

Defining “data” in conversations with students about the ethical use of learning analytics

Abi Brooker

University of Melbourne

Linda Corrin

University of Melbourne

Josie Fisher

University of New England

Negin Mirriahi

University of South Australia

In any conversation about the development of ethical standards for practice, it is vital that all stakeholders have a shared understanding of the main concepts in order to reach agreement. In the context of higher education and learning analytics, while many conversations are underway, it is less clear that such a shared understanding exists around the concept of “data”. In order to understand this situation more fully we conducted a study to investigate students’ perceptions of the ethical and privacy considerations related to the data that universities collect and use about them for the purposes of learning analytics. In this paper, we focus specifically on the understandings students have of the types of data that can be collected about them within the educational environment. The outcomes showed that there was a diversity of understandings, but that five main data types emerged. In developing a better understanding of the ways students understand data, it can assist institutions to have more effective conversations with students about the ethical use of learning analytics.

Introduction

The growing development of teacher- and student-facing learning analytics systems has prompted new discussions around the ethical use of student data in higher education. Specifically, the innovative nature of learning analytics and rapid increase in variety of student data being used means that new and complex questions are emerging for institutions about appropriate use of those data. In an age of big data and decision making based on ever-increasingly sophisticated algorithms it is not always clear how such uses fit within existing legal and ethical frameworks. Ensuring the ethical use of student data in this environment requires discussions involving all stakeholders in the implementation of learning analytics systems. However, it is only relatively recently that the student voice has been added to these conversations.

In this paper, we report on the initial outcomes of a study conducted across two Australian universities on students’ perceptions of the use of their data for learning analytics. In particular, we explore the understanding that students have of what is meant by the word “data” in the context of their educational experience in universities. The outcomes show that there is a diversity of understandings, what data is and what is actually

collected in the learning context. This research is important to inform the ways that we frame our conversations with students about data and the ethical considerations surrounding the use of such data by universities.

Background

As learning analytics initiatives gain momentum in the higher education sector, institutions are investing in technological approaches that collect, aggregate and utilise various data collected about students, through processes including enrolment and their use of institutionally-hosted learning technologies. Such use has complex implications for students in relation to ownership, reciprocity, privacy, and transparency of data. Within this environment, it is critical to consider students’ perspective on the collection and use of data pertaining to them. However, to date there are few studies that have included students in the conversation about the data they are willing to share and their understanding about how data are used.

Despite the ubiquity of data mining in everyday online contexts (e.g., social networking sites displaying personalised ads based on users’ Internet or Search



This work is made available under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) licence.

Engine history or location), it is not evident whether students are aware of the extent of data collection and data mining occurring in educational settings (Slade & Prinsloo, 2013). In addition, Crawford and Schultz (2013) posit that when institutions make decisions based on a person's data, the person has the right to question how their data informed that decision. For students to give informed consent for their educational institutions to collect and use their data, they should know what data is collected, its source, how it will be used, whether it will be shared with third parties, and how students' identity will be preserved (Slade & Prinsloo, 2013; Sclater, 2014).

To date, there have been limited studies that investigate students' understanding of the types of data collected of them and any issues they have concerning their privacy (Drachler et al. 2015). A comparative study involving students from the USA and UK found that the majority of students in both countries were happy for their data to be used to help improve their grades, although the percentage in agreement was smaller in the UK groups (Arnold & Sclater, 2017). This could be attributed to the fact that the US students had already been exposed to learning analytics tools in their institutions and therefore could more easily understand the type of data included and the benefit they can receive from such systems, whereas students in the UK did not have the same experience. In Germany, a study of 330 students found that students had mixed views about the data they were comfortable allowing learning analytics systems to use (Ifenthaler & Schumacher, 2016). While they were happy for data such as grades and course enrolment to be used, they didn't want log trail/clickstream data or personal information to be made available for learning analytics systems. In the Australian context, Roberts and colleagues (2016) investigated students' attitudes towards the use of their 'big data' for learning analytics purposes at one institution. Through focus groups with 41 undergraduate and graduate students it was found that students had limited knowledge about educational data and learning analytics. Further, despite recognising that the data could personalise their educational experience, students were concerned about invasions of their privacy and that data would be used without their informed consent (Roberts et al., 2016).

