Post-GDPR usage of students' 'Big-data' at UK Universities

FEARN, Carolyn and KOYA, Kushwanth <http://orcid.org/0000-0002-7718-1116>

Post-GDPR usage of students' Big-data at UK Universities

Carolyn Fearn[1] and Dr Kushwanth Koya[1*[0000-0002-7718-1116]]

[1]iSchool, College of Business, Technology & Engineering,
Sheffield Hallam University, Sheffield, S1 4WB, United Kingdom.
c.t.fearn@shu.ac.uk, k.koya@shu.ac.uk

Abstract. Higher education institutions are extensively using students' big-data to develop student services, create management or staff-led interventions and inform their strategic decisions etc. Following the implementation of the European Union's General Data Protection Regulation (GDPR) in 2018, there has been extensive uncertainty regarding the use of students' data. By conducting interviews with various University staff in the UK, this research aims to explore their understanding and usage of students' data, post-GDPR implementation. The findings indicate students' data is primarily used to build learning analytic tools and student-retention activities. Additionally, it was found that the understanding and usage of both big-data and GDPR differed across various Universities' stakeholders, and there is inadequate support available to these stakeholders. Overall, this research indicates the adoption of big-data based learning analytics requires comprehensive development and implementation policies to address the challenges of learning analytics. Therefore, this research proposes such an approach through co-creation with staff and students; institutional research and staff training.

Keywords: GDPR, big-data, learning analytics, higher education.

1 Introduction

1.1 Big-data in Higher Education

Big-data (BD) refers to large-scale data that are characterized by volume, velocity, veracity, variety and value [27]. Additionally, BD is defined as "building new analytic applications based on new types of data, in order to better serve your customers and drive a better competitive advantage" [4]. Challenges within higher education institutions (HEIs) have created an interest in BD and analytics as a potential solution to issues i.e. student-retention, personalized-learner support and changing pedagogy [1, 11, 26, 54]. Data created in HEIs through students' digital footprints provides an authentic reflection of real behavior, detailed insight into student performance and learning trajectories that could be used for personalized adaptive learning, and curriculum design [3]. However, it is irresponsible to believe more educational data always means better educational data and learning analytics (LA) possess limitations as well as multiple meanings [19].

Scholarly-works refer to issues in the use of BD, such as economic, legal, social and ethical, from both positive and negative aspects. Another concern is the automation of society in which actions are determined by behaviors and coercion, i.e. personalized advertising [2, 15, 31, 34, 39, 42]. The use of BD and LA needs strategic-leadership within any organization. "One of the biggest impacts of big-data will be that data driven decisions are poised to augment or overrule human judgement" (ibid, p.141). While the mining of BD in HE will support evidence-based research into enhancing learning and teaching, data taken out of context will lose meaning and value [7]. Furthermore, LA will only be effective if applied within course specific contexts rather than at institutional-level [50]. Careful consideration needs to be given to equality and inclusion when using BD, as within a retail-consumerist environment, not everyone engages with activities that BD tools can capture or analyze [25, 30, 48]. Not all students in HE leave the same type or volume of digital-footprint, this will vary between academic disciplines and the type of learner and their learning style [6]. In order to accurately use data to predict student success, or identify those at risk of withdrawal, the range and type of personal data that should be used needs to be more than just personal-biographical-data. A study suggests that the value of a degree is linked to personal cognitive motivations and economic benefits; therefore, using data identifying individual behaviors, such as critical thinking and social-emotional well-being will enhance the accuracy of predictions [51]. It is evident from the literature that increased data harvesting within HEIs offers the potential to improve student outcomes and retention [23]. However, what must also be taken into consideration is the compatibility of educational-datasets such as student's biographical, behavioral and curriculum data, and the capability of algorithmic approaches to interpret and present LA information.

1.2 GDPR and Learning Analytics

GDPR has brought clarity regarding the collection and use of personal data by presenting lawful bases for processing in the European Union (EU) [37]. The purpose of LA is for the benefit of students, either assisting them individually, or through aggregated data to improve educational experience more generally [23]. The Joint Information Systems Committee (JISC) recommends institutions should allocate specific responsibility within the organization to take accountability for the legal, ethical and effective use of LA [23, 24]. Models proposing the domain and application of LA consider six dimensions [19], however, no reference is made to challenges relating to the processes associated with LA i.e. the need for common datasets, data-quality or version-control [18, 20]. Additionally, application and compliance with data protection must be considered across each of the dimensions.

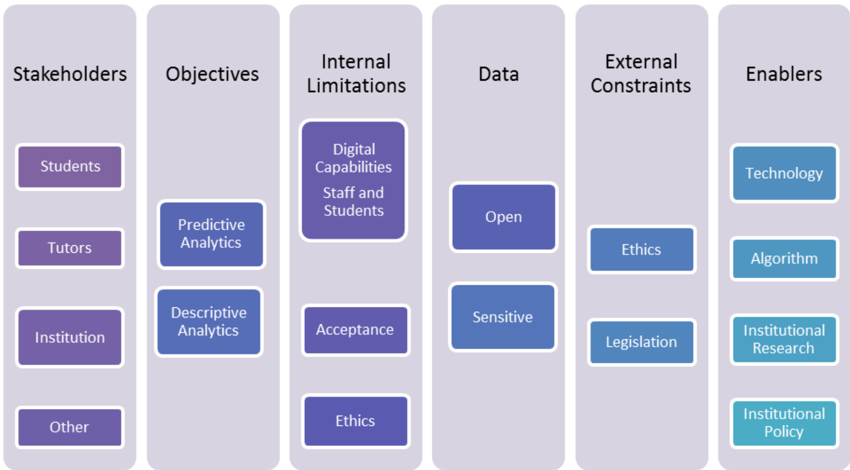

Fig. 1. Dimensions of LA. Adapted from Greller and Drachsler [19].

