Learning Analytics, Ethics and Participant Privacy

Dr. Brid Lane

IBAT College Dublin & Dublin City University (DCU), Ireland brid.lane@dcu.ie

Dr Eamon Costello

Dublin City University (DCU), Ireland eamon.costello@dcu.ie

Abstract

The field of Learning Analytics' ultimate promise is that it can turn insights into action so as to improve learning experiences and outcomes. Efforts have been made to consider the ethical implications of this field such that learners are properly protected in such research. However, surprisingly little research has been conducted into how researchers in the field actually handle issues of ethics and participant rights as evidenced in published studies. In this paper we address this gap by conducting a literature review and analysis of Learning Analytics research. We analysed 104 papers and found that a majority (60) made no mention of ethics or ethical approval for the research. We found differences in opt-in and opt-out policies for learners in the published studies. It was not always clear that full learner consent had been given in many studies. We highlight considerable absence of details in ethical reporting and recommend that future learning analytics research studies clearly detail the ethical activities followed.

Keywords: learning analytics, ethics, student privacy.

1. Introduction

Learning analytics collects data on users while they are engaged in the learning process and then converts these into actionable insights that when applied appropriately can improve that learning. They are useful in predictive analytics detecting if a student is not progressing as they should, and thereby boosting retention, and in tackling student disengagement. While there is no doubt that there is some value in the multitude of empirical research that has emerged, there is a need to consider the position of the students themselves. The digital footprints that students leave behind to inform the learning analytics data can be very large. This paper asks to what extent researchers have detailed their activities and processes in handling of student data. As researchers and teachers it is incumbent on us to ensure we are taking due consideration of our students right to have their data properly protected, their privacy respected, and their consent sought if their personal data is to be used in the provision of their education.

More formally, ethical responsibility might reasonably include ensuring students are aware that their data is being collected, how it is being used to further their education, and who has access to it, and that they agree with all this (Garaizar & Guenaga, 2014). The need for an ethical stance is widely proclaimed in the written literature. Sharon Slade and Paul Prinsloo have written extensively on ethics in learning analytics. In Prinsloo & Slade, 2013, they criticised Higher Education institutions in general for not going far enough. The suggestion is that existing policy frameworks tend to reflect what is legally required in terms of data privacy but might not be adequate in terms of the ethical challenges inherent in capturing and using student data in learning analytics activities. They compared ethical practices concerning student data across three MOOC providers (Coursera, EdX and Futurelearn) in Prinsloo & Slade, (2013), concluding that it is all too easy to become over-enthused with what learning analytics can offer, and then neglect the due attention required on details such as student consent.

They later highlighted that ethical concerns are too often not addressed in learning analytics activities in educational institutions, and offer a typology of sorts to address this (Willis, Slade & Prinsloo, 2016). This was supplemented by their 2017 piece, interestingly titled "An elephant in the learning analytics room – the obligation to act" which detailed ethical obligations shared between students and their chosen institution of study, and how they might be tackled (Prinsloo & Slade, 2017).

There is sometimes an unwritten assumption that students want to share their data with educational institution. This might not be the case. Addressing shared obligations is a most difficult task indeed if one party is unwilling to share their data, or there are unknowns on either side. Privacy policies around learning analytics are challenged when students are unaware of their privacy rights. The problem of education institutions being unaware of their obligations should be much reduced (at least in Europe) with the implementation of the Generalised Data Protection Regulation (GDPR). Legal obligations are likely an influencing factor in the design of learning analytics activities as educators and researchers try to balance privacy, pedagogy and technical development in their learning analytics endeavours (Hoel, Griffiths & Chen, 2017). It is a difficult balancing act, there are trade-offs even within the privacy mechanisms employed and no single technical solution to addressing the problem (Gursoy, Inan, Ecran-Nergis & Saygin, 2017).

1. Research Focus

We have outlined the importance of ethical aspects in learning analytics and the difficulties in ensuring they are sufficiently tackled. Nonetheless, there is a need to investigate if researchers in learning analytics are giving due consideration and attention to the ethical issues in their studies.

As one part of their research, Rebecca Ferguson and Doug Clow at the 2017 LAK Conference searched for evidence for quality in learning analytics research (Ferguson & Buckingham-Shum, 2012). They searched for the stem 'ethic-' in the 22 papers of the "Higher Education in the LACE Evidence Hub", and found that just three had explicitly considered ethics. This is surprising given that this hub (Learning Analytics Community Exchange, 2018) deals exclusively with learning analytics. Our research expands on this by taking a more in-depth search through a much wider corpus of research papers, across a range of academic journals. Our overarching research objective was to determine what information (if any) is given in empirical published research in the field of LA about the handling and management of student personal data. Specifically, we sought to determine if studies reported having received institutional approval from an ethical review board, whether they mentioned informed consent and how they treated anonymization of data.