Consistent across these three studies is the fact that students were given a definition of data and learning analytics at the start of the research and, in the case of the German and Australian studies, also exposed to examples of learning analytics systems. Providing such definitions and examples provides a similar baseline for all student participants' understanding in order to respond to certain questions in an 'informed' manner. However, as a result, it also obscures any understanding of students' previously held definitions of data as well as their knowledge of what data is collected and used about them by educational institutions. This is important to consider

as students often don't receive such a formal introduction to learning analytics and the scope of data collection before having to give consent to policies governing the use of their data for learning analytics systems.

What is 'data'?

The Oxford English Dictionary defines data as "facts and statistics collected together for reference or analysis" ("Data", 2017). In the context of education there are many facts and statistics that could potentially be collected about a student as they move through their degree. From the information that students provide in order to secure entry to the institution through to the information they give as they leave the institution and become an alumnus, the scope of data within higher education can be extensive. Added to this is the ever-increasing ability for data to be collected on students' activities in a multitude of online learning systems. The challenge that faces institutions implementing learning analytics is for students to be able to appreciate and understand the range of types of data that can be collected about them and how such data can be used. Such usage could be at the level of an individual or anonymised and aggregated to provide broader understandings of trends across student groups.

Typically, the statements written into student charters or statements on student privacy about the types of data that are collected and used are quite vague. This can include statements as broad as "data used in teaching and learning", "data for the provision of student services" or "personal information ... (collected) for a number of purposes". Although accurate, these broad definitions do not make clear to students the detail of the exact data that are collected or how they can be used across various university contexts. In contrast, in their *Policy on ethical use of student data for learning analytics*, the Open University (OU) in the UK provide a more specific definition for data. They explain that "data used for learning analytics typically falls into one of two categories: that captured at registration or at later points as a result of the student supplying information to the University (typically labelled as Student characteristic data), and that derived from ways in which the student engages with University systems as a result of their ongoing study (typically summarised as Study behaviour data)" (Open University, 2014, p.3). In the Australian context, Charles Sturt University (CSU) use a very similar definition to the OU, but add that data can also be collected from "information we are authorised to collect from other organisations (e.g. government agencies)" (Charles Sturt University, 2015a, p.1). As part of CSU's Learning Analytics Code of Practice the detail on the exact data is extended through the principle that "All users of the University's learning and teaching systems will have access to clear explanations of their rights and obligations with respect

to data from those systems” (Charles Sturt University, 2015b, p.7).

The current study

The current study provided an opportunity for participants to share and collaboratively discuss their knowledge of the term “data” and what they perceive the university collects and potentially uses about them. The study was driven by the research question: What do students understand about what, how and why their data is collected and used in higher education? In this paper, we focus on the first part of this question which considers what data students know and/or think is being collected and used about them. This enabled us to explore the ways students understand and define “data” within the higher education context. This is important to help maximise the effectiveness of conversations with students about learning analytics implementation and to use when designing institutional policies and procedures that provide an ethical environment for learning analytics.

Method

Six student focus groups were conducted at two Australian universities: four focus groups were conducted at institution A (a metropolitan university in Victoria) and two focus groups were conducted at institution B (a regional university in NSW). A convenience sampling approach was taken to recruit participants with participation open to all students at the two institutions. Each focus group included between 5 and 10 participants. Participants in the focus groups represented a range of academic disciplines (e.g., Nursing, Science, Engineering, Education, Psychology, Arts), year levels (undergraduate/postgraduate), genders and enrolment types (domestic/international). The focus group sessions were structured around three main discussion questions: (1) What data do you think the University collects about you? (2) What do you think the University uses this data for? and (3) What (if any) are the responsibilities of the University, when using your data?. The focus group facilitators did not give a definition of data at the beginning of the sessions, instead they allowed this to emerge throughout the discussion between student participants. The facilitators also did not use the term “learning analytics” as it was felt that students may not be familiar with this term and that giving a definition for this may influence students understanding of the definition of data. Where possible, the facilitators tried to control for bias that often emerge in focus group situations, such as dominance bias, by directly addressing quieter participants. A thematic analysis of the focus group transcripts was conducted to identify the main issues raised by participants.