Scholarly-works indicate that LA has the potential for improving teaching and learning [24]. However, the longitudinal impact of LA as a discipline is not clear, particularly within the UK and EU following the implementation of the GDPR. It broadens the term 'personal and sensitive data' to include 'online identifiers' such as IP addresses and cookies, genetic and biometric data [22]. The privacy-rights of individuals have been strengthened to include: stricter rules for obtaining consent as a legal basis for processing data; the right to have personal data erased; the right to have clear information regarding what data is being collected and how it is being processed; the right not to be subject to a decision based solely on automated processing of one's personal data [14, 56].

The GDPR was adopted by the EU in 2016, replacing the 1995 Data Protection Directive which was created at a time when the use of digital data were in infancy. As an act of UK law, the EU GDPR requirements will continue to apply after Brexit [14]. This covers EU-based organizations collecting or processing personal data of EU residents and organizations outside the EU for monitoring behavior or offering goods and services to EU residents. Organizations non-compliant with the regulations could be subject to the imposition of sanctions. There are therefore profound implications for UK-based HEIs. Hence, the collection and processing of personal data must be justified under one of the lawful bases provided by the GDPR, for example: meeting a legal obligation, collection is in the institution's legitimate interest, or required to fulfil contractual obligations with the student. Additionally, a clear affirmative action of consent from the student must be obtained, where interventions with individual students are made based on their analytics, a limitless right to withdraw consent is made available with clear accessible mechanisms [22].

The term 'student-engagement' can be characterized by a diverse set of systems and agents, spanning both the physical and digital-spaces [6]. Scholarly-research finds a weak-relationship between student-engagement and student outcomes, suggesting that to collect data about students' interactions with activities and services, physical or digital, may not be invaluable for predicting student outcomes. Nonetheless, there is a need for reciprocal sharing of appropriate and actionable information between students and their institution, allowing students to make informed decisions and act accordingly [47]. An individual's intention to remain within HE and perform to the best of their ability is influenced by their motivations, interests and behaviors [32]. An extensive body of research literature spanning more than four decades, indicates that students' level of integration in both academic (student assessment results and satisfaction with their academic experience) and social environment (extracurricular activities and peer-relationships) are major contributing factors to HE student-retention and attainment [40, 45, 46, 49]. Additional factors include: institutional commitment (academic and technical support, physical environment) and personal circumstances (financial, health and lifestyle) [1]. LA integrates various types of data i.e. learning and teaching behaviors, academic performance and socio-economic status to inform interventions for students' learning, and how tutors teach and design their curriculum [38, 53].

1.3 Need for Empirical Research

In summary, the literature reviewed refers to the focus of LA with phrases i.e. "intervention", "students at-risk" and "prediction", implying that analytics is concerned with students who are poised to fail. This use of language continues to present a culture of students as passive subjects, the objects of the flow of data, rather than as self-reflecting learners who could use LA data as a cognitive tool to evaluate their own learning processes and set their own goals. However, little is known of those students who fall into other categories, whose data presents them as "stable" or "good". Another common theme throughout the literature is communication, inclusion and engagement with all stakeholders. The successful implementation and use of LA relies upon collaboration with all stakeholders - staff, students and management with a clear strategic objective set by senior leaders within an organizational culture that is change inclined. Therefore the motivation of this research is to provide informed guidance regarding the implementation and use of LA and GDPR, explore the potential to use BD in the context of LA, understand the level of stakeholder involvement and training provided. Thus, the following research questions need to be addressed:

RQ1: Where is BD being used within the HE sector post-GDPR implementation?

RQ2: How BD assists in developing LA and consequently, its usefulness to relevant stakeholders?

2 Methodology

2.1 Data Collection

Participants were purposely selected, comprising of representatives from various HEIs where LA were being used due to their interest and knowledge of the research topic, and invitations were sent via professional institution networks to participate in one-to-one interviews, [52]. Nine participants agreed to participate in semi-structured interviews, with questions derived from literature.

A participant information-sheet, consisting of questions and themes to expect during the interview, was provided in advance for orientation. The themes of the questions reflect the overall objectives of this research, to understand the usage and comprehension of student 'BD' and GDPR. Interviews lasted between 30 to 45 minutes, were recorded and manually transcribed for analysis.

2.2 Data Analysis

The interviews were analysed using thematic analysis [8], a structured qualitative method applied to discover, interpret, analyse and communicate clusters of data within the text [12]. An inductive process of analysis was followed due to a small sample size and purposive sampling [43]. Braun & Clarke's (2006) process of thematic analysis as depicted in Figure 3 was applied. An iterative code checking process was additionally applied to ensure rigor and code maturity. The emerging themes were further refined to represent a specific definition and the context of occurrence. Although thematic analysis is flexible, this flexibility can lead to inconsistency and a lack of coherence when developing themes derived from the transcript data [21]. This was mitigated by creating a map to visualise the themes.

