Open Research Online

The Open University's repository of research publications and other research outputs

Making sense of learner and learning Big Data: reviewing 5 years of Data Wrangling at the Open University UK

Journal Item

How to cite:

Rienties, Bart; Cross, Simon; Marsh, Vicky and Ullmann, Thomas (2017). Making sense of learner and learning Big Data: reviewing 5 years of Data Wrangling at the Open University UK. Open Learning: The Journal of Open, Distance and e-learning, 32(3) pp. 279–293.

For guidance on citations see [FAQs](#).

© [not recorded]

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1080/02680513.2017.1348291>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Making sense of learner and learning Big Data: reviewing 5 years of Data Wrangling at the Open University UK

Bart Rienties*, Simon Cross, Vicky Marsh, Thomas Ullmann

Open University UK, Institute of Educational Technology, Milton Keynes, United Kingdom;

*Corresponding author. Email: bart.rienties@open.ac.uk

Abstract

Most distance learning institutions collect vast amounts of learner and learning data. Making sense of this “Big Data” can be a challenge, in particular when data are stored at different data warehouses and require advanced statistical skills to interpret complex patterns of data. As a leading institute on learning analytics, in 2012 the Open University UK (OU) instigated a Data Wrangling initiative. This provided every Faculty with a dedicated academic with expertise data analysis and whose task is to provide strategic, pedagogical, and sense-making advice to staff and senior management. Given substantial changes within the OU over the last 18 months (e.g., new Faculty structure, real-time dashboards, two large-scale adoptions of predictive analytics approaches, increased reliance on analytics), this embedded case-study provides an in-depth review of lessons learned of 5 years of data wrangling. Using semi-structured interviews with key stakeholders (10 senior managers/associate deans) and ten Data Wranglers (DWs), a clear mismatch was identified in terms of resources, expertise, and skills that can effectively address key needs from Faculties. Furthermore, inconsistencies in terms of reporting and responding to bespoke requests were noted by stakeholders. Given the essential role of DW for the OU, a new DW structure is proposed to ensure effective provision of in-depth, evidence-based data analyses, pedagogical insight, and actionable advice for Faculties. We will elaborate on the design of the new structure, its strengths and potential weaknesses, and affordances to be adopted by other institutions.

Keywords: learning analytics, big data, data wranglers, qualitative research

Making sense of learner and learning Big Data: reviewing 5 years of Data Wrangling at the Open University UK

Introduction

Many distance learning institutions and an increasing number of “traditional” universities have access to a wealth of data about their learners, which may provide a treasure trove to understand how their students are learning and how to optimise the institutional business processes ([Ferguson et al., 2016](#); [Sharples et al., 2016](#)). With an increasing amount of relatively static learner (e.g., demographics, previous educational) and more fluid, dynamic learning data (e.g., clicking behaviour, engagement in discussion forums, assessment contributions), educational institutions are provided with renewed opportunities to increase performance and retention and provide personalised learning on a large scale ([Bienkowski, Feng, & Means, 2012](#); [Tempelaar, Rienties, & Giesbers, 2015](#)).

For many distance learning institutions, making strategic sense and use of large amounts of data is common practice (e.g., [Ashby, 2004](#); [Calvert, 2014](#); [Richardson, 2006](#); [Simpson, 2004](#)). However, with the increased availability of data about students and their learning behaviour, advanced predictive modelling and visualisations, most distance learning institutions are progressively making use of principles of Learning Analytics (LA) to understand how to improve the core business ([Author A, 2016c](#); [Clow, 2013](#); [Ferguson et al., 2016](#); [Macfadyen, Dawson, Pardo, & Gasevic, 2014](#); [Sharples et al., 2014](#); [Sharples et al., 2016](#); [Tempelaar et al., 2015](#); [Verbert et al., 2011](#)). According to [Clow \(2014\)](#), “learning analytics is widely seen as entailing a feedback loop, where actionable intelligence is produced from data about learners and their contexts, and interventions are made with the aim of improving learning.”