Findings

Students’ understandings of the data collected by universities were varied and often lacked certainty. There

were five broad types of data that students recognised that the University collected about them. In some focus groups, these types of data emerged organically as part of the students’ discussion (e.g., “What we do in Moodle”). In others, they emerged in response to the facilitators’ prompting when it was clear that students were struggling to think of data beyond that explicitly given (e.g., “what about when you come to the library?”). The five types of data were:

1. Personal information

The most frequently identified type of data across all focus groups related to personal details provided by the student to the institution during enrolment, examinations, and other schemes such as scholarships or student support. This personal information included health information (“vaccination records”, “details of health care”), identifiable information (“our photos”, “address, emergency contact, phone number”), and official documentation (“police checks”, “my birth certificate”). Ambiguity around this type of data emerged in various contexts where data were collected. For instance, international students disagreed with each other about whether information about their financial circumstances were collected as part of their enrolment or as part of their visa application process. There was also uncertainty around whether and how personal, identifiable information was connected to other data sources.

2. Online activity

Students agreed that their university collected data about their use of online learning management systems and university-lead social media platforms (e.g., “What we are doing on Moodle”, “online revision tasks or just opening, reading the lecture slides”). They were less sure about the nature or detail of the data that was collected in these forums. Some students believed that every element of online work was collected, analysed and used by university services (e.g., “they check, I suppose, whether we’re accessing the test at the same time as someone else and giving similar answers”); whereas others believed that online activity were automatically collected into a databank but were not directly monitored (e.g., “it just like automatically collects in their database”).

3. Student feedback

In two discussion groups, there was strong consensus among student participants that the main data collected by their university was through feedback from students. They felt they provided a lot of feedback (e.g., “feedback for everything pretty much”). When prompted, they described providing feedback about their enrolment experiences, orientation week, their subjects, online support services, academic support services, and in research projects such as the current project. An ambiguity around providing feedback was whether and

how feedback offered voluntarily was connected to students' personal information. Some students were certain that feedback was anonymous (e.g., "they say we remain anonymous, yeah"), others were sure that it was identifiable (e.g., "actually they have our names ... because where you log in the LMS") and others were not sure (e.g., "I'm not sure if it's anonymous or not because it's linked to our LMS"). This was especially the case when students provided feedback online via their learning management system. Consequently, some students said they restricted feedback in order to avoid being targeted by their teachers. This was reflected in the following student's reservations about providing qualitative feedback in the online Student Experience Survey:

"I think maybe a lot of [students] would actually be reluctant to put any further comments ... [they] might think "oh that might get a bit personal" or "so-and-so might recognise me" so I'll just do the numbers [rating scales]. So that might actually count against that input and improving things."

4. Academic information

Academic information collected about the student included academic history (e.g., "transcripts", "previous study", "what school we came from") and information about current academic progress (e.g., "what subjects we're doing and our results"). Some students also identified the assignments themselves as data, and described instances where teachers had used their assignments as examples to show other cohorts of students, along with the grade and feedback. Students were comfortable with the way that universities collected and used their academic information, although they were unsure who had access to that data, for instance, whether teachers in other subjects could see their grades.