Participant	Institution characteristics	Designation
1	TEF Gold; Post-1992 University; 20000 students	Director of Teaching and Learning
2	TEF Bronze; 19000 students	Associate Dean
3	TEF Gold; 19000 students	Senior Lecturer and Academic Advisor (AA)
4	TEF Silver	Senior Lecturer and AA
5	TEF Silver	Senior Lecturer and AA
6	'withheld'	Head of Data Governance
7	TEF Silver; approx. 30000 students	Senior Lecturer and AA
8	TEF Silver; approx. 30000 students	Senior Lecturer and AA
9	'withheld'	Head of Information Governance and Data Protection

Fig. 2. Participant characteristics

To ensure further rigor, an inter-coder reliability test was conducted independent of the first coder. The second author coded all the interview transcripts and Cohen's Kappa was calculated to ensure repeatability of the emerged themes between independent coders [28]. This research approach is considered as interpretivist as it was about understanding the perceptions of participants, acknowledging that these observations will be subjective. Although it enables a deeper understanding of the participants' thoughts, perceptions and experiences in relation to the use of BD and GDPR, a further systematic review followed by a qualitative meta-analysis will confirm the findings [10, 41].

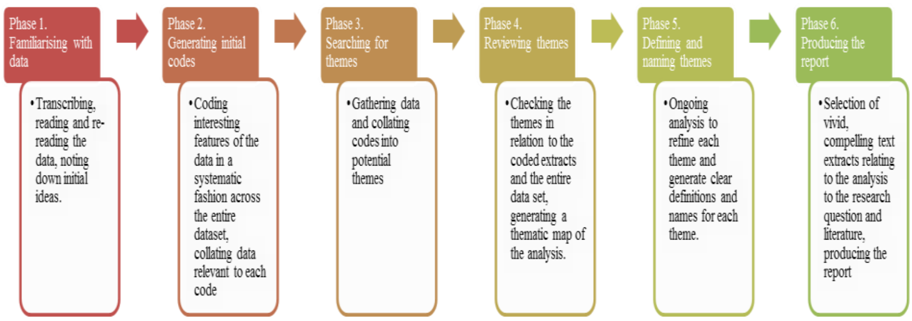

Fig. 3. Braun & Clarke’s thematic analysis framework [8].

3 Findings

3.1 Understanding the Term Big-data

The understanding of the term BD between the participants found consistent descriptions. Their understanding of BD relates to - the increasing availability and collection of data as a result of emerging technology capable of collecting large datasets; value and benefits of BD to inform decision-making; technical descriptions.

"different sources and different ways we can collect data now that we didn't used to have"

"how do we make it meaningful".

"we have a general rule of thumb, if it can fit on a laptop, it's not BD, that's from a computer science perspective".

Participants also reported the context of the data shouldn't be lost in any processing and acknowledge the increasing focus on data within the sector.

"..data driven society, being judged and governed against data today"

Four academic participants provided different responses. One reported no distinction between data and BD; *"data is data*". Two participants understood BD as large in volume and diverse in categories of student data, providing insight into their actions.

"all-encompassing data from all angles…how they interact with us"

One participant described BD as a new concept, with limited understanding, confined to its use as a marketing tool and unsure of its meaning within HE.

"came from a generation where BD is quite a new thing"

3.2 Current Use of Big-data to Support Students

At Institution Level. Within their institution to support student-retention, only one academic said BD is being used in-relation to attendance-monitoring and the triangulation of missed-sessions to inform interventions. The Institutional-strategy relating to student-retention has direct impact on academic advisors (AAs) who are presented with a dashboard containing individual student data: attendance-information, disability-statement and attainment-data. Academics are expected to engage and use this data and record notes of their interactions with students; this dashboard is not presented to students.

"….huge impact, the retention work we've done, the focus on retention - does impact on the way we do our job…"

"academic-tutors can enter and are expected to enter updates when they meet with students"

Participants mentioned their experience in a LA project which resulted in their University realising the potential of LA i.e. course delivery performance but acknowledged the challenge of integrating IT systems.

Data are used as *"proxies of which to judge success on different levels"*

Awareness of BD to support student-retention is mixed; two participants weren’t aware of existing practice or the use of BD and provided different responses, describing institutional data i.e. age, caring responsibilities, ethnicity and disability, are used to identify students at risk of withdrawing.

"it's a bit of an assumption that students in these categories would have retention and engagement issues….I have a student with a learning contract and a disabled brother, he's the most engaging student there is"

".. we get pushed on certain projects, the latest is about retention and achievement but of students from BAME backgrounds….last year we were pushed quite hard on commuter students - so they do use that information they only give it to me as and when they want me to use it."

The responses from institutional managers (IMs) varied; outlining a strategic management context regarding how data are used to structure strategic key performance indicators and the production of institutional retention reports, and attendance monitoring to indicate issues with non-attendance.

"So we are constantly evolving how we look at the data".

"..so that they (staff) can make interventions that would signpost them to support services and assist them in making sure that we retain the student"

Participant's use of big-data within their role

Within their role’s participants responded by reiterating BD was used for the purposes of statutory data reporting activities for management information. However, academics stated that they do not directly use BD, but relied on their relationships with the student.

"we build up a local level relationship with students".

