When the classroom becomes datafied:
A baseline for building data ethics policy and data literacies across higher education

Quando la classe diventa datificata:
una base di partenza per la costruzione di politiche sull'etica dei dati e l'alfabetizzazione all'uso dei dati nell'istruzione superiore

Bonnie E. Stewart* and Erica Lyons

Faculty of Education, University of Windsor, Canada, bstewart@uwindsor.ca*, elyons@uwindsor.ca

*corresponding author

ABSTRACT This paper overviews a summer 2020 pilot survey of educators' perspectives on the intersection of educational technology and datafication in higher education classrooms. The brief, international survey of university teachers used four proxy questions to frame a baseline snapshot of higher education teaching populations' knowledge, practices, experience, and perspectives on data and online learning: this paper focuses specifically on the results of the knowledge and practice questions. The paper suggests that, in the Emergency Remote Education (ERE) context generated by the COVID-19 pandemic, higher education instructors teaching online demonstrate patterns of limited knowledge and practice surrounding the data aspects of their classroom tools. The paper posits an urgent need for institutional and sector-wide policy and faculty development around data and online classroom tools, and for data ethics to be addressed as part of institutions' ERE transition online.

KEYWORDS Datafication; Data Literacies; Higher Education; Professional Development; Faculty Development; Data Ethics.

SOMMARIO Questo articolo presenta un'indagine pilota condotta nell'estate 2020 sulle prospettive degli educatori riguardo l'intersezione tra tecnologia educativa e datificazione nelle classi di istruzione superiore. Un breve sondaggio internazionale con docenti universitari ha utilizzato quattro domande "proxy" per inquadrare in un'istantanea di base la conoscenza, le pratiche, l'esperienza e le prospettive di una popolazione di insegnanti di istruzione superiore relativamente ai dati e all'apprendimento online: questo documento si concentra in particolare sui risultati delle domande volte a indagare la conoscenza e la pratica. L'articolo suggerisce che, nel contesto dell'Emergency Remote Education (ERE) generato dalla

pandemia COVID-19, i docenti di istruzione superiore che insegnano online esibiscono limitati modelli di conoscenza e pratica che circondano gli aspetti dei dati dei loro strumenti di classe. L'articolo postula il bisogno urgente di una politica istituzionale e settoriale e di sviluppo professionale sui dati e gli strumenti d'aula online, e che l'etica dei dati sia affrontata come parte della transizione ERE online delle istituzioni.

PAROLE CHIAVE Datificazione; Alfabetizzazione dei Dati; Istruzione Superiore; Sviluppo Professionale; Sviluppo Accademico; Etica dei dati.

1. INTRODUCTION

In an era of 'smart' surveillant devices and platforms and unprecedented online learning due to the COVID-19 pandemic, educators across higher education face a common learning curve: the systems we rely on for scholarship and teaching are increasingly datafied (Perrotta & Williamson, 2018). In the context of the pandemic, tools designed to extract behavioural data (Erickson, 2018) from digital experience stand in for the traditional four walls of the classroom. Yet this shift has been minimally addressed by higher education as a sector. Even as race and gender biases built into algorithmic decision-making become increasingly evident (Benjamin, 2019; Noble, 2018) and concerns emerge about the privacy implications of plagiarism detection (Brinkman, 2013; Morris & Stommel, 2017) and exam proctoring software (Kolowich, 2013; Swauger, 2020), students at academic institutions around the world are now subject to data collection from these and other digital learning platforms as they engage in daily coursework (Bozkurt et al, 2020). Although contemporary universities demand that scholars be intentional, careful, and ethical with research data and research subjects, most currently boast minimal policy or faculty guidance regarding the ethics of data and datafied platforms. Whether educators want to engage with data or not, their work both generates it and guides students into environments that mine it. This paper draws on a summer 2020 pilot survey of educators in university contexts as a first step in considering how institutions could foster understanding about digital learning tools and data amongst academic staff, and – by extension – students? Ultimately, the paper proposes that policy as well as faculty development initiatives are required.

The pilot survey was completed in August 2020, and was open to any university educator around the world who anticipated teaching online – at least in part – in September 2020. The survey goal was to establish a baseline picture of educators' understandings of the data implications of digital tools, in the midst of the Emergency Remote Education (ERE) response to COVID-19. The pilot focused on four elements of educators' relationships to data and digital classroom tools – their knowledge, practices, experiences, and perspectives. It asked respondents a single proxy question about each element, and correlated responses to demographic questions about geographic locale, academic role and status position, years of teaching in higher education, and experience with different online classroom tools. No single proxy question can fully represent complex knowledge or practice: the purpose of these proxies was to establish baseline patterns and to examine patterns' consistency across geography, academic status, and other demographic factors. The pilot survey captures timely – if limited – evidence, but offers grounding for a planned in-depth examination of the details of all four categories.