5. Resource usage

Students recognised that universities collected information about the resources and infrastructure that students use on campus. Often discussions around resource usage were initiated by the facilitator (e.g., "What about around campus?" "What about in the library?"). Most students readily identified activities with log recording such as using services accessed by their student ID cards (e.g., "using photocopiers and printers", "what books we borrow"). Only a few students identified university logs of location/usage of Wi-Fi networks as another form of data (e.g., "when you log into uni wireless", "when we connect to the Internet"). The students' discussions around these data logs were not straightforward. Some students were not sure about how or why log data was collected (e.g., "browsing on the uni WiFi, I am not sure to what extent that's monitored"); whereas others knew data were collected, but were not sure why (e.g., "they won't be interested in looking at my browser history, but like they do gain access to it"). Some

felt that the data logs were for censorship purposes (e.g., "they block things like Peer-to-Peer [software] so you can't pirate"), whereas others believed the data logs had no effect on their behaviour (e.g., "as long as you're not doing anything [wrong] you shouldn't, no reason to be worried really").

Discussion and conclusion

The outcomes of the focus groups offer a more nuanced understanding of what students understand about the various data collected about them. Students didn't always easily or readily come up with these data definitions - for some groups it took time. Sometimes, there was strong consensus about the types of data collected in different contexts (e.g., personal information given during enrolment). Although there was often less certainty about how some of this data was used in practice (e.g., log data from resources usage or their online activity on the LMS). There was also confusion about whether some forms of data could or should be collected and used. Whether and to what extent activity traces were collected through online learning systems prompted a variety of views across the focus groups. Occasionally, the students came to consensus on these understandings, but other times no such consensus was found. The ability for students to come to a consensus (or not) was a feature of the focus group environment and highlights that, in the current climate, the understanding of data by individual students may remain quite varied. With such variety within this small sample, we anticipate that further large-scale studies of students' perceptions will reveal even greater diversity in their understanding.

This work-in-progress paper reports the emerging themes from the first question of a broader study. As we investigate the rest of students' discussions, we anticipate that the types of interaction between them will continue to shape their understanding of more complex issues around the ways that universities use their data. However, the diversity of their understanding suggests that ongoing discussions with students about this issue need stronger clarification of data. This is to ensure that when students, teachers and administrators are engaging in more in-depth discussions about the best and most ethical ways to use data, that there is a shared understanding of what that data is.

References

Arnold, K. E., & Sclater, N. (2017). Student perceptions of their privacy in learning analytics applications. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 66-69). ACM.

Charles Sturt University (2015a) *Statement of Student Data Rights and Responsibilities*. Retrieved from https://www.csu.edu.au/_data/assets/pdf_file/0010/2160487/Statement-of-Student-Data-Rights-and-Responsibilities_final.pdf.

Charles Sturt University (2015b) *CSU Learning Analytics Code of Practice*. Retrieved from http://www.csu.edu.au/_data/assets/pdf_file/0007/2160484/2016_CSU_LearningAnalyticsCodePractice.pdf.

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *BCL Rev.*, 55, 93-128.

Data. (2017). In *English Oxford Living Dictionaries*. Oxford University Press. Retrieved from: <https://en.oxforddictionaries.com/definition/data>.

Drachsler, H., Hoel, T., Scheffel, M., Kismihók, G., Berg, A., Ferguson, R., ... & Manderveld, J. (2015, March). Ethical and privacy issues in the application of learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 390-391). ACM.

Open University (2014). *Policy on ethical use of student data for learning analytics*. Retrieved from <https://learn3.open.ac.uk/mod/url/view.php?id=85812>.

Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student Attitudes toward Learning Analytics in Higher Education: "The Fitbit Version of the Learning World". *Frontiers in Psychology*, 7, 1959. <http://doi.org/10.3389/fpsyg.2016.01959>.

Sclater, N. (2014). Code of practice for learning analytics: A literature review of the ethical and legal issues (pp. 1-64). Jisc. Retrieved from http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A- Literature_Review.pdf.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510-1529.

Contact author: Abi Brooker
brookera@unimelb.edu.au

Please cite as: Brooker, A., Corrin, L., Mirriahi, N. & Fisher, J. (2017). Defining "data" in conversations with students about the ethical use of learning analytics. In H. Partridge, K Davis, & J. Thomas. (Eds.), *Me, Us, IT! Proceedings ASCILITE2017: 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 27-31).

Note: All published papers are refereed, having undergone a double-blind peer-review process.