Even though LA is a relatively young field of research, several researchers have found that sophisticated learning analytics algorithms, visualizations, and data are at times difficult to make sense of and to interpret for non-specialists like teachers (Buckingham Shum et al., 2013; Drachsler & Greller, 2016; Gasevic, Rosé, Siemens, Wolff, & Zdrahal, 2014). As argued by Drachsler and Greller (2016, pp. 89-90) “A big challenge for Learning Analytics in this respect is the complexity of the data collection and algorithmic analysis processes. The applied technologies are not trivial and it can be rather difficult to provide non-technical educational stakeholders (learners, teachers, managers, and external parties like education authorities or parents) with an understanding of how and what data are being collected, how they are processed, and how reliable the results of the analysis are.”

One approach suggested is to help non-specialists to benefit from insights in learning analytics is to implement a structure of Data Wranglers (DWs). Data wrangling provides strategic, pedagogical, and sense-making advice to Faculties. According to Clow (2014), “DWs are a group of academics who analyse data about student learning and prepare reports with actionable recommendations based upon that data”. In practice, beyond data analyses DWs provide important “translation services” of Big Data insights to help teachers, instructional designers, curriculum managers, and senior management to make sense of complex data and to provide actionable insight and pedagogical support. With nearly five years of experience with DW, given some substantial changes at the OU (e.g., new Faculty structure, availability of real-time dashboards, increased reliance on analytics) using an embedded case-study methodology (Author A, 2015a; Yin, 2009) we will review how effective this DW structure was in helping non-specialists to understand learning analytics data and how to transform their practice. Furthermore, we aim to provide a new conceptual approach to data wrangling which is based upon getting the balance right in terms of key expertise, diversity, and ambassadorship while at the same time providing a solid core of

analytics expertise. As the OU is the leading organisation in the UK in terms of learning analytics according to the Higher Education Commission (2016), we hope that by sharing our lessons-learned of our DW approach we will spark a discussion amongst teachers, researchers, and managers how institutions can provide effective support structures to make sense of Big Data.

Data wrangling and intelligent use of data

Demand for actionable insights to help support teachers and senior management with module and qualification design is currently strong (Miller & Mork, 2013), in particular a desire for evidence of impact of “what works” (Ferguson et al., 2016). In a review of 66 learning analytics research papers between 2009-2015, Papamitsiou and Economides (2016) found an average positive effect of using learning analytics comparing quasi-experimental and RCT studies. However, in a large-scale review for the European Union, an inventory analysis by Ferguson et al. (2016, p. v) of 60 learning analytics tools, practices and policies highlighted that “there is a wide gap between the potential roles for learning analytics that are identified within the research literature as a whole and the dominant themes in learning analytics as they are put into practice by ICT/learning technology vendors, developers and researchers.” Similarly, in a recent review of 123 best-practice papers and research papers submitted to the LACE evidence-hub, Ferguson and Clow (2017) argued that “only two studies in the Evidence Hub provide evidence that analytics have prompted changes in teaching and support that have impacted on learners”. In other words, in the learning analytics literature there is a lack of evidence of what works, and more importantly what actually works in practice on a level of a module, faculty or institution as a whole.

In practice, teachers, managers, and researchers at most institutions have access to a substantial range of data about students and their learning behaviour. Over the last decade

several systems have been developed ([Ashby, 2004](#); [Inkelaar & Simpson, 2015](#); [Richardson, 2012](#)) for managing data tasks, such as logging assessment grades ([Author B, 2010a](#); [Calvert, 2014](#)), handling tutor feedback to students, monitoring VLE activity ([Tempelaar et al., 2015](#)), surveying students ([Ashby, 2004](#)) and capturing the pedagogic balances within a module ([Author B, 2012](#)). The nature of distance learning means that teaching at the OU is manifested primarily in three forms: in the embedded teaching and learning design in module materials, learning activities and assessment ([Author A, 2016d](#); [Conole, 2012](#)); in the detailed feedback tutors provide to each student after submission of each continuously assessed assignment, and in the direct distance teaching delivered by associate lecturers to their tutor groups (normally consisting of between 15-20 students) in (increasingly online) group tutorials ([Wolff, Zdrahal, Nikolov, & Pantucek, 2013](#)).