The study's knowledge proxy question focused on the location of LMS servers, while the practice proxy investigated the extent to which faculty report reading Terms of Service (TOS) for classroom tools. In the authors' Canadian context, "*there has been concern that Canadian student information and communications held on cloud servers in the USA may be accessible via the U.S. Patriot Act*" (Bates, 2019, section 8.9.2), hence we understood server location as an important indicator – among many potential options – of

knowledge about data and classroom tools. TOS engagement had emerged as a limited practice in a prior project with K-12 and preservice educators (Stewart, 2020), and we chose this as the practices proxy since we wanted to know whether the same held true for faculty.

The pilot revealed that approximately two-thirds of respondents did not know which country housed servers for their institutional LMS, and that not even 10% of respondents always read TOS. While it's known that TOS are generally ignored (Obar & Oeldorf-Hirsch, 2018) and these proxy findings are not surprising, they do suggest that higher education classrooms are datafying at a faster rate than educators are necessarily keeping up with. Amidst the ERE mass transition online, gaps in data knowledge and teacher practice demand attention.

Yet our premise is not that educators are derelict of duty in regards to data knowledge and practices: the landscape of data has to date been framed largely outside the purview of higher education instructors and instead as a technical issue (Raffaghelli & Stewart, 2020). Higher education has not yet grappled with the full complexity of the ethical responsibilities that data collecting tools pose. More, the sector has not focused on educating its educators or students about data. Thus we conclude that in this ERE moment, when digital platforms are serving as the classroom for many students around the world, ethics-focused policies that develop data knowledge and practices among educators and learners must be a priority for higher education.

2. BACKGROUND

Purposeful data collection in education can be a useful tool in promoting change, creating improvements, and supporting students. Educational data, including that generated by digital technologies and algorithmic platforms, can assist in the process of transforming educational policies, pedagogies and practices (Williamson, 2017). From recruitment and retention through student experience interfaces, online program management (OPM) infrastructure services, and digital learning, many aspects of higher education are increasingly datafied, with data drawn upon as a guide for action and interventions. Institutions collect demographic, financial, library, and academic record and performance data as well as behavioral data based on institutional Learning Management System (LMS) platforms and plagiarism monitoring tools, and in some cases, regulatory compliance and health data (Braunack-Mayer et al, 2020). Student data is also gathered by the many third-party digital platforms and tools that students are encouraged or expected to use in online courses. Data on student activity and engagement can generate new insights regarding the experiences of students, and serve as a foundation for systemic change (Long & Siemens, 2011).

However, the rise of narratives focused on improving student learning via data-driven decision-making led to the surfacing of ethical issues and privacy concerns regarding institutional use of digital data, including issues of student surveillance in virtual learning environments (VLEs) (Land & Bayne, 2005) and in learning analytics (Slade & Prinsloo, 2013; Willis, Slade, & Prinsloo, 2016).

Datafication also can be fraught with bias and unintended consequences. The algorithmic logics that govern datafied decision-making can reinforce opaque and discriminatory practices: scholars have raised questions of whether algorithmic assessments of student behaviour are actually accurate, particularly in relation to high-stakes issues such as plagiarism (Weber-Wulff, 2019). In summer 2020, MIT had to take a highly-cited dataset offline when it was found to have trained Artificial Intelligence (AI) and machine learning systems to use sexist and racist terms to describe everyday images (Chadwick, 2020).

Even where benefits are promised, there may be risk. MacCarthy (2014) overviewed predictive analytics-based interventions used to identify students at risk of not graduating; they were found to be somewhat successful at reducing drop-out rates, but also to bypass and even undermine privacy norms by collecting swaths of information not covered by consent. The risks that datafied classroom platforms pose also include

the exploitation of such student data in private markets, pedagogical reductionism and overemphasis on what is countable (Williamson, Bayne, & Shay, 2020), and stigmatization of students via cross-campus data sharing (Benjamin, 2019; Marachi & Quill, 2020). During the pandemic, students have been required to install location apps or mandatory wearables in what Tufekci (2020) calls the 'surveillance theatre' of COVID monitoring, exposing students to potential leaks, hacking, and misuse of data. More broadly, datafication can lead to students being refused funding for education because automated lending models deem their zip codes risky (O'Neil, 2016) or to racialized students being falsely identified or unrecognized entirely by tools that rely on facial recognition technologies (Jones, 2020). The 'digital redlining' (Gilliard & Culik, 2016) that tends to restrict community college students' access to digital services and searchable information compared to those using university systems demonstrates how datafication's impacts are seldom transparent or visible, even to those they constrain.

