Aware, but Don't Really Care: Student Perspectives on Privacy and Data Collection in Online Courses

Phu Vu, University of Nebraska at Kearney
Megan Adkins, University of Nebraska at Kearney
Shelby Henderson, University of Nebraska at Kearney

Abstract

The purpose of this study was to examine student viewpoints about privacy and personal data collection in online courses in U.S. higher education settings. Results of data analysis revealed that students were aware that their learning behaviours (such as login frequency, pages viewed or clicked, and learning profiles) could be monitored and recorded by their instructors. Additionally, they were not concerned about their learning behaviours being monitored, recorded, or collected for academic research, and used for instructional/teaching improvement purposes. There was no evidence of significant difference between students' gender (female and male) in terms of their awareness and concern about their privacy in online learning settings.

Keywords: online learning; elearning; student privacy; data analytics

Introduction and literature review

According to the annual New Media Consortium Horizon Reports (2016, 2017, and 2018), learning analytics has the potential to accelerate higher education technology adoption to help educational institutions increase student retention, improve student success, and reduce the burden of accountability. Although online learning has been appreciated by many researchers and educators, there are concerns about student privacy of online activities and ethics in relation to the data analytics collected (Greller & Drachsler, 2012). Boundaries can be crossed easily, and the privacy and security of students' information could be easily breached in the process of collecting large amounts of data (Drachsler & Greller, 2016; Pardo & Siemens, 2014). The purpose of this study was to examine student viewpoints about privacy and personal data collection in online learning settings.

Learning analytics is defined as the collection, analysis and use of large amounts of student data and information to understand learner behaviours and contexts (both digital and analogue) to improve the educational outcomes of students and to increase institutional effectiveness and efficiency. Analytics has grown in importance due to the increased pressure for higher education to reform and provide educational platforms online to educate students in remote settings (Rubel & Jones, 2016). Learning analytics began to be used when universities were able to track their online learners through their learning management systems (LMSs). These systems provide data sets that help universities to create online courses, provide an avenue to deliver learning materials, and collect information about enrolment numbers. As technology and the way it was used evolved, so did the development of additional analytics that could be integrated with an LMS. This included social learning applications and learning to use real-time data to help organisations improve their infrastructure. Moreover, as LMSs are implemented in higher

educational institutions, additional analytics are collected. Examples of these analytics are found in online navigation—including the sites students visit, how long they are on them, whether they complete the tasks, and how long they hover over question options during a test. Student data is also traced throughout the university to track library activity and use of facilities on campus (e.g., campus recreation facilities). All of the data-based systems create an infrastructure for universities to use the information they gather about students' activities.

Another analytical tool used is the process of changing complex data into meaningful patterns and values. This is called big data. Big data is a resource for capturing, storing, distributing, managing, and analysing larger data sets with diverse structures (Daniel, 2015). It encompasses analytical techniques such as descriptive analytics and mining/predictive analytics. Big data analytics tools have sophisticated functionality to facilitate student information integration and provide insights to help universities meet the market needs and future market trends, and thus improve the quality of educational and financial performance. Big data is not a new or isolated phenomenon; it is part of a long evolution of capturing data. The concept of big data began in 1999, when the number of online devices and the potential for them to communicate with each other started to expand. During the birth of Web 2.0, the term "big data" emerged as a result of the large volume of data available and the potential for it to be used to help company analytics. The application of learning analytics and big data is credited with helping to improve learning performances and retain students (Becker et al., 2017; Gandomi & Haider, 2015; Romero & Ventura, 2017; Siemens & Long, 2011; Vu, Meyer, & Capero, 2016).

In recent years, as the collection of learning analytics and big data has become more popular in higher education, the issue of privacy and security of student information has been seriously questioned. A leading concern about the collection of student data is that it could end up in a state or national governmental database, which would allow many people to access students' private information and data (Picciano, 2012). Another concern about the collection of data, especially for students' parents, relates to the ability for anyone else to access a student's private information. This information could include their name, address, student identification number, email address, and phone number (Rubel & Jones, 2016). A leak in the database could expose an individual's private information to misuse—causing undue burden or humiliation.

