists how these competencies influence the use and acceptance of learning analytics systems. For example, Trepte, Dienlin, and Reinecke (2013) report that students who frequently use social media tools are more open to disclose personal information in online environments. Therefore, we assume that releasing any personal data for learning analytics systems is related to the students' percentage of use of the Internet for learning (Hypothesis 2a) and social media (Hypothesis 2b).

Students' trust and control with regard to online systems in general and learning analytics systems in particular may be another factor guiding the use and acceptance of learning analytics (Ennen, Stark, & Lassiter, 2015; Nam, 2014). We also assume that student's willingness to provide personal data is related to their anticipated control over data (Hypothesis 3).

Last, students may disclose personal data for learning analytics systems if the overall benefits for learning are greater than the assessed risk of releasing personal data (Culnan & Bies, 2003). We assume that releasing personal data for learning analytics systems is related to the anticipated benefits from a specific learning management system (Hypothesis 4).

3. METHOD

3.1 Participants and Design

The study was designed as an 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 removing 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 studied in the Bachelors program (80%), with 20% of the participants studying in the Masters program. The average course load in the current semester was 5 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.

3.2 Instruments

3.2.1 Study Interest Questionnaire

The study interest questionnaire (FSI; Schiefele, Krapp, Wild, & Winteler, 1993) includes 18 items (Schiefele, Krapp, Wild & Winteler, 1993) which focus on study-related interest such as feeling- and value-related valences as well as intrinsic orientation (Cronbach's α = .90). 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).

3.2.2 Technology Affinity Scale

The technology affinity scale (TAS) focuses on information behavior to indicate educationally relevant activity, such as information seeking and sharing (Mills, Knezek, & Wakefield, 2013). TAS consists of 22 items which 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) (Cronbach's α = .645).

3.2.3 Control over Data Scale

The control over data scale (COD) focuses on access, control, and use of data in learning analytics systems, including four subscales: 1. Privacy of data (PLA; 5 items; Cronbach's α = .78), 2. Transparency of data (TAD; 8 items; Cronbach's α = .72), 3. Access of data (AOD; 11 items; Cronbach's α = .83), and 4. Terms of agreement (TOA; 6 items; Cronbach's α = .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).

3.2.4 Sharing of Data Questionnaire

The sharing of data questionnaire (SOD) focuses on specific personal information participants are willing to share in learning analytics systems, such as 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 α = .74).

3.2.5 Demographic Information

Demographic information included age, gender, Internet usage for learning and social media, years of study, study major, course load, etc.

3.3 Learning Analytics Systems

Three different examples of learning analytics systems were presented to the participants. The first example was based on the Course Signals project including simple visual aids such as completion of assignments, participation in discussion (Pistilli & Arnold, 2010). 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 comparison across study major and university. The third example provided detailed insights into learning and performance including personalized content and activity recommendation (e.g. reading materials), self-assessments, predictive course mastery, suggestions for social interaction, and performance comparisons. Participants rated each of the examples regarding acceptance of the learning analytics system and expected benefits for learning (ALA; 10 items; Cronbach's α = .89).

3.4 Procedure

Over a period of two weeks in June 2015, students were invited to participate in the laboratory study which included three parts. In the first part, participants received a general introduction regarding learning analytics and use of personal data in digital university systems. Then they completed the study interest questionnaire (FSI; 18 items; 8 minutes) and the technology affinity scale (TAS; 22 items; 10 minutes). In the second part, participants were confronted with three different learning analytics systems. After a short time to familiarize with each of the learning analytics system, they were asked to rate acceptance and expected use for learning of the learning analytics systems as well as to compare the three different systems (30 minutes). In the third part, participants completed the control over data scale (COD; 30 items; 20 minutes) and the sharing of data questionnaire (SOD; 28 items; 20 minutes). Finally, participants reported their demographic information (14 items; 7 minutes).

4. RESULTS

Table 1 shows the zero-order correlations among the variables with regard to the first set of hypotheses. Students' *study year* was negatively related to their *course load*, as was their percentage of *Internet use for social media*. Students' *study year* was positively related to their percentage of *Internet use for learning*, as was their anticipated *control over data*. Their *study interest* was related to their anticipated *control over data*. Additionally, their percentage of *Internet use for learning* was positively related to their anticipated *control over data* as well as their expected *benefits of the learning analytics system*. Finally, students' anticipated *control over data* was positively related to their expected *benefits of the learning analytics system*.

A hierarchical regression analysis was used to determine whether study related variables (SY, CL, FSI), Internet usage (IUL, IUS), control over data (COD), and expected benefits of learning analytics systems (BLA) were significant predictors of *sharing of data for a specific learning analytics system* (SOD; dependent variable). Table 2 shows the four steps of entering data into the equation. The final regression model explained a statistically significant amount of variance in *sharing of data* (SOD), $\Delta R^2 = .370$, $F(7, 329) = 28.58$, $p < .001$.