2. Method

We adopted a systematic literature review approach (Petticrew & Roberts, 2009), (Okoli, 2015) for this study. Systematic literature reviews can be seen as a rigorous approach to analysing a large but distinct corpus of literature on a topic. They give an overview of that topic, can identify weaknesses or gaps in understanding, give a structured approach that may help mitigate potential researcher bias, and as they are clearly documented and are potentially more reproducible.

A systematic literature review first specifies a set of inclusion criteria to describe the topic (Okoli, 2015). The inclusion rules here specified that studies meet five criteria: (1) be empirical research i.e. we excluded theoretical, commentary or work in progress pieces (2) be clearly related to learning analytics research, (3) be written in English, (4) be published between 2016 and 2017 inclusive, (5) be published in a peer reviewed journal or conference proceedings.

We chose to use the Scopus database as our search tool. One of the advantages of using Scopus as as source over others is the reliable metadata it contains about articles (Dawson, Gašević, Siemens & Joksimovic, 2014). In order to be indexed by Scopus journal articles and conference proceedings must meet various criteria related to the research quality, including clear peer review policies, editorial board appointments, ethics policies, plagiarism checks etc. All of the major sources of MOOC related research are indexed in Scopus such as relevant ACM proceedings (Learning Analytics and Knowledge, Learning @ Scale), Springer Lecture Notes in Computer Science (European MOOCs Stakeholder Summit Proceedings; European Conference on E-learning) IEEE proceedings (Learning with MOOCS Conference). Scopus also has a very wide coverage of the relevant journals in the field.

Following a trial of various search terms, we developed a search query comprised of the key term "learning analytics" and common variations thereof ("metrics", "learning metrics", "learning prediction") to search for the relevant papers within Scopus. The abstracts of these papers were then examined to determine their inclusion in the final corpus. The search terms used to examine these papers to screen for inclusion were as per Table 1 below.

Table 1. Search terms used to screen papers for inclusion.

Stem	Variations included
Ethic	Ethics, Ethical
Priv	Privacy, Private
Permi	Permission, Permitted
Consent	Informed Consent
Anon	Anonymity, Anonymous, Anonymised, Anonymized
Pseud	Pseudonym
Transpar	Transparency, Transparent
Legal	
Law	
GDPR	

3. Results

Similar to Dawson et al (2014) we found a large number of conceptual papers as well as the empirical research pieces. Of these papers that we found, those that didn't implement empirical research in LA involving student data were split between position papers and other opinion-style pieces, literature reviews, editorials, technology proposals, framework proposals, panel discussions, workshop reports, and those whose primary data was not learning analytics data, but rather some other data e.g. survey data.

The screening criteria to select the papers, and the subsequent searching within them returned 130 papers. Of the 53 empirical papers for 2016, 42 were available in full-text to us. Of the 77 empirical papers for 2017, 62 were available to us in full-text. The total number of papers suitable was therefore 104.

60 of the papers made no mention of any of our stem terms. There could be many explanations for this. It is possible that nothing at all was done to address the ethical issues involved. Alternatively, due consideration might have been given but the authors considered ethics as of lesser importance and thereby chose not to include them. They may have been taken care of by institutional procedures separate to the given research and thereby not directly in the domain of the researchers. Whichever,

it is nonetheless concerning for those reading these papers to see no details at all given on how students data was handled and managed from an ethics perspective.

Just 9 of the research papers talked about consulting students about the research that involved their data. All but one of these reported seeking and receiving informed consent from their respondents. One merely informed students that their data was to be included. In this particular case, the researchers made a point of saying their institution does not allow students to opt-out of having their data used for learning analytics research. It should be noted that this was the only one of the papers that went into detail in the explanations given to their students on exactly how their data was used for learning purposes.

The stem for "pseud" returned only three papers. 21 papers (including one of the "pseud" papers) reported they anonymised either the student names or their data that could identify them. While most just mentioned that anonymisation was done, some were quite specific on how this was done, pointing to the algorithms that were used in the anonymisation process. One considered the student's ID Number to be a sufficient anonymiser.

There was specific mention of obtaining institutional approval for the research in 7 cases. These varied across the institutions' ethical policy process, ethical committee, privacy process, institutional review board, and human ethics committee. Some merely mentioned this approval had been sought. One, dealing with students studying a medical domain, gave more specific details of what was involved. Health data is considered sensitive personal data and so it is to be reasonably expected that this domain would be detailed in their handling of any personal data, including student data.

4. Discussion & Conclusion

From the above results, it appears that more could be done by researchers in learning analytics in their handling of student data. Under the EU GDPR, students have a right to know what data is being collected on them and how it is being used. A large number of the papers we analysed did not say that they did this. There is a risk then of the possibility of operating without due consideration for the legal requirements and just the ethical ones.

