
Taking Student Data for Granted? A Multi-Stakeholder Privacy Analysis of a Learning Analytics System

Kaiwen Sun

School of Information
University of Michigan
Ann Arbor, MI 48109, USA
kwsun@umich.edu

Christopher Brooks

School of Information
University of Michigan
Ann Arbor, MI 48109, USA
broosch@umich.edu

Abraham H. Mhaidli

School of Information
University of Michigan
Ann Arbor, MI 48109, USA
mhaidli@umich.edu

Florian Schaub

School of Information
University of Michigan
Ann Arbor, MI 48109, USA
fschaub@umich.edu

Sonakshi Watel

School of Information
University of Michigan
Ann Arbor, MI 48109, USA
sonakshi@umich.edu

Abstract

Learning analytics systems measure, collect, and analyze learners' data, and provide reports or interventions based on the results in order to understand and optimize learning behaviors and outcomes. As learning analytics is becoming an integral part of higher education, it is essential to understand the privacy and ethics of the access, use, and analysis of learners' data. We conducted semi-structured interviews with different stakeholders of a specific learning analytics system in operation at the University of Michigan, called Student Explorer. We find that all stakeholders have various attitudes, perceptions, concerns, and needs toward privacy, indicating an existence of discrepancies and tensions due to stakeholders' different roles in the learning analytics process. Based on our findings we argue for better involvement, integration, and engagement of all stakeholders in learning analytics systems.

Author Keywords

Learning analytics; privacy; student data; higher education.

Introduction

Educational institutions these days are getting "smarter" by using learning analytics systems for various purposes and stakeholders: there are teaching systems for instructors [4]; learning platforms for students [11]; performance tracking tools for academic advisors [2]. These data-driven systems

are becoming part of educational experiences, hopefully generating positive learning outcomes. At the same time, associated privacy expectations, concerns, and needs of the stakeholders involved in the learning analytics process are not well studied. Moreover, learning analytics systems ultimately use massive amounts of learners' data, in which case the access, ethical use, benefits, issues, quality, and management of the data need to be considered carefully.

To understand different stakeholders' privacy perceptions of learning analytics systems and the ethics of using learners' data, we conducted 32 semi-structured interviews with three different groups of stakeholders at the University of Michigan across various departments on a system called Student Explorer [5]. Our findings indicate both consistent and inconsistent privacy attitudes and perceptions among stakeholders toward Student Explorer and the use of learners' data.

Related Work

Learning analytics systems use the data of students and it is crucial to analyze the expectation of students from the system. Students should play an active role in learning analytics systems. They can act as agents by collaborating and defining their own expectations from the system [9]. Ifenthaler & Schumacher [3] found that students are comfortable sharing data based on context. Students are willing to share course related data, like test scores and course enrollment details, but are not comfortable sharing personal information, such as online behavior, medical data, social media data etc., with learning analytics systems.

Privacy concerns alter student behavior. Dawson [1] found that students change their writing style when informed of institutional surveillance. Furthermore, students expect to have opt-in and opt-out options [10].

Prinsloo & Slate [7] propose ethics of care for learning analytics. These ethics aim at involving individuals and groups during their personal data collection process, providing them with knowledge about how their data is used, and granting them access to that data. Another recurring theme across research is maintaining transparency. This means that students should know how, why, with whom, and for what reasons their data is collected, used or shared [8]. Slate & Prinsloo [9] also underline that learning analytics is a moral practice focused on understanding not measuring.

Our research work will contribute to the literature by presenting qualitative research on privacy perceptions and ethics of data usage of three groups of stakeholders: developers, advisors and students, toward a learning analytics system at the University of Michigan.

What is Student Explorer?

We conducted our study in the context of Student Explorer [5], a learning analytics system deployed at the University of Michigan. Student Explorer is an early warning system that leverages Learning Management System data from Canvas (by Instructure) to create actionable intelligence that assists academic advisors in identifying students at risk of academic jeopardy in order to facilitate outreach to these students [6]. It helps advisors create timely interventions to direct students toward resources or behavioral changes that help facilitate future student success.

There are several groups of relevant stakeholders for Student Explorer: the *design and development team members* who built the system, *academic advisors* who are the primary users of the system, and all the *students* at the University of Michigan whose data is used in the system.

Study Design

We conducted 32 in-person semi-structured interviews between October 2017 and February 2018 with three groups of key stakeholders of Student Explorer: developers (4 male), advisors (4 female, 4 male), and students (9 female, 11 male) at the University of Michigan. All the interviews were audio recorded and lasted 35-45 minutes on average. Our study was approved by our IRB.

To gain diverse perspectives, we recruited academic advisors and students from multiple colleges and departments across the university. Particularly for student participants, we aimed to balance gender, school year, and majors. In recruitment, we did not mention that the focus of the study was privacy to reduce self-selection bias. Similarly, we began the interviews by asking them about their role, familiarity and use of Student Explorer, including privacy questions in the process. For student participants, we started the interviews by asking about their academic advising experiences, then introduced them to Student Explorer to learn about their perspectives.

Qualitative data analysis was conducted by three research team members, who worked together to identify preliminary themes and establish two codebooks. Both codebooks were collaboratively and iteratively refined through independent coding until we reached sufficient inter-rater reliability (Fleiss' $\kappa = .76$). One researcher then finished coding and re-coding the interviews.

Findings

Here, we summarize key findings in two areas: stakeholders' perception and understanding of the data usage in the learning analytics system and their privacy concerns and needs regarding the system.