In spite of these discriminatory impacts, datafication in higher education has not been subject to the same kind of ethical scrutiny or policy governance that regulates research practices. Institutions have built and benefited from analytics models without necessarily applying ethical governance models. While scholars have highlighted the need for ethics and equity lenses regarding data (Braunack-Mayer et al., 2020; Johnson, 2019), higher education's focus on "*the technical side of data engagement, encompassing business-oriented, positivistic approaches*" (Raffaghelli, 2020, para. 2) has tended to reinforce the construct of data as objective or neutral. And although the COVID-19 pandemic has led to a rise in online classes and a surge in online proctoring (Flaherty, 2020), no parallel rise in policy communications about the data implications of these trends appears to have emerged at the sector level as an ERE response.

As a result, educators and students may not even be aware of the data collection taking place in their learning contexts, or of how their data is used. While the idea that educators should use data to improve their teaching and students' performance was part of the early data-driven decision-making narrative, the responsibility to actually educate instructors or students themselves about data policies or practices is not one that institutions have undertaken, on the whole.

At the higher education level, few mechanisms currently exist to support academics in building key data literacies. These include understanding data sources, risks, and the implications of data collection over the long-term (Maybee & Zilinski, 2015), but also the more complex reflexive and critical socio-technical infrastructures underpinning data use (Pangrazio & Selwyn, 2019; Raffaghelli, Manca, Stewart, Prinsloo, & Sangrà, 2020). The siloing of technical knowledge in many contemporary universities works against this: institutional educational technologies tend to fall under the purview of Information Technology (IT) or other alt-academic departments and specialists, even where tools are to be used across a campus community. When communications do happen regarding institutional data policies, their meaning is likely to be obfuscated by the terminology of educational technologies' Terms of Service (TOS), which tend to be highly technical and outside most instructors' areas of expertise.

The implications of datification for educators and scholars, whose 'valued self-identity' Winter (2009) describes as grounded in commitment and competence, are significant. As Williamson (2020) notes in a broad mid-COVID review of the impacts of datafication and automation on higher education in the UK, a datafied academy not only privileges a technical model of teaching and learning that runs counter to more relational and critical pedagogies (Williamson, 2020), it also risks deprofessionalizing academic staff, and demanding a process of reskilling or reprofessionalization. Building capacity in data is increasingly required for educators who wish to maintain self-identity as competent within the shifting sector of higher education. Moreover, if educators, as knowledge workers, are not knowledgeable about the conditions and infrastructures of the contemporary classroom, then the construct of shared governance within higher edu-

cation is undermined. However, Raffaghelli's (2017) four dimensions of conceiving professional learning for complex systems, with its framework of competencies and scenarios, institutional strategies and policies, environments, resources, and activities enabling self-direction, and showcases of success, may be an important path for the capacity-building this study calls for.

3. METHODS

The research question that underpinned this study was "*What are the perspectives and practices regarding educational technologies and data of university educators who expect to use online teaching tools during the 2020/2021 academic year*?" The instrument was a short 5–10-minute online questionnaire containing 12 questions, one-third of which were demographic while others focused on tools and on knowledge, practice, experiences, and perspectives related to data and classroom tools. Our team designed the instrument as a foundation for a forthcoming in-depth study, ultimately focused on potential faculty development and policy interventions. The pilot survey was circulated online for three weeks during August 2020. Anyone currently teaching or expecting to teach in 2020/2021 in a university setting worldwide was eligible to participate, whatever their academic status. Further research building on this survey will dig deeper into its proxy categories, and will compare and contrast perspectives from instructors at different types of institutions as well as in different academic roles and status positions.

3.1. *Sampling*

Our goal in terms of sample size was to obtain at least 252 responses, for a diverse and representative sample. The 252 minimum sample size was calculated based on a priori executed using the G*Power tool. We calculated that for validity, approximately 28 individuals in each major variable category were needed (ie. teaching role: four categories; years of experience: five categories).

The decision to use 28 was the result of this G*Power calculation where correlation was set to .3, alpha level was set to .05, and power was set at .999. This resulted in a 237-participant sample being needed for .999 statistical power. Our sample size of 339 is statistically powerful, with a power of .999.

Since the purpose of the study was to survey teaching staff from a variety of geographic locales, a participant pool was not an optimal recruitment tool. Thus, spaces where educators can be contacted online were utilized for recruitment: Twitter, Facebook, and also institutional email and mass email listservs. The latter two were important, as many educators who may not have established social media networks or digital practices are teaching online during COVID-19, and these educators' perspectives were important to try to recruit.