"I have tried and trusted methods of asking and talking to the student very much on a one to one conversational basis getting to know them"

"The role is also tasked with overlaying a governance framework on the use of data in order to shift the University's culture towards being data driven and regarding data as assets."

3.3 Presentation of Learning Analytics Data to Staff and Students

Presentation to Students. Participants mostly responded that data are collected but not presented to students; one member was unsure if data are presented to students but would favour a "*transparent approach*".

Presentation to Staff. Two participants reported that the data are not presented to staff, but would welcome this approach to "*aid conversations with students"*

One participant responded that data are presented to staff in the form of a "*Personal Academic Tutor Dashboard*", although felt this was a "*crude*" presentation of data.

One academic expanded their response and suggested they wished to see individual and cohort profiles and information about cohort experiences.

"*Again it comes with doing that as an AA, comes with a health warning though, that people might just pull off the data and think well I don't really need to see them, I'll just send them an email…But behind every piece of data there is a story and it would be that that I would be frightened that got missed, so yes I can see the benefits but it comes with a health warning"*

Support for Staff and Students Using Big-data. All participants responded that there is a devolved specialist team within their institution that have ownership for data with support provided typically by one administrative colleague. Two participants specified resistance or lack of engagement from academic colleagues in using the data and one participant specifically referred to the need for institutional change, preferring personal academic tutorials rather than the need for support using data dashboard.

"it is pretty basic, it doesn't need much interpretation, it's about institutional change, it's more about making people do this, getting round the idea of personal academic tutorials"

"…teaching is just a bit of a bind, they would rather be doing their research, they'd rather be doing anything other than standing in a classroom. So how those students perform…not really interested".

All academics responded that support is not available, specifying the need for training and time to undertake the training. The academics also report that the focus of training should cover the reasons for using data and the consequences, rather than the mechanics of manipulation of data.

"*good question - there really isn't much support available to staff, because the data isn't given to staff*".

"it's also why would you use it, and what's the issue what's the flip side of using it erroneously."

Support was also required for academics regarding how to deal with students.

"*the person is more important than the data, you need to know how to deal with the person, and support for AAs as well…there can be some very sad stuff that happens to students and that can be quite difficult to deal with."*

3.4 Use of Predictive Learning Analytics

When asked if their institutions plan a move towards using predictive analytics (PA). One participant reported a limited institutional understanding and their ethical concerns regarding its use.

"I'm cynical about prediction because from my research, what I've shown and demonstrated is that students are so much more complex than prediction and I worry about prediction from an ethical point of view"

A project at one participant's institution had attempted predicting degree classifications, resulting in a model that was 70%-80% accurate. Participants commented that if student predictions were to be used there must be transparency regarding the algorithms. One participant said the use of PA should be applied at an institutional level. One academic responded that a manual local predictive process identifying students within specific categories and monitoring their attendance is current practice.

"You can present the prediction to students, but I think it would have to be done with blinkered eyes, and have it vetted before showing it to the student."

"So predictive analytics are quite useful but again my view on anything like that is putting people in boxes…..it should come with a caveat…".

Participants reported the essentiality of good relationships with students and the presentation of predictions "*should never take away the human element*". Presentation of predictive data to staff was not reported as an issue.

"...predictive is great for an AA to have a measure of who they are dealing with and to be mindful of it - actually sharing that with the student I can potentially see that as being counterproductive.."

"going back to predictive….we need to do it on us…if we keep doing what we are doing, we are like the dodo, we are going to work ourselves to extinction because we don't understand, even with big-data with these analytics, we will see them coming in, but we won't measure against that student we will measure against our standards of teaching and our standards of engagement, rather than what they need.".

3.5 The Impact of the GDPR on Existing Practice

Most academics reported there had been no impact on practice. IMs did add that the introduction had "*unfortunately*" resulted in the stopping of activities relating to the analysis of student data and, additionally reported that academic staff are accustomed to working with student numbers rather than names. As individuals using BD, participants reported practical adjustments regarding data-storage and privacy.

"*the first thing we say to our students who comes to us with a problem is what is your student number…so we are programmed in that way and our students are programmed in that way*".

However, GDPR according to IMs was on institutional policy and governance, bringing clearer rationale for using data to fulfil the requirements of student contracts. The GDPR has also highlighted the need for institutions to focus on data quality to avoid distress being caused to the data subject when using their personal data.

"*ensuring compliance with legislation processing data fairly and lawfully and looking after the rights of data subjects*"

A reported advantage of the introduction of the GDPR was the requirement for an Information Asset Register to document institutional data.

"*we will for the first time know what data we have, where it is and why we use it*".

3.6 Developing LA

Co-creation with Students. Five participants responded students are not involved with projects that use their personal data. The remaining participants indicated that students are invited to participate and contribute. One reported that their project had been developed through a user-centred design approach with students employed as ambassadors leading workshops to ensure the project was driven by what students said.

"there's a systematic literature review that shows that about 6% of student facing LA projects published have shown that have actually looked at working with students to design stuff. I think that mine is one of the first projects to do it fully.".

One participant expanded their response by commenting that students are in an environment where "*not sure they care*" about how their data are used, seeing their data in a social media context is their "*environment*" and the "*norm*" for them.

Co-creation with Staff. In general, the participants responded that staff are, or have been, included with institutional projects. The pilot LA project led at one HEI had been developed collaboratively with two academic members of staff.