Moreover, the collection of learning analytics and big data may contribute to students feeling they are being watched, causing some students to avoid certain courses if they know data will be collected and their privacy could be compromised (Picciano, 2012). This is a huge concern, as learning analytics was developed to improve learning, not hinder it. Researchers and teachers must consider students' privacy if quality studies are to be completed and the learning environment improved. Compromised students' privacy could be detrimental for student learning and the success of the institution (Watters, 2011). If the use of learning analytics and big data continues, researchers and institutions will need to find a more secure alternative to protect the privacy and security of students' information when it is collected and distributed for research purposes.

As presented above, there are a number of viewpoints about using learning analytics and big data in higher education. However, researchers and educators all seemed to come to a consensus about how students' data should be protected. One way to protect student privacy is to tell them what data will be collected and give them authority, before the researcher proceeds, to refuse collection of some or all of that information (Watters, 2011). Students who approve the collection of their data should be given details about what data will be collected, how it will be accessed, and who will be able to see it. They should also be able to request error corrections in the analytics (Pardo & Siemens, 2014). To build student trust, the data management procedures should be clearly outlined, so they understand the process. They should also be told how long data will be stored, where it will be stored, who will have access to it, and how it will be

destroyed after the stated time. The purposes and benefits of a study relating to data analytics should be explained so they understand why their data is needed, and how the information gathered could help educational research. The principles of privacy and ethics for both the particular country and educational institution in which research is occurring must be followed (Black, Dawson, & Priem, 2008; Daniel, 2015; Lewis, Kaufman, & Christakis, 2008). It is vital that the principles of transparency, student control, security, accountability, and assessment are followed by all researchers who collect, analyse, and report learning analytics.

Although researchers, educators, and educational administrators have raised concerns about student privacy and ethical issues over the practice of collecting massive amounts of student data, little research has been conducted into student perspectives. Therefore, the purpose of this study was to examine student viewpoints in relation to privacy and data collection in online learning settings.

Research method

The primary purpose of this exploratory study was to use quantitative research strategies (via an online survey) to investigate students' stances on privacy issues in their online courses. More specifically, the study aimed to answer the following research questions.

1. Were students aware that their learning behaviours in online learning settings can be monitored and recorded by their instructors?
2. Were students concerned about instructors using their learning data for academic or research purposes?

Research instrument

A tow-part online survey was used for data collection. The first part of the survey collected demographic information that included the participant's gender (female or male) and educational level (undergraduate or graduate). The second part of the survey had six statements with a five-point scale ranging from one to five. The instrument was piloted with 55 participants who were undergraduate- or graduate-level students in one of the researchers' colleges. After collecting participants' responses in the pilot, the researchers conducted a reliability test to measure the consistency of the questions in the instrument. To be more specific, the reliability test examined whether the six questions in the second part of the survey measured related aspects of the issue under investigation. The resulting alpha values are reported in
Table 1.

Table 1 Reliability test statistics

Cronbach's Alpha	Cronbach's Alpha based on standardised items	No. of items
.88	.84	6

Cronbach's Alpha showed a value of .84 which, according to Mallery & George (2003) indicated that the survey items had good internal consistency.

When they had completed the pilot, researchers sent the survey to the listserv of the Association for Educational Communications and Technology (AECT) and the alumni listserv of Online Learning Consortium (OLC), asking colleagues in the listservs to help recruit students to complete the online survey. The researchers also reached out to the Student Affairs Offices at three large public research universities in the U.S. midwest to help share the survey request to

students who were enrolled in online courses at their institutions. The survey was available for students to complete for 3 months and the researchers sent a follow-up message each month to remind colleagues to share the survey link with their students. The goal of the research team was to receive at least 1,500 student responses so the next phase of data analysis and reporting could be conducted. The researchers did not conduct a power analysis to determine the sample size of the participants in this exploratory study.

Participants

The survey was administered online over 3 months and yielded 1,752 responses from students across the U.S. higher educational institutions. One hundred and five incomplete responses were removed from the response pool before data analysis was completed. The total eligible number of responses was 1,647. The resulting participants' demographics are reported in Table 2.

Table 2 Participants' demographic information

	Students' gender		Students' educational level	
	Female	**Male**	**Graduate**	**Undergraduate**
Total legitimate survey participation	923	724	219	1,428
	1,647			

Results

Research question 1

Were students aware that their learning behaviours in online settings can be monitored and recorded by their instructors?