Learning analytics are now increasingly taken into consideration at the OU when designing, writing and revising modules, and in the evaluation of specific teaching approaches and technologies ([Author A, 2016b](#)). A range of data interrogation and visualization tools developed by the OU supports this ([Author A, 2016d](#); [Author B, 2012](#); [Calvert, 2014](#)). The use of such data can range from investigating the student experience of assessment, and contrasting the effectiveness of changes in learning design with respect to observable activity on the VLE ([Author A, 2016a, 2016b](#)), to investigating retention and learning issues associated with concurrent study (studying more than one distance learning module at once), collaborative learning, online tuition, and wikis.

DW work fits within existing annual cycles of data reporting and monitoring. Most notably, the publication of an annual Key Metrics Report (KMR) about module performance and profile information in the autumn provides DWs with specific data with which to engage Faculties. Based upon these KMRs, DWs develop specific bi-annual DW reports per Faculty

that provide key data insight and pedagogical advice to improve teaching and learning. According to [Clow \(2014\)](#), “data from these sources is aggregated using a SAS data warehouse, and exported to a Tableau workbook for each Faculty. The DW use these workbooks as their primary data investigation tool, and to generate some charts and visualization, but also use the data sources directly where appropriate, and produce their own charts in excel”. In some instances, bespoke datasets have been created for additional analysis. In a conceptual model of DW, [Clow \(2014\)](#) indicated that initially a DW helps to present data from all sources to all users. In a mature stage of Data Wrangling, many users will make use of data sources directly and the DW is facilitating those users who need additional support, as illustrated in Figure 1.

➔ Insert Figure 1 about here

An informal evaluation amongst 22 stakeholders in 2013 amongst seven Faculties indicated that many were positive about the support provided by DWs, whereby the reports provided good starting points for discussion about how to improve pedagogy and the core business ([Clow, 2014](#)). At the same time, the quality of the reporting was varied and heavily dependent on the individual skills of the DW and his/her relation to the respective Faculty. Similarly, the quality of output was dependent on the quality of data analysis, which in 2013/4 was heavily relying on descriptive data and visualizations ([Clow, 2014](#)).

In the last three years, like many distance learning institutions the OU has made substantial investments in learning analytics, whereby there is a surge of new data-intensive approaches and units working to understand how to maximize student retention (see for example [Author A, 2016a, 2016d, 2016e](#); [Calvert, 2014](#); [Wolff et al., 2013](#)). For example, the new OU University Strategy focusses on *Students First*, whereby resources should focus on what helps students to successfully complete a module and degree ([Open University, 2016](#)).

For five years DWs were at the forefront of considering data from different sources (e.g., student satisfaction data, VLE data, learning design, retention), the landscape of the OU and learning analytics in particular has rapidly evolved. First of all, with the increased focus on retention and learning analytics, a range of OU initiatives have increasingly put actionable, (near) real-time data, visualisations, and insights at the forefront of pedagogical change. For example, in Analytics4Action ([Author A, 2016a](#)) teachers and module teams are provided with near-real time data about the students' journeys, and are encouraged to pro-actively react within presentation on trends on data, or redesign key "bottlenecks" for the next implementation. Working together in small teams with pedagogical experts, technologists, and instructional designers, module teams can actively identify key bottlenecks in learning design based upon learning analytics data and insight, and where needed address these in presentation. This agile way of module production and implementation is a rather radical departure from previous OU module design, whereby most resources were dedicated to designing a new module, with limited opportunities for redesign ([Author A, 2016a](#)). In a parallel development, OU teachers now have access to near-real time predictions of whether students are going to submit their next assignment ([Hlosta, Herrmannova, Zdrahal, & Wolff, 2015](#)) and whether they will still be present at the next fee point ([Calvert, 2014](#)). As a result of real time dashboards and innovation processes like Analytics4Action, many module teams, teachers and senior management are actively engaging with real-time data ([Author A, 2016a](#)). Second, with the increased demand for data interpretation and understanding, many Schools and Faculties have hired their own data interpreters and/or involved other data experts to provide quick, discipline specific insight. Third, the OU organizational structure has substantially changed, from moving from seven to four Faculties, and the move to bring innovation units together in one Learning and Teaching Innovation portfolio.