In terms of specific demographic targets, our aim was to obtain responses that offered some perspectives from outside the m/f gender binary, and that represented varying roles within the academy, varying levels of teaching experience and experience with technologies, plus a diversity of disciplines and geographic locales. Ultimately, 339 participants completed the survey (Table 1). The participants identified as working in 25 countries around the world, within a variety of faculties and with a wide range of years of teaching experience: 19% of respondents had taught less than 5 years, while almost 20%, had taught more than 21 years. Most had some specific prior experience teaching with educational technologies, and all were anticipating using digital tools within the 2020-2021 academic year. The largest number of participants were from Canada (38%) and the United States (40%), which was anticipated as the research team is based at a Canadian university. The other 22% were based predominantly in Europe, though with multiple responses from Egypt, Turkey, Australia, and Mexico, among other countries. The participants were 62.5% full-time faculty, with 32.5% identifying as male, 63% as female, and a small subsection self-identifying an alternate gender designation or selecting "prefer not to answer." The majority identified as teaching in the

Humanities and Social Sciences (54.9%), about a quarter were from Education (24.5%), followed by Sciences (11.5%), Other (including Nursing, Health Sciences, Library Science, and Interdisciplinary Studies) (11.8%), Business (7.1%), Engineering (4.1%), and Law (2.1%).

N=339		
Gender Identification		
Identify as Female	214	32.4%
Identify as Male	110	63.1%
Self Identify	4	1.2%
Prefer Not to Answer	11	3.2%
Country		
Africa	16	4.7%
Australia & New Zealand	7	2.1%
Canada	129	38.05%
Europe	34	10.0%
Mexico	7	2.1%
Middle East and Asia	4	1.2%
South America	4	1.2%
United States	137	40.4%
Other	1	0.3%
Teaching Role		
Full-Time (ie. Tenured, tenure-track, permanent)	212	62.5%
Contract Faculty (ie. Sessional, adjunct, contract, casual, temporary, visiting scholar)	85	25.1%
Staff/Ancillary with a teaching component	31	9.1%
Alternate Status	20	5.9%
Years of Teaching Experience		
5 Years or less	63	18.6%
6-10 Years	82	24.2%
11-15 Years	69	20.4%
16-20 Years	58	17.1%
21+ Years	67	19.8%

Table 1. Survey Participants' Demographic Information[1].

[1] For Teaching Role n is more than 339 because 9 participants selected multiple teaching roles and both of these selections were accounted for.

3.2. *Data Analysis*

A frequency analysis was done on gender identification, country of teaching experience, teaching role, and years of teaching experience to obtain a summary of demographic information (Table 1). A frequency analysis was then done on responses to the practices proxy question about reading TOS (Table 2). The findings outlined below are based specifically on bivariate analysis of years of experience and how often participants read the full TOS of new educational technologies (Table 3), then teaching role and how often participants read TOS (Table 4). This was followed by a bivariate analysis of years of experience and the knowledge proxy question about server location (Table 5), plus teaching role and location of LMS (Table 6).

These specific bivariate analyses offered insight into the intersection of varying experience levels and teaching roles with the knowledge and practice proxies. The bivariate analysis was done in SPSS using cross tabulations and custom table tabulations utilizing the Bonferroni correction to assist with errors.

4. RESULTS

The first tables below offer a frequency analysis of how often participants read the full TOS of new educational technologies. The remaining tables present the bivariate analysis of the variables of years of experience and teaching role in comparison to the knowledge and practice proxy questions. Tables 3 to 5 provide an actual count of respondents and then indicate the row % (the proportion of people in a column category from among those in the row), and column percentage (the proportion of people in each row from among those in the column). For example, in Table 3, 18.80% of people with 5 years or less teaching experience read the TOS privacy agreement 90% or more of the time and 9.5% of those that read the TOS 90% of the time or more have been teaching for 5 years or less).

How often have you read the full Terms of Service (TOS) privacy agreements of new educational technologies before using them with students?	90% of the time or more	32	9.4%
	50-89% of the time	42	12.4%
	10-49% of the time	63	18.6%
	Less than 10% of the time	202	59.6%

Table 2. How often the full Terms of Service (TOS) privacy agreements of new educational technologies were read before use with students.

		5 Years or less	6-10 Years	11-15 Years	16-20 Years	21+ Years
How often have you read the full Terms of Service (TOS) privacy agreements of new educational technologies before using them with students?	**90% of the time or more**	6	7	4	6	9
	Row Total N %	18.8%	21.9%	12.5%	18.8%	28.1%
	Column Total N %	9.5%	8.5%	5.8%	10.3%	13.0%
	50-89% of the time	11	12	5	8	6
	Row Total N %	26.2%	28.6%	11.9%	19.0%	14.3%
	Column Total N %	17.5%	14.6%	7.20%	13.8%	9.0%
	10-49% of the time	9	17	16	12	9
	Row Total N %	14.3%	27.0%	25.4%	19.0%	14.3%
	Column Total N %	14.30%	20.70%	23.20%	20.70%	13.40%
	Less than 10% of the time	37	46	44	32	43
	Row Total N %	18.30%	22.80%	21.80%	15.80%	21.30%
	Column Total N %	58.70%	56.10%	63.80%	55.20%	64.20%

Table 3. Years of teaching experience vs. how often the full Terms of Service (TOS) privacy agreements of new educational technologies was read before use with students.