Question 1 investigated students' awareness that their instructors could monitor and record most of their learning behaviours in online learning settings. To find the answer to this research question, the researchers added three sub-questions to the survey to ask about three aspects of learning activities, including students' login frequency, pages viewed or clicked, and learning profiles. In a typical LMS (a license- or subscription-based LMS such as Blackboard or Canvas, or an open-source LMS such as Moodle or Sakai CLE), course instructors could access these three aspects. Depending on course roles and institutional decisions (usually the responsibility of a department or division of information systems or educational technology), course instructors might have more or fewer access privileges to their students' data. However, within the scope of this research project, the researchers focused on only three standard accessible areas of students' data in their online courses. Those three sub-questions were presented in the form of a five-level Likert scale, a bipolar scaling method that measures positive and negative responses to statements with a range of: (1) completely aware, (2) aware, (3) neither aware nor not aware, (4) aware, (5) completely aware.

The results are reported in Table 3.

Table 3 Students' responses to whether they were aware that instructors could use their learning data for academic or research purposes

N	Statements	Mean	SD	Completely aware	Not completely aware
1	Are you aware that your learning activity (such as login frequency) in your online courses could be seen and recorded by your instructors?	1.50	1.32	1	5
2	Are you aware that your learning activity (such as pages viewed or clicked) in your online courses could be seen and recorded by your instructors?	1.75	1.95	1	5
3	Are you aware that your learning profiles in your online courses could be seen and recorded by your instructors?	2.73	1.82	1	5
Mean total		1.99			

As shown in Table 3, students were aware that their instructors could monitor and record learning behaviours in the three aspects of login frequency, pages viewed or clicked, and learning profiles. The researchers broke student responses down into gender and educational levels to examine whether there was any difference in their awareness between genders (male and female), and undergraduate and graduate levels by conducting an unpaired t-test. The results are reported in Table 4.

Table 4 Descriptive statistics about students' responses in terms of gender

		N	Mean	Std. deviation	Std. error mean
Gender	Male	724	1.97	1.37	1.25
	Female	923	2.00	1.85	1.14

An unpaired t-test was conducted to compare female and male students' responses to the first research question. The two-tailed P value equals 0.7153 with t = 0.37, df = 1645, and standard error of difference = 0.082. By conventional criteria, this difference is considered to be not statistically significant. In another words, these results suggest that there was no difference between genders in term of students' awareness of their learning behaviours being monitored and recorded by their instructors.

The researchers originally also planned to evaluate whether there was any significant difference between graduate- and undergraduate level students in terms of their awareness of their learning behaviours being monitored and recorded by their instructors. However, because there was a huge disparity in the total number of responses between the two cohorts of students (219 responses from graduate students and 1,428 from undergraduate students), the researchers did not examine this aspect of the research question.

Research question 2

Were students concerned about instructors using their learning data for academic or research purposes?

Question 2 examined whether students were concerned about the fact that their instructors could use their learning data for academic or research purposes. To answer this research question, the researchers included three sub-questions with three domains of students' data, including their login frequency, pages viewed or clicked, and learning profiles. The survey used a five-level Likert scale, measuring positive and negative responses to statements with the range of: (1) not really concerned, (2) not concerned, (3) neither not concerned nor concerned, (4) concerned, (5) really concerned.

The results are reported in Table 5.

Table 5 Students' responses to whether they were concerned about the potential for their instructors to use their learning for academic or research purposes

N	Statements	Mean	SD	Not really concerned	Really concerned
4	Are you concerned that most of your learning behaviours in your online courses (such as login frequency, page viewed and learning profile) can be monitored and recorded by your instructors?	2.10	1.55	1	5
5	Would you be concerned if your instructors collected your learning data in your online courses without revealing your personal information (name, gender . . . etc) for academic or research purposes?	2.45	1.75	1	5
6	Would you be concerned if your instructors collect your learning data in your online courses without revealing your personal information (name, gender . . . etc) for instructional/teaching improvement purposes?	2.15	1.25	1	5
Mean total		2.23			

The data in Table 5 indicates that students were quite neutral about their learning behaviours being monitored, recorded, or collected for academic or research purposes, and used for instructional/teaching improvement. The researchers divided students' responses into gender and educational levels as they did for Question 1, and conducted an unpaired t-test to examine whether there was any significant difference in student answers between male and female, and undergraduate and graduate levels. The results are reported in Table 6.