However, the advent of real-time data and dashboards is resulting in a need for Faculties to understand the reasons behind perceptible changes, the increasing quantity of data requires greater expertise in the selection the “most appropriate” data for analysis, and in response to the Faculty restructure potential gaps in awareness of previous trends may emerge. In a way, the central position of the DW in Figure 1 might have changed in the last three years as a result of these organizational change processes and practice. Therefore, in this study we will reflect on five years of Data Wrangling and how institutions might use the lessons-learned from the OU to support the increased need to provide timely data insight and intelligence that is tailored to the users’ needs.

Research Questions

In this mixed method study, an embedded case study at the OU was undertaken to determine what worked well in our DW provision after five years according to key stakeholders and users of DW output. An embedded case-study approach ([Yin, 2009](#)) examines the characteristics of a single individual unit (recognising its individuality and uniqueness), in our case data wranglers, senior management and the OU organisation. [Yin \(2009\)](#) emphasised that a case study investigates a phenomenon in-depth and in its natural context. A case-study approach is particularly useful when the “problem” under investigation is complex, interlinked, and agents might have different perspectives and interpretations of the problem. Therefore, the purpose of a case study is to get in-depth information of what is happening, why it is happening and what are the effects of what is happening ([Author A, 2015a](#)). Reviewing five years of data wrangling at the OU, we will address the following questions:

RQ1: How has data wrangling worked in practice, and in particular what has worked well?

RQ2: What elements of the data wrangling provision can be improved, and why?

RQ3: What “ways of working” by data wranglers could further strengthen strategic, pedagogical, and sense-making advice of Big Data to staff and senior management

Method

Setting

The Open University UK (OU) is one of the front leaders in experimenting, testing, evaluating, and implementing learning analytics at scale (Higher Education Commission, 2016). Most of its 400 modules are provided in a distance learning format to its 170.000+ students. A unique feature of the OU is its open entry policy, whereby any learner can start at the OU. With a change in government funding (Universities UK, 2013), the OU is under increased pressure to ensure that not only the OU is open to diverse groups of learners with potentially lower previous qualifications, but also to ensure that those students successfully complete a degree.

Instruments

Semi-structured interviews

23 interviews were conducted (10 senior managers; 3 professional support staff; 10 data wranglers) lasting on average 30 minutes, whereby a semi-structured interview approach (Lichtman, 2013) was followed consisting of five key questions (i.e., 1. What do you consider to be your role as data wrangler? 2. What do you think works well in data wrangling? 3. What do you think can be improved in data wrangling? 4. If you would redesign the structure of data wrangling, how would it look like? 5. What statistical techniques and/or qualitative approaches are you proficient in?). Depending on the role of the stakeholder the questions were slightly altered to match the respective context. The first author (who had no prior involvement with DW) conducted all interviews, and afterwards the

first and second author (a current DW) analysed the transcribed qualitative data independently from each other using emergent themes analysis ([Lichtman, 2013](#)), before the analysis results were compared and contrasted arriving at a final result. By independently conducting the interviews by an “outsider” and contrasting the initial findings and themes with an experienced DW, we followed qualitative research guidelines of credibility and trustworthiness by [Twining, Heller, Nussbaum, and Tsai \(2017\)](#). All data was anonymized by the first author to ensure that no interviewee could be identified. Therefore, we cannot report on the specific job role, gender, and Faculty of a particular interviewee, as this would be easily identifiable for people involved in data wrangling.