		Full-time faculty	Contract faculty	Staff or An-cillary with a teaching component	Alternate status
How often have you read the full Terms of Service (TOS) privacy agreements of new educational technologies before using them with students?	**90% of the time or more**	**18**	**9**	**4**	**1**
	Row Total N %	56.3%	28.1%	12.5%	3.1%
	Column Total N %	8.5%	10.6%	12.9%	5.0%
	50-89% of the time	**26**	**10**	**4**	**3**
	Row Total N %	61.9%	23.8%	9.5%	7.1%
	Column Total N %	12.3%	11.8%	12.9%	15.0%
	10-49% of the time	**32**	**21**	**9**	**3**
	Row Total N %	50.8%	33.3%	14.3%	4.8%
	Column Total N %	15.1%	24.7%	29.0%	15.0%
	Less than 10% of the time	**136**	**45**	**14**	**13**
	Row Total N %	67.3%	22.3%	6.9%	6.4%
	Column Total N %	64.2%	52.9%	45.2%	65.0%

Table 4. Teaching role vs. how often the full Terms of Service (TOS) privacy agreements of new educational technologies was read before use with students.

		5 Years or Less	6-10 Years	11-15 Years	16-20 Years	21+ Years
Do you know which country/ countries house the servers and data for your current institutional LMS?	**Yes**	**19**	**30**	**20**	**25**	**22**
	Row Total N %	16.4%	25.9%	17.2%	21.6%	19.0%
	Column Total N %	30.2%	36.6%	29.0%	43.1%	32.8%
	No	**44**	**52**	**49**	**33**	**45**
	Row Total N %	19.7%	23.3%	22.0%	14.8%	20.2%
	Column Total N %	69.8%	63.4%	71.0%	56.9%	67.2%

Table 5. Years of teaching experience vs. awareness which country/countries store data for current institutional LMS?[2]

[2] Only 315 of the 339 of participants responded that they had used LMS in the past so only these participants were asked this question.

			Full-Time Faculty	Contract Faculty	Staff or Ancillary with a teaching component	Alternative status
Do you know which country/ countries house the servers and data for your current institutional LMS?	**Yes**		**68**	**30**	**18**	**5**
		Row Total N %	58.6%	25.9%	15.5%	4.3%
		Column Total N %	32.1%	35.3%	58.1%	25.0%
	No		**144**	**55**	**13**	**15**
		Row Total N %	64.6%	24.7%	5.8%	6.7%
		Column Total N %	67.9%	64.7%	41.9%	75.0%

Table 6. Teaching role vs. awareness which country/countries store data for current institutional LMS?[3]

Overall, results indicate that the majority (59.6%) of participants read the TOS on new classroom technologies less than 10% of the time (Table 2) and 78.2% of participants read the TOS privacy agreement less than half the time. Only a small percentage (9.4%) read the TOS privacy agreements often. Within the majority of participants that rarely read the TOS, there was not a substantial indication of variance among those with differing ranges of experience (Table 3). Whether educators had been teaching for fewer than five years or more than 21 years, TOS privacy agreements were not a significant consideration for most when integrating new technology. Around 55% of respondents with fewer than 5 years, 6-10 years, 16-20 years and more than 21 years experience read the TOS privacy agreement less than 10% of the time. The only slight variations from this 55% range were those with 11-15 years of experience or more than 21 years of experience. Within the group with 11-15 years of experience, 63.8% stated that they read the TOS less than 10% of the time. A slightly higher variance was found in those with more than 21 years of teaching experience as 64.5% of these participants read the TOS privacy agreement less than 10% of the time.

In terms of teaching role, educators in staff or ancillary roles with a teaching component were most likely to read the TOS privacy agreement (Table 4). However, still only 26% of staff and ancillary educators read TOS more than 50% of the time. The majority in all teaching roles read the TOS less than 10% of the time but the small number of respondents who wrote in an alternate self-identification of role were the least likely (70.6%) to read the TOS, with full-time faculty not far behind (64.3%).

The results also indicated that while 92.9% of participants had used an institutional LMS for teaching, the majority (65.8%) of those surveyed were not aware of where the data and servers of these LMS were housed. Awareness did not vary significantly among respondents with differing levels of teaching experience, though a slightly higher percentage of those in the 11–15-year range of teaching experience were unaware of server and data hosting location (71%) (Table 5). Awareness also varied somewhat depending

[3] Only 315 of the 339 of participants responded that they had used LMS in the past so only these participants were asked this question.

on teaching role (Table 6). More than 50% of educators who identified as staff and ancillary status within their university knew about server and data housing, while of the participants who identified as alternative status, 76.5% said they did not know the location of servers or data for their LMS.