Table 6 Descriptive statistics about students' responses in terms of gender

		N	Mean	Std. deviation	Std. error mean
Gender	Male	724	2.14	1.95	1.50
	Female	923	2.32	2.35	1.75

After the completion of the unpaired t-test, a two-tailed P value equalled 0.097 with t = 1.66, df = 1645, and standard error of difference = 0.108. By conventional criteria, this difference is not considered to be statistically significant. In other words, the statistical values indicate no difference between male and female students in terms of their concern about their learning behaviours being monitored and recorded, being collected for academic or research purposes, and being used for instructional/teaching improvement efforts.

Discussion and conclusion

The results of this study have provided important insights into graduate and undergraduate students' attitudes towards data privacy issues in online learning environments in the U.S. Participants' responses to the online survey revealed that they were aware that their learning behaviours—such as login frequency, pages viewed or clicked, and learning profiles—could be monitored and recorded by their instructors. There was also no significant difference between genders in terms of students' awareness of their learning behaviours being monitored and recorded by their instructors. The researchers found that this was in line with previous studies about students' awareness of their privacy in the online learning environment (Doring, Hodge, & Heo, 2014; Lorenz, Sousa, & Tomberg, 2013; May, Fessakis, Dimitracopoulou, & George, 2012; May, Iksal & Usener, 2016; Yang & Wang, 2014). One possible interpretation is that, although students' ages were not known in this study, it may be assumed that most of them were from either the millennial generation or generation X, and were familiar with social media such as Facebook, Snapchat and/or Instagram, and therefore already knew about issues relating to users' data privacy (Gogus & Saygın, 2019).

The second finding of the study—that students were not concerned about potential use of their learning performance data by their instructors for teaching and/or research purposes—echoes recent studies about students' concerns about their privacy in online learning settings (Doring et al., 2014; Kokolakis, 2017). However, while previous research reported that gender could determine the extent to which students were concerned about their data privacy (Barak & Gluck-Ofri, 2007; Cockcroft & Clutterbuck, 2001; Petronio, 2002), this study found that gender was not a discriminating factor. It is suggested that future studies could explore the differences between the findings of this and studies reported earlier. In-depth studies using methods of observation are also recommended to extend our understanding of online learners' information-sharing preferences and actual practices. In addition, this study examined students' viewpoints on the potential use of their learning performance data only, by their instructors for teaching and/or research purposes. It did not examine responses to use by the university or institution for other non-commercial purposes. It would be interesting to find out whether students' perspectives are still the same.

One of the contributions of this study was to add the perspective of those students who attended an U.S. university into the literature. Previous studies surveyed students from Australia, China, Japan and France about this topic, but did not attempt to examine whether female and male

students had different perspectives. As mentioned by Cockcroft and Clutterbuck (2001), researchers in this area have identified a number of factors that influence individual attitudes to information privacy—such as gender, age, culture, socio-economic status and even country. Including U.S. student perspectives in the literature will help broaden our understanding about the issue of students' data privacy in online learning environments.

We might not be able to generalise the results of this study to the population as a whole due to its small and convenient sample size, but they are likely to be of interest to university educators, researchers, system developers and policy makers who are collecting, tracking, and analysing data that relates to students' learning performances for research and/or education. The study outcomes could also address researchers and advocates' concerns about online privacy (Black et al., 2008; Daniel, 2015; Lewis et al., 2008). It is argued that the practice of collecting and analysing students' performance data in online courses for instructional improvement and/or research or academic purposes should not be seen as a threat to students as long as (1) students are aware and informed of any tracking processes when they access their LMS, and (2) their personal identities will not be revealed. In other words, a compromise between tracking students and protecting their privacy is still important when collecting data.

References

Barak, A., & Gluck-Ofri, O. (2007). Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior, 10*, 407–417. doi:10.1089/cpb.2006.9938

Becker, S. A., Cummins, M., Davis, A., Freeman, A., Hall, C. G., & Ananthanarayanan, V. (2017). *NMC Horizon report: 2017 higher education edition* (pp. 1–60). The New Media Consortium.