" reflect on how it went and what needs to be improved for the following year it is based on data and evidence not just on subjective opinion".

Academic staff at another institution had been involved with the introduction of their attendance monitoring system. However, the participants also reported resistance from some academic staff within their institution.

"...you'll always get the ones who....resisters....we call those CAVEs - colleagues against virtually everything."

Institutional Approach to LA. Participant responses differed when asked about the Institutional approach to LA. One did not have any knowledge of the institutional approach and all other responses reflected different approaches i.e. driven by improving student engagement and student outcomes as opposed to focussing on students at risk.

Data Strategy. One IM referred to a data strategy written by a steering group consisting of staff from central directorates with knowledge and expertise in the use of data, although the data strategy doesn't refer to use of data for LA. The knowledge of an Institutional data strategy within the academic responses varied, with three participants having little or no awareness of a strategy and the remaining were aware.

"I'm going to have to say yes, it's a big enough organisation to need one....has it been particularly well shared - not so much".

Ethical issues of using big-data

Participants reported their concerns regarding the ethical use of BD and the impact this may have on all stakeholders.

"does the university have capacity….that's the key thing because once you open this you can then start to identify students at risk - if you can't then do anything about it then that's the biggest problem".

All participants reported concerns regarding data transparency; how the data are collected, processed and applied to predict outcomes.

"students are so much more complex than prediction and I worry about prediction from an ethical point of view."

Participants reported that Institutional discussions and a corporate approach regarding the ethical use of data is required.

"it feels like here we have almost just thought we just need a system and the actual cultural element and how it is going to be adopted by the front line users has not been explored enough."

4 Discussion

This research aims to explore within the context of the new GDPR legislation, how student data is utilised at UK HEIs. The study identified key themes as listed in Figure 4.

4.1 Understanding the Term Big-data

The term 'BD' was recognised by all participants who provided a range of descriptions; from no distinction between data and BD, to the variety, volume and rate at which data are available, partially aligning with current scholarly-works [4, 27]. This theoretical perspective is supported by the IMs and some academics who believe in technology advancements in HE. However, some academics were concerned regarding the use of data to monitor and judge academic performance, specifically referring to the TEF, as it is considered a proxy for success, a view shared in recent literature [55].

Category	Theme	Inter-coder reliability (Cohen's Kappa); Landis & Koch (1977) interpretation
Understanding 'BD'	Volume	0.63; Substantial agreement
	Variety	
	Value (to improve student services and offer intelligence)	
Use of 'BD'	Custom student services	
	Management or staff led interventions	
	Monitor not predict	
	Report student performance	
	Academic advisory	
Impact of GDPR	Storage of personal information	0.77; Substantial agreement
	Stopped analysing student data	
	Inundation of consent requests from private companies who work with the specific HEI	
	Changes to administrative practice to obtain student consent	
	Staffing change (recruitment)	
	Clarity on the definition of consent	
	Not much effect	
Retention research	Limited understanding or knowledge in the area of student retention	0.33; Fair agreement
	BD contributed towards retention of commuter students.	
LA	Student permission is a must	0.68; Substantial agreement
	Institutional planning, teaching and learning practice	
	Do not underestimate personal relationships	
	Provide assurance to students and staff	
	Transparency of the algorithm	
	Possibility of negative profiling	
	Should be combined with face to face interaction	
Support available to staff	No standards on how the student data in decision support systems	0.61; Substantial agreement
	Increase data variety under single student profiles to offer better support to students	
	Large quantities & types of student data, including digital footprints collected.	
	Inconsistent interpretations with current data (needs standardisation)	
	Training provision for staff involved at all levels (academic advisory, data management and analysis	
	Very little collaboration between staff or students to agree on the practices with the data collected.	
	Little understanding of GDPR	
	Lack of institutional direction	
	Lack of co-creation	
	Influence student behaviour with intelligence gathered from data.	

Fig. 4. Summary of findings

The potential value of BD to support student-retention is recognised by most participants, with caveats that data should be viewed within the context of a students' circumstances and should not replace the professional staff-student relationships. However, other benefits such as changing pedagogy and personalised learning were not mentioned as reported in earlier works [3]. This research partially supports a view which reports that institutional adoption of analytics is hindered by lack of a data-driven mindset [35]. All participants demonstrated a data-driven mind-set and an acknowledgement that data has value; however, academics reported the lack of available data. In summary, the understanding and value of BD within HE is recognised, but a clear institutional-strategy regarding data usage in LA is needed [23, 24, 36, 50].

4.2 The Current Use of Big-data

Academics responses were mixed about the use of data within their role and their institutions. Some reported that they do not use it, but later spoke about activities they undertake, using student personal data they have collected. It is interesting to observe that participants did not class student personal data as 'BD'. In general, most participants were aware of institutional data being used to monitor and report an aggregated performance of students. Two academics reported that their institutions' use attendance data to inform interventions preventing student withdrawals.

In summary, the sector uses BD to support student-retention and engagement activities. Some HEIs use attendance monitoring data, not predictive data, to trigger interventions. Hence, it can also be argued that the use of one set of data does not fit with earlier suggested definitions [27]. Although vast amounts of data are collected across the sector, this research finds that the lack of a data strategy is common across the HE sector.