As the vast majority (78.5%) of respondents were based in Canada and the US, the survey primarily reflects the reality in the contexts of these two countries. However, there were no significant differences in knowledge or practice emerging from any of the other countries from which we had more than one data point. The majority of participants from all countries read the TOS privacy agreement 10% or less of the time, and participants from all countries were fairly equivalent in knowledge of server and data location. The only exception was the United States, where 84.6% did not know where data and servers were housed. We were curious whether availability of faculty development or professional learning opportunities correlated with educators' knowledge or practices, however, the majority (77.3%) of those surveyed were offered professional development/learning opportunities as part of their paid role. The content of faculty development offered was not explored in this study.

5. DISCUSSION

Findings from the knowledge and practice proxy questions of this pilot survey indicate some of the limited ways in which a broad international sample of higher education instructors engage with core elements of student data privacy. In the ERE context, with so many educators online, this is an important snapshot, reflecting a sector that has not educated educators about data aspects of tools that have come to constitute a large number of its classrooms.

Obviously, there are significant limitations to the findings of this pilot survey, particularly with the intentionally narrow focus that this paper takes. The two proxy questions highlighted do not and cannot represent the myriad of possible means of surveying faculty's knowledge and practices around data. That means that the conclusions simply point to particular common patterns, rather than representing specific discoveries, per se. Faculty respondents who indicated they do not read TOS or know the location of LMS servers may know many other things about the data aspects of their classroom tools: our follow-up research will investigate these questions in detail. But the pilot nonetheless demonstrates that they do not know – and therefore cannot tell students – whether their classroom data is subject to other countries' data privacy laws. This matters, even simply as a broad pattern. In this ERE context the fact that so few faculties are likely to be discussing this issue critically with a cohort of students who are online like never before is not insignificant. Few societal institutions outside universities have any capacity or incentive to help students engage critically with the basics of their own data privacy, in a time when data is a valuable commodity. Thus, the minor gaps in knowledge and practice highlighted by this pilot survey nonetheless point to a major responsibility on the part of the higher education sector to step up to prepare its campus communities to understand the data implications of their classroom tools.

Higher education is where people go to learn to think critically about emergent challenges in society, and the sector has a duty of care to students not to simply outsource core functions to data-mining platforms without critical capacity-building among faculty and students. During COVID-19, where the campus experience is currently constituted in large part through datafied classroom tools, it is especially important that institutions are transparent about privacy rights and potential harms. Yet these two proxy question findings suggest that – across geographic and academic status differences – faculty are unlikely to be sources of information for students about the TOS or data risks of the tools they use to teach.

Two core recommendations emerge from this pilot research study. The first is that campus-wide and sec-

tor-wide professional development in the area of data literacies be undertaken. Plain language "what you need to know as an educator" and "what you need to know as a student" communications related to ethical data use for institutionally procured tools could address core knowledge issues. Succinct information such as which country the servers for a given tool are housed in, as well as why this matters and what it might mean for international students studying online from other countries, could address data privacy concerns. Designed and communicated effectively, brief caveats overviewing what educators and students need to know to build an informed practice related to data and classroom tools could be a foundation for deepening campus knowledge regarding data.

These approaches would need to be designed collaboratively, across campus divisions between what is framed as technical and what is framed as pedagogical. As Tsai & Gasevic (2017) have noted, there are a shortage of pedagogy-based approaches to dealing with the learning barriers that learning analytics identify. Extrapolated beyond analytics to datafication in its myriad of forms, we found few exemplars within higher education of pedagogy-based approaches to dealing with problems and learning curves identified as 'technical.' Yet the complex growth demanded of educators in adapting to the ERE teaching context and specifically to new data knowledge and practices will require authentic professional learning rather than simply content delivery (Webster-Wright, 2009). Educators need institutional and sector-wide support in developing critical data literacies, in order to teach students in turn.

But faculty development around institutionally-procured tools cannot address all issues that datafication creates. Data is collected on students through non-institutional digital platforms such as YouTube, Twitter, Kahoot, Padlet, PhET, Piktochart, Quizlet. The TOS of these tools are public, but rarely written to be read by non-specialists. As a result, even where educators may be interested in ensuring student data privacy, TOS can be difficult to decipher and apply to specific situations. Risks may not be evident to educators who are not content experts in both law and networked digital practice simultaneously.

