Balancing Privacy and Open Science in the Context of COVID-19: A Response to Ifenthaler & Schumacher (2016)

Joshua Rosenberg[1] and K. Bret Staudt Willet[2]

[1]*University of Tennessee, Knoxville*
[2]*Michigan State University*

Abstract

Privacy and confidentiality are core considerations in education. At the same time, using and sharing data—and, more broadly, open science—are increasingly valued by editors, funding agencies, and the public, and learning analytics systems depend upon ready access to student data, the use of which might infringe upon the rights of students. This manuscript responds to an empirical investigation of students' perceptions of the use of their data in learning analytics systems by Ifentahler and Schumacher (2016). We summarize their work in the context of the COVID-19 pandemic and the resulting shift to digital modes of teaching and learning by many teachers. We use the tension between privacy and open science to frame our response, and argue that whether one is carrying out empirical research or deploying a learning analytics system, balancing the (often seen as) competing values of openness and privacy is important and that doing so begins with a deep understanding of the specifics of the context. We conclude with a call for educational technology scholars to meet the challenge of studying learning

(and disruptions to learning) in light of COVID-19 while protecting the privacy of students in ways that go beyond what Institutional Review Boards consider to be within their purview.

Privacy and confidentiality are core considerations in education and should be protected at all costs. Also, sharing data—and, more broadly, open science—is increasingly valued by editors, funding agencies, and the public (van der Zee & Reich, 2018). For instance, during the COVID-19 pandemic, sharing data may have enabled researchers to address questions not directly related to the original purposes for which the data were collected (Doughton, 2020).

In addition, privacy often exists in tension with the utility of learning analytics systems (Chen & Zhu, 2019). For example, during COVID-19, schools have rapidly shifted to digital modes of teaching and learning, prompting researchers to evaluate the effectiveness of this response. At such a time of crisis, data about student learning, including data collected through learning analytics systems, are especially important for understanding disparities caused or amplified by the pandemic (Kuhfeld & Tarasawa, 2020). Although emergency modes of learning may invite more educational technologies into the classroom or altogether change our understanding of what constitutes "a classroom," the increased presence of these technologies also renews questions about how educational data are collected and used (Watters, 2020). Data, even data collected for valorous purposes in learning analytics or open science, can be misused by those with power toward destructive ends (D'Ignazio & Klein, 2020).

1   An Empirical Study of Students' Perceptions of Privacy

In an empirical study, Ifenthaler and Schumacher (2016) asked students about their perceptions of privacy in learning analytics. A key insight from this research—similar to findings from Fiesler and Proferes' (2018) study of participants' privacy in empirical research—is that many of the practices in which researchers engage are not desired by student participants. Additionally, students did not perceive all collected data equally: there was variability in which data sources students were comfortable sharing. For instance, although more than 80% of respondents said they would not mind their course enrollment data to be used in learning analytics systems, less than 25% wanted their parents' educational level
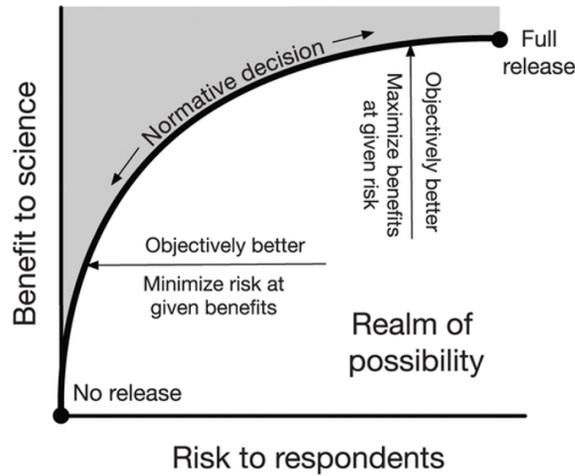
Figure 1: Lundberg et al.'s (2019) model for balancing risk to respondents with societal (science) benefits. *Note.* Figure reproduced from Lundberg et al. (2019).

to be used, and less than 10% were comfortable with their medical data being used.