Following (Dawson, Gašević, Siemens & Joksimovic, 2014) we are not claiming that the research we looked at was necessarily actually unethical. Rather, we are noting that the authors gave too little and in many cases no information at all, about what ethical activities or procedures they followed. Yet, as per Colledge (2014) we recognise that there needs to be recognition that balancing privacy and ensuring good quality usable data is a trade-off. It is arguably too strict to insist students allow their data to be used in learning analytics, and so an opt-out is needed. At any rate, legislation such as the GDPR specifically requires that data subjects specifically opt-in to having their data collected, and are unlikely to tolerate any coercion on students to share their data. Yet, without this student data, the value arising out of learning analytics is greatly reduced. Similarly, students need to be made aware that their activities are being recorded, why it is being recorded and how it will be used. However, this is not an easy task to do clearly and with precision, so that students feel in control of their privacy without overwhelming them with privacy policies and procedures [LAK, 2017], such that they opt-out of sharing that data.

Across all the publications we looked at, just one talked at length about legislation. Garaizar and Guenaga (2014) discussed movement towards increased regulation in data practices, citing the Europe Directive 95/46/EC, extended with directive EUP 2002. Publishing in 2014, they were too early for the GDPR and the promises that holds. Nonetheless, they expressed concern that technology and the uses it is put to simply changes too fast for the law to keep up. As a result, they recommend that the

learning analytics community should self-regulate, calling for a framework to guide ethical and privacy-related issues in learning analytics research to help in doing this, and thereby build confidence among the various parties involved in LA. Slade and Prinsloo (Willis, Slade & Prinsloo, 2016) and (Prinsloo & Slade, 2017) have created such a framework however we have shown here that despite such tools there is an under-reporting of ethical issues in the literature.

Several possibilities exist for furthering the research. One dimension that could be added to continue the results of this paper concerns the geographical provenience of the authors of the reviewed papers, looking at any relevant difference in how authors take into consideration ethical issues in learning analytics based on their geographical provenience. Other possibilities are to compare different forms of learning analytics, the contexts in which learning analytics are used, varying types and formats of student consent.

This study makes a contribution to the debate around ethical issues in Learning Analytics by highlighting the absence of ethical treatment in study design as evidenced in the published literature. Our recommendation is that research funders and editors of published papers put processes in place to make reporting of ethical issues standard in such research.

5. References

Colledge, L. (2014). Snowball metrics recipe book. *Amsterdam: Snowball Metrics Program Partners, 110*.
Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current State and Future Trends: A Citation Network Analysis of the Learning Analytics Field. *LAK 2014.* Indianapolis, IN, USA.
Ferguson, R., & Buckingham-Shum, S. (2012). Social Learning Analytics: Five Approaches. *2nd International Conference on Learning Analytics & Knowledge (LAK).* Vancouver, BC. Canada. Retrieved from http://oro.open.ac.uk/32910
Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. *LAK '17.* Vancouver, BC, Canada.
Garaizar, P., & Guenaga, M. (2014). A Multimodal Learning Analytics View Of HTML5 APIs: Technical Benefits And Privacy Risks. *Technological Ecosystems for Enhancing Multiculturality (TEEM).* Salamanca, Spain.
Gursoy, M. E., Inan, A., Ercan Nergiz, M., & Saygin, Y. (2017). Privacy-Preserving Learning Analytics: Challenges And Techniques. *IEEE Transactions on Learning Technologies, 10(1)*, 68-81.
Hoel, T., Griffiths, D., & Chen, W. (2017). The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems. *LAK '17.* Vancouver, BC, Canada.
Learning Analytics and Knowledge *(LAK) Conference (2017).* Retrieved September 29, 2019, from http://lak17.solaresearch.org
Learning Analytics Community Exchange (LACE) (2019): http://evidence.laceproject.eu. Retrieved September 29, 2019
Okoli, C. (2015). A Guide To Conducting A Standalone Systematic Literature Review. *Communications of the Association for Information Systems, 37.* Retrieved from https://search.proquest.com/docview/1763749504.
Petticrew, M., & Roberts. (2009). *Systematic Reviews in the Social Sciences.* Malden, MA, USA: Blackwell.
Prinsloo, P., & Slade, S. (2013). An Evaluation Of Policy Frameworks For Addressing Ethical Considerations. *Third International Conference on Learning Analytics and Knowledge Learning Analytics (LAK).* Leuven, Belgium.
Prinsloo, P., & Slade, S. (2015). Student Privacy Self-Management: Implications For Learning Analytics. *The Fifth International Conference on Learning Analytics and Knowledge (LAK).* Poughkeepsie, New York. Retrieved from https://pdfs.semanticscholar.org/09d7/56d7a66f002f5c06b05237c3fc162b61a653.pdf
Prinsloo, P., & Slade, S. (2017). An Elephant In The Learning Analytics Room - The Obligation To Act. *LAK '17.* Vancouver, BC, Canada.
Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. *Education Technology Research Development*, 881-901.