Thus, our second recommendation is that data ethics guidelines and policy also need to be part of a sector-wide approach to addressing the datafication of the classroom. Our pilot survey demonstrates that, in the midst of the COVID-19 ERE transition, there are pervasive knowledge and practice gaps about data and data privacy among higher education teaching staff, internationally. This is an ethical issue. Both institutional and corporate data analytics will be generated through higher education's unprecedented online engagement during the ERE, and realities with real, human consequences also stand to be created in the process. Thus, even while continuing to gather information about specifics of faculty and student knowledge, practices, perspectives, and experiences, institutions have a responsibility to prioritize the development of policy and governance guidelines about data ethics and harm minimization for all campus stakeholders, in this ERE moment and context.

Policy needs to be both pedagogy- and procurement-focused, and as with effective professional and faculty development for the complex socio-technical reality of datafication, demands the collaboration of interdepartmental and cross-disciplinary teams, including pedagogical specialists, technology specialists, librarians, and legal and procurement experts. While the context of the COVID-19 pandemic and the ERE online pivot may make collaboration and shared governance challenging, ERE also means that at many institutions the campus experience is constituted via data-mining platforms without policies governing extraction and surveillance. This is not only an abdication of the academic tradition, but also an invitation to vendors to simply cut out the unwieldy middleman of higher education and deliver increasing direct online credentials at an increasing scale. If higher education as a sector does not address the incursion of aggressive corporate interests into its classroom structures as part of its ERE response, it risks being unable to reign in those interests later, or to recapture public trust in the ethics it claims to espouse and uphold.

6. CONCLUSION

It is in higher education's own interest, then, to develop new and effective data governance policies, practices, and communications across the sector, with clear student-centered guidelines on the ethical use of data at their core. Policies and practices need not be designed to prevent the use of learner data, but harm minimization must be at the core.

It is also in higher education's interest to begin to educate its educators in the area of data literacies. The study suggests that educators' knowledge and practices about data are misaligned with the environments in which we teach, at this moment. We assert that institutions of higher learning could be sites for changing this global reality through sustained inter-departmental data literacy development initiatives and communications, particularly if these efforts are complex and practice-focused; in effect, designed to challenge the instrumentalism that has accompanied and underpinned much of the embrace of datafication. Further research will explore educators' practices, knowledge, understandings and perspectives in greater depth, as a step towards trying to foster a culture within which higher education uses of data are ethical, legible, and transparent to students and staff alike.

7. REFERENCES

Bates, A.W. (2019). Teaching in a digital age: Guidelines for designing teaching and learning. Victoria, BC: BCCampus. Retrieved from https://opentextbc.ca/teachinginadigitalage/

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim Code. Cambridge, UK: Polity.

Braunack-Meyer, A.J., Street, J.M., Tooher, R., Feng, X., & Scharling-Gamba, K. (2020). Student and staff perspectives on the use of big data in the tertiary education sector: A scoping review and reflection on the ethical issues. *Review of Educational Research, 90*(6), 1-36. doi: 10.3102/0034654320960213

Brinkman, B. (2013). An analysis of student privacy rights in the use of plagiarism detection systems. *Science and Engineering Ethics*, *19*(3), 1255-1266.

Bozkurt, A., Jung, I., Xiao, J., Vladimirschi, V., Schuwer, R., Egorov, G., ... & Rodes, V. (2020). A global outlook to the interruption of education due to COVID-19 Pandemic: Navigating in a time of uncertainty and crisis. *Asian Journal of Distance Education*, *15*(1), 1-126.

Chadwick, J. (2020, July 2). MIT apologises after a giant dataset it was using to teach AI how to recognise people and objects in images was found to be assigning racist and misogynistic labels. *The Daily Mail.* Retrieved from https://www.dailymail.co.uk/sciencetech/article-8483929/MIT-pulls-racist-misogynistic-dataset-offline.html

Erickson, K. (2018, July 17). The future of network effects: Tokenization and the end of extraction. *Medium*. Retrieved from https://medium.com/public-market/the-future-of-network-effects-tokenization-and-the-end-of-extraction-a0f895639ffb

Flaherty, C. (2020, May 11). Big proctor. *Inside Higher Education*. Retrieved from https://www.insidehighered.com/news/2020/05/11/online-proctoring-surging-during-covid-19

Gilliard, C. & Culik, H. (2016, May 24). Filtering content is often done with good intent, but filtering can also create equity and privacy issues. *Common Sense*. Retrieved from https://www.commonsense.org/education/articles/digital-redlining-access-and-privacy

Johnson, J.A. (2019). Constructing data ethically. *New Directions for Institutional Research*, *2019*(182), 35-50. doi: 10.1002/ir.20306

Jones, W.D. (2020). Racial profiling goes high tech with facial recognition. *IEEE Spectrum.* Retrieved from https://spectrum.ieee.org/tech-talk/computing/software/do-you-have-the-right-complexion-for-facial-recognition