Table 1. Descriptives and zero-order correlations for study related variables, Internet usage variables, and data as well as learning analytics related variables ($N = 330$)

Variable	1	2	3	4	5	6	7
1. Study year (SY)	-						
2. Course load (CL)	-.378***	-					
3. Study interest (FSI)	-.008	.071	-				
4. Internet use for learning (IUL)	.123*	-.076	.014	-			
5. Internet use for social media (IUS)	-.156**	.023	-.066	-.032	-		
6. Control over data (COD)	.141**	-.038	.111*	.290***	.007	-	
7. Benefits of learning analytics system (BLA)	.076	-.017	-.009	.630***	-.006	.362***	-
M	3.58	5.36	2.99	35.00	32.95	2.71	3.13
SD	2.30	1.70	.28	21.21	20.43	.39	.97

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 2. Regression analysis predicting sharing of data on study related variables, Internet usage, control over data, and expected benefits of learning analytics systems ($N = 330$)

	R^2	ΔR^2	*B*	*SE B*	β
Step 1	.038	.029			
Study year (SY)			.538	.170	.186**
Course load (CL)			-.081	.231	.726
Study interest (FSI)			-.094	1.295	.942
Step 2	.322	.311			
Study year (SY)			.432	.145	.149**
Course load (CL)			.010	.195	.002
Study interest (FSI)			-.127	1.093	-.005
Internet use for learning (IUL)			.165	.014	.525***
Internet use for social media (IUS)			.040	.015	.122**
Step 3	.352	.340			
Study year (SY)			.366	.143	.127*
Course load (CL)			-.005	.191	-.001
Study interest (FSI)			-.609	1.077	-.026
Internet use for learning (IUL)			.149	.015	.474***
Internet use for social media (IUS)			.037	.015	.114*
Control over data (COD)			3.144	.806	.185***
Step 4	.383	.370			
Study year (SY)			.373	.140	.129**
Course load (CL)			-.035	.186	-.009
Study interest (FSI)			-.376	1.054	-.016
Internet use for learning (IUL)			.106	.018	.339***
Internet use for social media (IUS)			.037	.014	.113*
Control over data (COD)			2.339	.813	.138**
Benefits of learning analytics system (BLA)			1.606	.400	.234***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Specifically, students' study year (SY) positively predicted their willingness to share personal data for a specific learning analytics system (SOD), indicating that the higher the study year (SY), the higher the students' liberality to provide personal data for educational purposes. Accordingly, Hypothesis 1a is accepted.

ISBN: 978-989-8533-43-2 © 2015

The percentage of Internet usage for learning (IUL) and social media (IUS) positively predicted the students' release of personal data for learning analytics purposes (SOD), indicating the higher the usage of the Internet for learning and social media, the higher their disposition to share personal data for learning analytics systems. Hence, Hypotheses 2a and 2b are accepted.

The student' awareness about control of data (COD) positively predicted their preparedness to share personal data for a specific learning analytics system (SOD), indicating that the higher the awareness about the control of personal data, the higher their disposition to share personal data for learning analytics systems. Thus, Hypothesis 3 is accepted.

The expected benefits of a learning analytics system (BLA) positively predicted the students' release of personal data for learning analytics purposes (SOD), indicating the higher the expected benefit of the learning analytics system, the higher the readiness to provide personal data for learning analytics purposes. Consequently, Hypothesis 4 is accepted. As shown in Table 2, no significant correlations were found for course load (CL) and study interest (FSI). So, Hypotheses 1b and 1c are rejected.

5. DISCUSSION

At a time of growing interest in learning analytics systems of 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 extend, students are the producers of data used in learning analytics systems, however, passive recipients of information provided in dashboards (Prinsloo & Slade, 2014).

The findings of this exploratory study highlight an overall interest of students in learning analytics systems. As students mature in their higher education studies they seem to be more aware of the context of sharing educational data (Bailey, Ifenthaler, Gosper, Kretzschmar, & Ware, 2015). To make the benefits of learning analytics and emphasize the need of sharing data within the learning analytics system to first year students, tutoring systems and/or training sessions need to be implemented accordingly.

Students spend a large amount of time for using the Internet for learning and social media activities. Not surprisingly, spending time on the Internet is associated with the openness of sharing data for learning analytics systems. This effect may be explained by trust students generate with regard to online systems in general and learning analytics systems in particular (Ennen et al., 2015). The relationships between perceived control over personal data and expected benefits as well as sharing personal data is closely related to the phenomenon of trust (Nam, 2014). These findings indicate that a high computer literacy is prerequisite for the acceptance of learning analytics as well as the willingness to share data and should be systematically trained.