4.3 Impact of the GDPR

From the participants' perspectives, the impact of GDPR is minimal. Most referred to changes to data-storage on their systems, one participant referred to restrictions on data usage as a result of GDPR. Albeit GDPR does permit analysis of data, it is the actions taken as a result of the data that are affected [22]. There appears to be misconceptions within the sector between academic and IMs regarding the implications of the GDPR. The IMs reported positive impacts, including clarity on data usage and consent. Across the sector, an impact of the GDPR has been the changes in administrative practice at the point of obtaining student consent to collect and use their data. The IMs cited another impact with the introduction of new roles i.e. Chief Information Officer and Head of Data Governance; both roles were identified due to the need for expertise and knowledge of data management and accountability within their institutions.

4.4 Institutional Research on Student-Retention

Although the need to understand student-retention has been discussed in literature for more than four decades [45, 46, 49], this research indicates limited research to understand factors affecting student-retention. Only one participant reported institutional research that subsequently informed changes to their practice. Participants mentioned small-scale internal projects, but no impact or change to practice was reported. Without Institutional knowledge of such factors, data are incomplete and therefore analysis will be subject to misinterpretation and bias [5]. Earlier research [32, 51], suggests that understanding an individual's motivations for studying, using behavioural data i.e. motivations, critical thinking and social-emotional well-being, enhances the accuracy of predictions of student withdrawal or attainment. However, this was not mentioned by any participant.

4.5 Data Presentation and Support for Staff and Students

The presentation of individual student data to staff appears inconsistent. All participants acknowledged that a large amount and variety of student data are collected, but not all is presented to staff. However, if presented, some concerns were expressed by academics regarding misinterpretation of data, leading to inconsistent practice. HEIs should consider comprehensively the provision of guidance and training for use of BD.

One participant referred to 'data-experts', implying that they would not need or require training. However, if HEIs were to introduce a LA solution, the use of data would be very different to current practice. Training for all staff that access and use such data would be critical for effective implementation. This research finds that very little collaboration with staff or students in the development of any solution using student personal data to support student-retention or attainment has taken place, in line with prior investigations [13, 17].

In summary, the prominent concerns raised during the interviews were a combination of lack of institutional direction and strategy regarding the use of data, limited knowledge of the GDPR, lack of co-creation with end users, personal ethical and moral perspectives of how student data should be used. The overall perspective of participants was that students should be entitled to see their data used at institutional level.

4.6 Use of Predictive Learning Analytics

As stated above, participants were in favour of presenting students with their data, although concerns were expressed when asked about showing predictions to students. To have their data presented which predicts their withdrawal or failure could be seen as demotivating, and possibly inaccurate if based on a stereotypical approach of categorising students. Some participants expressed their desire to talk to their students before presenting predictive data, whereas others were adamant that students shouldn't see. These concerns relate to a lack of knowledge regarding how PA works and a lack of transparency in the predictive modelling algorithm, as discussed in earlier works [29].

4.7 Developing Learning Analytics

Participants reported several common perspectives, including the need for a coherent institutional approach and policy, clear guidance and support for users of LA data, and collaboration with staff and students is important. Although it is suggested that students should be engaged as collaborators with a LA solution [9], this research suggests that academics should also be involved - specifically with clarifying the institutional purpose. This inclusion could provide the assurance and address academic concerns regarding the ethical and erroneous use of BD [44, 57].

It should also be recognised that whilst LA could support student-retention, it could also be used to inform institutional planning, teaching and learning practice. Use of LA in this context would lead to innovation and change, which as a result of institutional resistance to change could be considered a risk [16, 33].

4.8 Ethical Issues of Using Big-data

This investigation finds the surveillance and profiling of students is a concern for academics; as suggested in earlier works [29]. Personal observations by academics of their students also suggested that student behaviours do not always follow the path that data have predicted; to them it is more important to retain the personal relationship.

Participants also expressed concerns regarding individuals' access to, and use of LA data, as this was seen as the most variable risk. The digital capabilities and confidence to diagnose a student's situation and take follow-up actions was highlighted as a very individual undertaking and where significant inconsistences would occur. The findings indicate mixed practice within the sector regarding the sharing and use of BD, with an underlying desire for an approach to adopting LA.

5 Conclusion

This investigation of whether post the GDPR, BD could be used within HE. The study has generated evidence that BD based LA is, or has been, used within HEIs to mainly support student-retention. However, only one HEI is currently using a single source of attendance monitoring data to support student-retention, whereas other HEIs are using BD additionally for attainment, management information, business-modelling and quality assurance. This research concludes that the implementation of GDPR has had little impact on existing practice within UK HEIs in their use of data, academic participants only reporting changes to practice in data-storage. IMs cited that the GDPR was a positive move enabling greater clarity on data collection and usage. Participant responses indicated there is a gap in knowledge and application of GDPR. The introduction of the GDPR had had an impact on staffing levels at one HEI with the appointment of a Chief Information Officer and a Head of Data Governance with responsibility for data governance and compliance.

In general, all participants described various possible datasets for predictive modelling, although all expressed concerns regarding its application. Despite the literature available in the field, a larger study would conclusively indicate how BD identifies students at risk of withdrawal. All participants cited that it would be beneficial to present student data to academics to support and inform their role in providing academic guidance. Several benefits were cited, including: being able to see collated student profile information and students' course engagement. Additionally, participants also expressed concerns relating interpretation of data by colleagues, the perceived volume of work and the impact on other areas of the University to support students. Overall, participants believed that the presentation of student data would be of significant benefit for academics, but training and support would be required to ensure a consistent institutional approach to support students.