Kolowich, S. (2013, April 15). Behind the webcam's watchful eye, online proctoring takes hold. *The Chronicle of Higher Education.* Retrieved from http://chronicle.com/article/Behind-the-Webcams-Watchful/138505/

Land, R. & Bayne, S. (2005). Screen or monitor. In Bayne, S., & Land, R. (Eds.), *Education in Cyberspace* (pp. 165-179). London, UK: RoutledgeFalmer. doi: 10.4324/9780203391068

Long, P. & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30-32. Retrieved from https://er.educause.edu/~/media/files/article-downloads/erm1151.pdf%20

MacCarthy, M. (2014). Student privacy: Harm and context. *International Review of Information Ethics, 21*, 11-24

Marachi, R., & Quill, L. 2020. The case of Canvas: Longitudinal datafication through learning management systems. *Teaching in Higher Education 25*(4), 418–434. doi: 10.1080/13562517.2020.1739641

Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1-4.

Morris, S., & Stommel, J. (2017). A guide for resisting edtech: The case against Turnitin. *Hybrid Pedagogy*, *15*. Retrieved from https://hybridpedagogy.org/resisting-edtech/

Noble, S.U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York, NY, US: NYU Press.

Obar, J. & Oeldorf-Hirsch, A. (2020). The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society, 23*(1), 128-147. doi: 10.1080/1369118X.2018.1486870

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. London, UK: Allen Lane Penguin Books.

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media and Society*, *21*(2), 419–437. doi: 10.1177/1461444818799523

Perrotta, C., & Williamson, B. (2018). The social life of Learning Analytics: Cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, *43*(1), 3-6. doi: 10.1080/17439884.2016.1182927

Raffaghelli, J.E. (2017). Exploring the (missed) connections between digital scholarship and faculty development: a conceptual analysis. *International Journal of Educational Technology in Higher Education, 14*(20). doi: 10.1186/s41239-017-0058-x

Raffaghelli, J.E. (2020). Is data literacy a catalyst of social justice? A response from nine data literacy initiatives in higher education. *Education Sciences*, *10*(9), 233-253. doi: 10.3390/educsci10090233

Raffaghelli, J., Manca, S., Stewart, B., Prinsloo, P, & Sangrà, A. (2020). Supporting the development of critical data literacies in higher education: building blocks for fair data cultures in society. *International Journal of Educational Technology in Higher Education*, *17*(58), 1-22. doi: 10.1186/s41239-020-00235-w

Raffaghelli, J., & Stewart, B. (2020). Centering complexity in "educators' data literacy:" A critical review of faculty development literature. *Teaching in Higher Education, 25*(4), 435-455. doi: 10.1080/13562517.2019.1696301

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529. doi: 10.1177/0002764213479366

Stewart, B. (2020). The Open Page Project: Putting digital learning principles into practice for pre-service educators. *Journal of Teaching and Learning, 14*(1), 59-70. doi: 10.22329/JTL.V14I1.6265

Swauger, S. (2020). Our bodies encoded: Algorithmic test proctoring in higher education. *Hybrid Pedagogy.* Retrieved from https://hybridpedagogy.org/our-bodies-encoded-algorithmic-test-proctoring-in-higher-education/

Tsai, Y.S., & Gašević, D. (2017). Learning analytics in higher education -- challenges and policies: A review of eight learning analytics policies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on LAK '17*. ACM Press, 233–42. Retrieved from http://dl.acm.org/citation.cfm?doid=3027385.3027400

Tufekci, Z. (2020, September 4). The pandemic is no excuse to surveil students. *The Atlantic.* Retrieved from https://www.theatlantic.com/technology/archive/2020/09/pandemic-no-excuse-colleges-surveil-students/616015/

Weber-Wulff, D. (2019). Plagiarism detectors are a crutch, and a problem. *Nature*, *567*(7749), 435-436.

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research, 79*(2), 702-739. https://doi.org/10.3102/0034654308330970

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice.* London, UK: SAGE.

Williamson, B. (2020). *The Automatic University A review of datafication and automation in higher education.* University and College Union. https://www.ucu.org.uk/media/10947/The-automatic-university/pdf/ucus_the-automatic-university_jun20.pdf

Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives. *Teaching in Higher Education, 25*(4), 351-365. doi: 10.1080/13562517.2020.1748811

Willis, J., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: A typology derived from a cross-continental, cross-institutional perspective. Educational Technology Research and Development, 64. doi: 10.1007/s11423-016-9463-4

Winter, R. (2009). Academic manager or managed academic? Academic identity schisms in higher education. *Journal of Higher Education Policy and Management, 31*(22), 121–131.