2   Both-And Thinking For Participants' Privacy and Open Science

Lundberg et al. (2019) highlighted a broader tension between privacy and open science that further illustrates the import of Ifenthaler and Schumacher's (2016) learning analytics study. Due to calls for scientists to share their work in an open way to build trust in findings, Lundberg et al. (2019) wrestled with how much data to share and with whom. They did not prescribe a recipe to follow; instead, their solution was deeply contextualized in their project, discipline, and understanding of the risks to participants. Lundberg et al. developed a decision-making model for balancing privacy and open science (Figure 1), creating a tiered system of sharing some data publicly and other data through a gated application system.

In summary, Ifenthaler and Schumacher's (2016) highlighted the importance of privacy considerations in learning analytics, and Lundberg et al. (2019) presented a way to move forward with open science values while also respecting the privacy

and rights of participants.

3 Implications for Educational Technology Designers and Researchers

Our key implication is that educational technology designers and researchers should consider open science and privacy as values that need to be balanced in their work. The tension between open science and privacy parallels the tension between the utility of learning analytics and the privacy of students (Chen & Zhu, 2019). Whether one is carrying out empirical research or deploying a learning analytics system, balancing begins with a deep understanding of the specifics of the context (Greenhalgh et al., 2021; Kimmons & Veletsianos, 2018), including—as Ifenthaler and Schumacher (2016) point out—what students think of the specific data being collected.

We offer several thoughts on how researchers can achieve this balance in their work. First, researchers should choose to listen to students, especially when collecting sensitive data for learning analytics or empirical research. Instructional designers and researchers should also keep in mind that some data are more private than others (Ifenthaler & Schumacher, 2016) and assume that participants are not inclined to share health-related data (e.g., daily wellness checks, COVID-19 diagnostic tests). Both instructional designers and researchers should be clear in communicating to teachers and students what data they are collecting, as well as offer opportunities for participants to learn more and give feedback regarding their wishes for how such data are used. These recommendations are likely more stringent than what the Family Educational Rights and Privacy Act (FERPA) or an Institutional Review Board (IRB) may require. Therefore, instructional designers and researchers should consider adherence to these guidelines as a necessary-but-insufficient step in protecting participants' privacy.

Second, to date, open science is not widespread in education (van der Zee & Reich, 2018). However, there are important reasons for educational technology researchers and designers to share data and related materials more openly than in the past. At the same time, participants' privacy is paramount and must not be compromised. In our recent work (Greenhalgh et al., 2020, Staudt Willet & Carpenter, 2020), we have publicly shared code for analysis on GitHub and data on Open Science Framework (osf.io), but did so in different ways, depending on

the nature of the data we used. We deemed Twitter data to be more sensitive and thus created a carefully anonymized version of the dataset (Greenhalgh et al., 2020). We also created an application process for other researchers to request access to the original data. For access, we required a project description and strategies for protecting the privacy of participants in the data. In contrast, in a different study, we shared Reddit data in their original form, because the platform norm is that Reddit users' profiles are not typically identifiable with their offline identities (Staudt Willet & Carpenter, 2020). Minimizing participants' risks may require sharing no data; maximizing open science requires sharing all data. However, in many cases, neither extreme is tenable. A way forward is to balance privacy and open science, holding risks to participants as one consideration and the benefits of sharing data as another.

Finally, researchers should seek out guidelines, such as Prinsloo and Slade's (2018) for the informed consent process. They argued that researchers should consider how *certain* the outcome of a learning analytics application is, and how much *risk* is posed to participants, in order to determine how the consent process should take place. Future work may extend Lundberg et al's (2019) model for balancing privacy and open science. Future work should also build upon Chen and Zhu's (2019) findings and call to balance the practical utility of learning analytics with students' rights by articulating strategies and guidelines for *how* learning analytics systems can do this. The targeted guidance available with respect to the consent process is a start. Furthermore, a rubric that addresses multiple stages of the learning analytics process may be especially helpful for scholars—including ourselves—seeking to balance the benefits of openness with the importance of protecting students' privacy in their work.