5.1 Limitations

The sample size is the main limitation to this study. Two IMs and seven academics do not represent the HE sector; hence a larger comprehensive study would offer more insight. Participant responses did not differentiate between types of student, for example: year of study, undergraduate, postgraduate, distance-learning, part-time.

5.2 Recommendations

While the research supports the power and use of BD, it is apparent how this knowledge is translated into interventions, and whether these interventions are effective at supporting students, are key questions. The research indicates that the use of BD to support student-retention post the GDPR is possible, but not in isolation; it is the actions and interventions that have an impact, together with student engagement with their academic community and the willingness to respond to guidance that maybe drawn from their LA data. Implementation of LA must be supported by:

- Co-creation of a LA approach designed with staff and students.
- A legal and ethical institutional-strategy, and purpose for using BD, informed by appropriate investigations.
- Commitment to data-quality and the collection of relevant datasets to accurately inform the LA solution.
- Commitment to enhance digital capabilities of staff.
- A framework of training and support for the role of Academic Advising that includes the GDPR.

- Change management plan to identify and address cultural issues.
- Commitment by institutional leadership to adequately resource the support services required to deliver interventions to all students that would benefit their academic journey.

6 References

1. Alblawi, A. S., & Alhamed, A. A.: Big-data and learning analytics in higher education: Demystifying variety, acquisition, storage, NLP and analytics. In *2017 IEEE Conference on Big-data and Analytics (ICBDA)* (pp. 124-129). IEEE. (2017, November).

2. Athey, S.: Beyond prediction: Using Big-data for policy problems. *Science (New York, N.Y.), 355*(6324), 483 (2017).

3. Baker, R., & Yacef, K.: The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM | Journal of Educational Data Mining*, *1*(1), 3-17 (2009). Retrieved from https://tinyurl.com/y65m9cel

4. Bertolucci, J.: *Big-data: A practical definition.* San Francisco: UBM LLC. (2013).

5. Bienkowski, M., Feng, M., & Means, B.: *Enhancing Teaching and Learning Through Educational Data Mining and LA: An Issue Brief.* Office of Educational Technology, U.S. Department of Education (2012). Retrieved from https://tinyurl.com/y6znls4l

6. Boulton, C. A., Kent, C., & Williams, H. T. P.: Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar' university. *Computers & Education; Computers & Education, 126*, 129-142 (2018).

7. Boyd, D., & Crawford, K.: Critical Questions for Big-data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679 (2012).

8. Braun, V., & Clarke, V.: Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101 (2006).

9. Buchanan, E. A.: *Internet research ethics: Past, present, and future* Wiley-Blackwell (2011).

10. Card, N. A.: *Applied meta-analysis for social science research*. Guilford Publications (2015).

11. Daniel, B.: Big-data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology, 46*(5), 904-920 (2015).

12. Denzin, N., & Lincoln, Y. S.: The SAGE Handbook of Qualitative Research, Thousand Oaks, CA Sage (2011)

13. Dollinger, M., & Lodge, J.M.: *Co-creation strategies for LA*. In Proceedings of the 8th International Conference on LA and Knowledge (LAK '18). ACM, New York, NY, USA, 97-101 (2018).

14. European Data Protection Supervisor (2018). Retrieved from https://tinyurl.com/ydy882pe

15. Farah, B.: Big-data - what data and why? *Journal of Management Policy and Practice*, *17*(1), 11-17 (2016).

16. Faraj, S., Pachidi, S., & Sayegh, K.: Working and organizing in the age of the learning algorithm. *Information and Organization, 28*(1), 62-70 (2018).

17. Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S.: *Setting LA in context: Overcoming the barriers to large-scale adoption* Association for Computing Machinery (2014).

18. Foster, J., McLeod, J., Nolin, J., & Greifeneder, E.: Data work in context: Value, risks, and governance. *Journal of the Association for Information Science and Technology*, *69*(12), 1414-1427 (2018).

19. Greller, W., & Drachsler, H.: Translating learning into numbers: A generic framework for LA. *Educational Technology & Society*, *15*(3), 42-57 (2012).

20. Gűlbahar, Y, & Ilgaz, H.: Premise of LA for Educational Context: Through Concept to Practice. Bilişim Teknolojileri Dergisi, 7 (3) (2014). Retrieved from https://tinyurl.com/y29uueem

21. Holloway, I., & Todres, L.: The status of method: Flexibility, consistency and coherence. *Qualitative Research, 3*(3), 345-357 (2003).

22. Information Commissioner's Office (ICO):. *Guide to the General Data Protection Regulation (2018).* Retrieved from https://tinyurl.com/y9jpbxmh

23. Jisc.: *LA in Higher Education: A review of UK and international practice* Full report (2016). Retrieved from https://tinyurl.com/j4qcasg

24. Jisc.: *Code of practice for LA. Setting out the responsibilities of educational institutions to ensure that LA is carried out responsibly, appropriately and effectively*(2018). Retrieved from https://tinyurl.com/huwpqrm

25. Johnson, S. L., Gray, P., & Sarker, S.: Revisiting IS research practice in the era of big-data. *Information and Organization*, *29*(1), 41-56 (2019).