From a holistic point of view, learning analytics may provide multiple benefits for higher education institutions and for involved stakeholders and different data analytics strategies can be applied to produce summative, real-time and predictive insights (Ifenthaler, 2015). For example, students may use *summative learning analytics* implemented as an interactive dashboard to analyze learning outcomes of individual courses after completing a semester of study or track their progress towards self-defined goals (e.g., credit points). Students may also be able to compare their own learning paths and outcomes between individual units or courses. This may enable students to understand their learning habits and to adjust their learning strategies as well as private habits in order to be successful in their studies. On the same dashboard or within a learning management system, students may receive *real-time learning analytics information* based on their currently available data. Automated interventions may point them to learning materials and tips for progressing further in a particular study unit. Students may take self-assessment on a specific topic and receive just-in-time feedback or get recommendations to participate in online discussions or connect to peers using preferred social media. *Predictive learning analytics* for students may help to optimize the learning path in a specific study unit by providing them probabilities of success when choosing a particular pathway. Such predictions are expected to increase the overall engagement and success rates of students (Ifenthaler, 2015).

However, reliable and valid learning analytics systems require rich and current information of students including personal characteristics and preferences, academic performance, educational pathways and logfiles of various online learning systems. If the underlying learning analytics algorithms do not have access to the required information, the above-described benefits cannot be produced. While higher education institutions implement learning analytics systems, students may find themselves in a dilemma situation concerning the divulgence of personal information for learning analytic systems. In order to overcome such a dilemma situation, it is necessary to provide students transparency of the implemented learning analytics system and its underlying algorithms, as well as clear guidelines towards access, analysis, control, ownership and use of relevant data.

6. CONCLUSION

Remaining questions such as who should get access to which data, where and how long will the data be stored, which analyzes and deductions are conducted and are the students aware of the data collected from them need to be discussed in prospective research. From an instructional design point of view, research may focus on usability, personalization, and adaptivity of learning analytics systems. Understanding these factors may be crucial for implementing learning analytics systems at higher education institutions. Student's computer literacy is expected to be a prerequisite for using learning analytics systems. Professional development for learning analytics including transparency of underlying algorithms and involving all relevant stakeholders in the development and implementation phases may help to increase trust and acceptance in the systems.

Students are more than shattered bits of information given 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, however, they are not accredited and far from being unbiased, comprehensive, and valid.

REFERENCES

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.

Bailey, M., Ifenthaler, D., Gosper, M., Kretzschmar, M., & Ware, C. (2015). The changing importance of factors influencing students' choice of study mode *Technology, Knowledge and Learning, 20*(2), 169–184. doi: 10.1007/s10758-015-9253-9

Berland, M., Baker, R. S., & 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

Bowles, A., Fisher, R., McPhail, R., Rosenstreich, D., & Dobson, A. (2014). Staying the distance: students´ perception of enablers of transition to higher education. *Higher Education Research & Development, 33*(2), 212–225.

Culnan, M. J., & Bies, R. J. (2003). Consumer privacy: Balancing economic and justice considerations. *Journal of Social Issues, 59*(2), 323–342. doi: 10.1111/1540-4560.00067

Ennen, N. L., Stark, E., & Lassiter, A. (2015). The importance of trust for satisfaction, motivation, and academic performance in student learning groups. *Social Psychology of Education*. doi: 10.1007/s11218-015-9306-x

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.

Heath, J. (2014). Contemporary privacy theory contributions to learning analytics. *Journal of Learning Analytics, 1*(1), 140–149.

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Thousand Oaks, CA: 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

Kennedy, G., Judd, T., Churchward, A., Gray, K., & Krause, K. (2008). First year students' experiences with technology: are they really digital natives? *Australasian Journal of Educational Technology, 24*(1), 108-122.

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.

Mills, L. A., Knezek, G. A., & Wakefield, J. S. (2013). Understanding information seeking behavior in technology pervasive learning environments of the 21st century. *Turkish Online Journal of Educational Technology, 12*(4), 200–208.

Nam, C. W. (2014). The effects of trust and constructive controversy on student achievement and attitude in online cooperative learning environments. *Computers in Human Behavior, 37*, 237–248. doi: 10.1016/j.chb.2014.05.007

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*(1), 119–157.

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.

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.

Schiefele, U., Krapp, A., Wild, K. P., & Winteler, A. (1993). Der „Fragebogen zum Studieninteresse“ (FSI). *Diagnostica, 39*(4), 335–351.

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, NY: New York University Press.

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.

Trepte, S., Dienlin, T., & Reinecke, L. (2013). Privacy, self-diclosure, social support, and social network site use. Stuttgart: University of Hohenheim.

Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review, 4*(5), 193–220.