Black, E. W., Dawson, K., & Priem, J. (2008). Data for free: Using LMS activity logs to measure community in online courses. *The Internet and Higher Education, 11*(2), 65–70.

Cockcroft, S., & Clutterbuck, P. (2001). Attitudes towards information privacy. *ACIS 2001 Proceedings*, 20. Retrieved from https://aisel.aisnet.org/acis2001/20/

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904–920.

Doring, A., Hodge, A., & Heo, M. (2014). Online learners and their self-disclosure preferences. *Journal of Information Technology Education: Research, 13*, 163–175. Retrieved from http://www.jite.org/documents/Vol13/JITEv13ResearchP163-175Doring0517.pdf

Drachsler, H., & Greller, W. (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* (pp. 89–98). ACM. doi: 10.1145/2883851.2883893

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society, 15*(3), 42–57.

Gogus, A., & Saygın, Y. (2019). Privacy perception and information technology utilization of high school students. *Heliyon*, *5*(5), e01614. https://doi.org/10.1016/j.heliyon.2019.e01614

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Lewis, K., Kaufman, J., & Christakis, N. (2008). The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, *14*(1), 79–100.

Lorenz, B., Sousa, S., & Tomberg, V. (2013). Privacy awareness of students and its impact on online learning participation: A case study. In T. Ley, M. Ruohonen, & A. Tatnall (Eds.), *Open and social technologies for networked learning* (pp. 189–192). Springer: Berlin, Heidelberg.

Mallery, P., & George, D. (2003). *SPSS for Windows step by step: A simple guide and reference*. Boston, MA: Allyn & Bacon.

May, M., Fessakis, G., Dimitracopoulou, A., & George, S. (2012). A study on user's perception in e-learning security and privacy issues. In *2012 IEEE 12th International Conference on Advanced Learning Technologies* (pp. 88–89). IEEE.

May, M., Iksal, S., & Usener, C. A. (2016). Learning tracking data analysis: How privacy issues affect student perception on e-learning? In *8th International Conference on Computer Supported Education (CSEDU) Vol 1* (pp. 154–161).

New Media Consortium Horizon Reports. (2016, 2017, 2018). Retrieved from https://library.educause.edu/search#?publicationandcollection_search=New%20Media%20Consortium%20(NMC)

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438–450.

Petronio, S. (2002). Boundaries of privacy: Dialectics of disclosure. New York, NY: State University of New York Press.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20.

Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7*(1), e1187. doi: 10.1002/widm.1187

Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, *32*(2), 143–159.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), 30.

Vu, P., Meyer, R., & Capero, J. (2016). Models of administration for online learning programs in the U.S. higher education institutions. *Journal of Applied Educational and Policy Research, 2*(1). Retrieved from https://journals.uncc.edu/jaepr/article/view/460

Watters, A. (2011). *Data science is a pipeline between academic disciplines*. Retrieved from http://radar.oreilly.com/2011/08/data-science-social-science-academic.html.

Yang, F., & Wang, S. (2014). Students' perception toward personal information and privacy disclosure in e-learning. *Turkish Online Journal of Educational Technology, 13*(1), 207–216.

Biographical notes

Phu Vu

vuph@unk.edu

Dr. Phu Vu is an associate professor in the Department of Teacher Education in the University of Nebraska at Kearney, U.S.A., where he teaches courses mainly in the Instructional Technology graduate program. His research interest is in game-based learning, learning analytics, and online learning.

Megan Adkins

adkinsmm@unk.edu

Dr. Adkins is an associate professor in the Department of Kinesiology and Sport Sciences in the University of Nebraska at Kearney, U.S.A. Her research focuses on teacher preparation; science, technology, engineering, math (STEM); and social emotional learning of underserved populations. She has completed numerous peer-reviewed articles and national presentations on these topics. Dr. Adkins teaches method courses in physical education. She focuses student preparation on experiential learning through a homeschool physical education teaching lab, STEM, and SEL after-school programming that she has developed.

Shelby Henderson

hendersonsj@lopers.unk.edu

Shelby Henderson is a graduate assistant in the Department of Teacher Education, and is a graduate student in Clinical Mental Health Counselling at University of Nebraska at Kearney, USA.