26. Jokhan, A., Sharma, B., & Singh, S.: Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, 1-12 (2018).

27. Kune, R., Konugurthi, P. K., Agarwal, A., Chillarige, R. R., & Buyya, R.: The anatomy of BD computing. *Software: Practice and Experience*, *46*(1), 79-105 (201).

28. Landis, J. R., & Koch, G. G.: The measurement of observer agreement for categorical data. *Biometrics*, 159-174 (1977).

29. Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J.: Identification of 'at risk' students using LA: The ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology, Research and Development*, *64*(5), 957-968 (2016).

30. Lerman, J.: Big-data and its exclusions. *Stanford Law Review Online*, *66*, 55.63 (2013). Retrieved from https://tinyurl.com/y69qkogk

31. Lim, S., Woo, J., Lee, J., & Huh, S. Y.: Consumer valuation of personal information in the age of big-data. *Journal of the Association for Information Science and Technology*, *69*(1), 60-71 (201).

32. Liu, M., Kang, J., Zou, W., Lee, H., Pan, Z., & Corliss, S.: Using data to understand how to better design adaptive learning. *Technology, Knowledge and Learning*, *22*(3), 271-298 (2017).

33. Macfadyen, L. P., & Dawson, S.: Numbers are not enough. why e-LA failed to inform an institutional strategic plan. *Educational Technology & Society, 15*(3), 149-163 (2012).

34. Maciejewski, M.: To do more, better, faster and more cheaply: Using BD in public administration. *International Review of Administrative Sciences, 83*(1), 120-135 (2017).

35. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H.: Big-data: The next frontier for innovation, competition and productivity. McKinsey Global Institute (2011). Retrieved from https://tinyurl.com/y5rrh58x

36. Mayer-Schönberger, V., & Cukier, K.: *Big-data : A revolution that will transform how we live, work and think*. London: John Murray (2013).

37. Miltgen, C. L., & Smith, H. J.: Falsifying and withholding: exploring individuals' contextual privacy-related decision-making. *Information & Management*, *56*(5), 696-717 (2019).

38. Na, K. S., & Tasir, Z.: Identifying at-risk students in online learning by analysing learning behaviour: A systematic review. In 2017 IEEE Conference on Big-data and Analytics (ICBDA) (pp. 118-123). IEEE (2017, November).

39. Nersessian, D.: The law and ethics of BD analytics: A new role for international human rights in the search for global standards. *Business Horizons; Business Horizons, 61*(6), 845-854 (2018).

40. Pascarella, E. T., & Terenzini, P. T.: Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *The Journal of Higher Education, 51*(1), 60-75 (1980).

41. Petticrew, M., & Roberts, H.: *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons (2008).

42. Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C.: Student attitudes toward LA in higher education: "the fitbit version of the learning world".(report)(author abstract). *Frontiers in Psychology, 7* (2016).

43. Saunders, M., Lewis P., Thornhill A.: *Research methods for business students* (Seventh edition. ed.). Harlow: Pearson (2016).

44. Slade, S., Prinsloo, P., Haythornthwaite, C., de Laat, M., & Dawson, S.: Learning Analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510-1529 (2013).

45. Social Market Foundation.: *On Course for Success? Student-retention at university* (2017). Retrieved from https://tinyurl.com/y6nuolju

46. Spady, W.: Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange, 1*, 64-85 (1970).

47. Subotzky, G., & Prinsloo, P.: Turning the tide: A socio-critical model and framework for improving student success in open distance learning at the university of south africa. *Distance Education, 32*(2), 177-193 (2011).

48. Tempini, N.: Till data do us part: Understanding data-based value creation in data-intensive infrastructures. *Information and Organization, 27*(4), 191-210 (2017).

49. Tinto, V.:Research and practice of student-retention: What next? *Journal of College Student-retention, 8*(1), 1-19 (2006).

50. Universities UK: *The Funding Environment for Universities 2015. The Economic Role of UK Universities (2015, June).* Retrieved from https://tinyurl.com/yyzrgxqy

51. Van, D. Z., Denessen, E., Cillessen, A. H. N., & Meijer, P. C.: Domains and predictors of first-year student success: A systematic review. *Educational Research Review; Educational Research Review, 23*, 57-77 (2018).

52. Wellington, J. J.: *Educational research: Contemporary issues and practical approaches* (Second edition. ed.). London: Bloomsbury Publishing (2015).

53. Williams, P.: Squaring the circle: A new alternative to alternative-assessment. *Teaching in Higher Education, 19*(5), 565-577 (2014).

54. Williamson, B.: The hidden architecture of higher education: Building a BD infrastructure for the 'smarter university'. *International Journal of Educational Technology in Higher Education, 15*(1), 1-26 (2018).

55. Wilsdon, J.: *Deliver us from rankers*.(2019, April). Retrieved from https://tinyurl.com/y4exo8y6

56. Wu, P. F., Vitak, J., & Zimmer, M. T.: A contextual approach to information privacy research. *Journal of the Association for Information Science and Technology* (2019). In-press.

57. Xie, K., Wu, Y., Xiao, J., & Hu, Q.: Value co-creation between firms and customers: The role of big-data-based cooperative assets. *Information & Management*, *53*(8), 1034-104 (2016).