Perception and Understanding of Data Usage

Most of the participants in both advisor and student participants groups had some sense that Student Explorer (SE) does not generate data on its own but either relies on data from a learning management system (called Canvas) or instructor input to calculate student performance. The tricky place with Canvas being the data source is that not every single class on campus uses Canvas, and different instructors use it differently depending on how they structure the classes and assignments. Under such circumstances, most of the participants from all three groups – including developers working on the system – believed the data in SE to be neither 100% accurate, nor a complete representation of students' academic performance. Developer participants acknowledged that the algorithm uses more of a 80:20 kind of process to calculate the learners' data, which is not a complete capture and calculation of all the data situations. Advisor and student participants stated that the data might be accurate in number but not in showing how students are doing, especially for classes that do not have frequent assignments, are graded based on feedbacks and comments rather than scores, do not update grades often, or have gameful learning elements, etc. Moreover, most of the advisor and student participants believed students' academic performance was more than numbers and graphs, pointing out that there are a lot of other factors to consider that are less quantifiable, such as classroom engagement, written evaluation, career interest, personal goals, soft skill evaluation, impacts from students' personal life, etc. None of the advisor participants fully trusted the data in SE and they mentioned using the data with a "grain of salt" or as a "conversation starter" during advising session with students to find out more about how they were doing.

Privacy Concerns and Needs

According to developer participants, SE is not a student-facing system, and its primary users are advisors. Consequently, none of the 20 student participants had ever heard or seen SE before the interview, nor had they been informed of their data being used in SE. Meanwhile, as the user of SE, some advisor participants did not know if students knew about SE, or whether students had been informed of their data being used in SE. When asked if students should be informed that SE is using their data, some developer and advisor participants acknowledged students' right to know how their data is used; regarding student consent, they considered it to be "a part of a broader conversation in the university."

For student participants, there were three different perspectives: most of them expressed that students should be informed and should have a say when SE uses their data. Students should have options such as being asked for permission, having an opt-out option of the learning analytics system, being informed what and when systems are using students' data and how it is used, have the choice to share their data with certain stakeholders (administrators, management teams, development teams) or not, and having their data be anonymous in certain contexts. A few student participants preferred not to have a say of their data because they believed learning analytics systems to be purposeful, useful and helpful in the educational context; to ensure a complete data environment and maximize the overall benefits, all students should participate in it. The rest of the student participants either did not care as much or were unsure because they considered the university as a "big government" who is in control of student data anyway, so the students have gotten used to following the university's arrangements and considered it a hassle if asking every single student for data usage permission.

Overall, each group of stakeholders has their unique perceptions and experiences with Student Explorer: the developers present themselves as good stewards of the learners' data, and they understand the importance of addressing learners' data privacy in the learning analytics process, but ultimately this topic should involve higher management teams at the university for a broader discussion. Although the advisors are the primary users of Student Explorer, they stay very focused in their advising role not knowing much about how Student Explorer works with learners' data; they acknowledge the students' right to be informed on how their data is used, but could not draw a conclusion on student-opt-out option. Meanwhile, advisors haven't considered themselves taking an active role to help address privacy considerations in learning analytics systems. As for students, while they are aware the university is constantly using their data, none of them had been informed that their data was being used in Student Explorer; most of the students expressed they have the rights over how their data is used, so the university should provide an opt-out option or ask students for permission.

Discussion and Future Work

This work reveals privacy perceptions of three groups of key stakeholders regarding a learning analytics system, identifying both tensions and agreements among stakeholders regarding the use of learners' data in this system. Our findings can inform the design of learning analytics systems, in particular with respect to privacy protections and needs for transparency. Depending on the context, additional stakeholders' perspectives, such as instructors and administrators, should also be considered in the design of learning analytics systems.

Acknowledgements

This research was partially funded by the University of Michigan's Office of Academic Innovation and School of Information.

References

1. Shane P Dawson. 2006. The impact of institutional surveillance technologies on student behaviour. *Surveillance & Society* 4, 1/2 (2006), 69–84.
2. Anna Lea Dyckhoff, Dennis Zielke, Mareike Bültmann, Mohamed Amine Chatti, and Ulrik Schroeder. 2012. Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society* 15, 3 (2012), 58.
3. Dirk Ifenthaler and Clara Schumacher. 2016. Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development* 64, 5 (2016), 923–938.
4. Andrew E Krumm, R Joseph Waddington, Stephanie D Teasley, and Steven Lonn. 2014. A learning management system-based early warning system for academic advising in undergraduate engineering. In *Learning Analytics*. Springer, 103–119.
5. Steven Lonn and Stephanie D. Teasley. 2014. Student Explorer: A Tool for Supporting Academic Advising at Scale. In *Proceedings of the First ACM Conference on Learning @ Scale (L@S '14)*. ACM, 175–176.
6. Leah P Macfadyen and Shane Dawson. 2010. Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education* 54, 2 (2010), 588–599.
7. Paul Prinsloo and Sharon Slade. 2014. Student data privacy and institutional accountability in an age of surveillance. In *Using Data to Improve Higher Education*. Springer, 197–214.
8. Paul Prinsloo and Sharon Slade. 2015. Student privacy self-management: implications for learning analytics. In *Proceedings of the fifth international conference on learning analytics and knowledge*. ACM, 83–92.
9. Sharon Slade and Paul Prinsloo. 2013. Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist* 57, 10 (2013), 1510–1529.
10. Sharon Slade and Paul Prinsloo. 2015. Student perspectives on the use of their data: Between intrusion, surveillance and care. *European Journal of Open, Distance and E-learning* 18, 1 (2015).
11. George Veletsianos, Royce Kimmons, and Karen D French. 2013. Instructor experiences with a social networking site in a higher education setting: Expectations, frustrations, appropriation, and compartmentalization. *Educational Technology Research and Development* 61, 2 (2013), 255–278.