4   Conclusion

Data on student learning are *especially* important during educational disruptions related to COVID-19. However, schools' collection of health-related data from teachers and students poses new risks. Even before the current pandemic, scholars from learning analytics (Arnold et al., 2020) and educational technology (Greenhalgh et al., 2021; Kimmons & Veletsianos, 2018; Krutka et al., 2019) have called for researchers to consider privacy issues beyond what is in the purview of FERPA or the IRB. The contribution of Ifenthaler and Schumacher's (2016) work to these renewed ethical conversations in learning analytics is to ensure students'

preferences are taken into consideration, especially when applying Lundberg et al.'s (2019) model for balancing privacy and open science.

5   References

Arnold, K., Ham, M. Pappas, R., & Rehrey, G. (2020). Learning analytics principles of use: Making ethics actionable. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK20)* (pp. 1-5). `https://dl.acm.org/doi/proceedings/10.1145/3375462`

Chen, B., & Zhu, H. (2019, March). Towards value-sensitive learning analytics design. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK19)* (pp. 343-352). `https://dl.acm.org/doi/10.1145/3303772.3303798`

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Doughton, S. (2020). Drugs touted by Trump, blood from recovered patients: Seattle scientists seek coronavirus cures. *The Seattle Times*. `https://www.seattletimes.com/seattle-news/health/seattle-scientists-fast-track-potential-treatments-for-covid-19/`

Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media + Society, 4*(1), 1-14. `https://doi.org/10.1177/2056305118763366`

Greenhalgh, S. P., Koehler, M. J., Rosenberg, J. M., & Staudt Willet, K. B. (2021). Considerations for using social media data in learning design and technology research. In E. Romero-Hall (Ed.), *Research methods in learning design & technology*. Routledge.

Greenhalgh, S. P., Rosenberg, J. M., Koehler, M. J., Akcaoglu, M., & Staudt Willet, K. B. (2020). Identifying multiple learning spaces within a single teacher-focused Twitter hashtag. *Computers & Education, 148*(4). `https://doi.org/10.1016/j.compedu.2020.103809`

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development, 64*(5), 923-938. `https://doi.org/10.1007/s11423-016-9477-y`

Kimmons, R., & Veletsianos, G. (2018). Public internet data mining methods in instructional design, educational technology, and online learning research. *TechTrends, 62*(5), 492-500. `https://doi.org/10.1007/s11528-018-0307-4`

Krutka, D. G., Heath, M. K., & Staudt Willet, K. B. (2019). Foregrounding technoethics: Toward critical perspectives in technology and teacher education. *Journal of Technology and Teacher Education, 27*(4), 555-574. `http://learntechlib.org/p/208235/`

Kuhfeld, M. & Tarasawa, B. (2020). *The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement*. Northwest Evaluation Association. `https://www.nwea.org/research/publication/the-covid-19-slide-what-summer-learning-loss-can-tell-us-about-the-potential-impact-of-school-closures-on-student-academic-achievement/`

Lundberg, I., Narayanan, A., Levy, K., & Salganik, M. J. (2019). Privacy, ethics, and data access: A case study of the Fragile Families Challenge. *Socius, 5*, 1-25. `https://doi.org/10.1177/2378023118813023`

Prinsloo, P., & Slade, S. (2018). Student Consent in Learning Analytics: The Devil in the Details?. In: J. Lester, C. Klein, A. Johri, & H. Rangwala (Eds.), *Learning analytics in higher education: Current innovations, future potential, and practical applications* (pp. 118-139). Routledge.

Staudt Willet, K. B., & Carpenter, J. P. (2020). Teachers on Reddit? Exploring contributions and interactions in four teaching-related subreddits. *Journal of Research on Technology in Education, 52*, 216-233. `https://doi.org/10.1080/15391523.2020.1722978`

Watters, A. (2020). The ed-tech imaginary. In M. Gresalfi & I. S. Horn (Eds.), *The interdisciplinarity of the learning sciences: Conference proceedings* (Vol. 1; p. 4). International Society of the Learning Sciences. `http://hackeducation.com/2020/06/21/imaginary`