Document analysis

In line with [Bowen \(2009\)](#) and recommendations by [Yin \(2009\)](#) to use rich and diverse sets of data, the authors also used document analysis to compare and contrast previous 30+ bi-annual DW reports, bespoke requests, notes, KRMs, and supplementary materials (e.g., reports and publications on learning analytics projects and practices at the OU). The document analysis helped to identify common patterns as well as difference in practice.

Results

RQ1 Data wrangling in practice and what worked well

There was a general agreement amongst the ten DWs about the role of the Data Wrangler and what it entails in practice during the interviews. Several DWs indicated that DWs provide a bridge and act as a communicator between Faculties and data. There was a need to respond to Faculties to investigate data, what matters to them, solving data queries and problems that Faculty staff have that provide value. Data Wrangler 4 (DW4) indicated that DWs are a first point of contact for Faculties in terms of data insight. Similarly, DW1 indicated that DWs work together with senior managers on what data is leading to positive and negative learning experiences and retention, while at the same time addressing specific requests from module teams and senior managers with more complex questions across/within

modules and programs. DW6 highlighted that capacity building for Faculty staff to interpret and to make sense of data was important. Stakeholder 2 indicated that the relationship with the Faculty was essential, whereby DWs should provide understanding of data, and support bespoke services to the Faculty where it is practical.

While many stakeholders indicated a similar perspective on the role of DWs, all stakeholders interviewed indicated a mismatch between the intended role of DW and practice. Stakeholder 6 mentioned that “I thought that DWs would answer strategic questions and provide deep-dives into data based upon requests from Faculties. In addition, I thought that DWs would bring in pedagogical expertise and recommendations to help Faculties. The actual practice is focused on describing data rather than analysing interlinked data, whereby there are real inconsistencies in measurements used and approaches. Furthermore, there is limited reference to other OU programs and projects that work along similar paths and notions to serve the university.” Similarly, Stakeholder 4 indicated that DWs were expected to benchmark at an institutional level and provide recommendations based upon evidence from other Faculties.

Stakeholder 7 indicated that “in theory the role of the DW is to interrogate data on a bi-annual manner and work together with the Faculty to investigate key issues and triangulate data sets that lead to pedagogical insights. In practice there is a complete spectrum of approaches, whereby only a few DWs get to grips with one or two data sources. Furthermore, there is a lack of engagement with the "right" people in Faculty”. Stakeholder 3 indicated that DWs need to be able to understand the details of data and be able to present and communicate these details in a coherent manner to staff, who have mixed abilities to interpret data and visualizations.

Most DWs indicated that their relationship with their respective Faculty was essential, and mostly this worked well. Several DWs had regular meetings with Faculty staff and senior managers, which helps DWs to be a “translator” of the respective Faculty to get to the right question. DW9 indicated that “[our unit] has loads of data and there is a clear need to make sense of the data, as there are often mismatches in results and understandings”.

All DWs indicated that the ambassador role to the central unit within the Faculties is important. This ambassador role was essential to develop trust with the Faculties to help them to interpret the bi-annual reports and to formulate clear bespoke requests that can be effectively answered. This was also highlighted by the stakeholders, who regarded the client relationship to work well to reasonably well in some Faculties, while in other Faculties the experiences were rather mixed or not meeting expectations.

RQ2 What can be improved in Data Wrangling

Several DWs indicated time pressures and resource issues to complete the bi-annual reports in time and to report quickly and effectively to bespoke requests. In particular dealing with the revision requests of the first drafts of the bi-annual reports by the Faculties were considered to be labour intensive. Furthermore, several DWs indicated that it was difficult to get access to the “right” datasets to answer key questions from Faculties. In particular, our analysis of the interviews and reviewing the bi-annual reports across the various Faculties indicated that most analyses were focused on individual, cross sectional data sets (e.g., looking at retention data within one Faculty in autumn semester 2014), rather than conducting advanced quantitative analyses (e.g., Regression, Structural Equation Modelling, Multi-level analysis) to link various datasets together (e.g., demographics, VLE, Satisfaction and retention across modules of several Faculties between 2013-2016) to get a thorough understanding of the underlying factors of a particular retention percentage, and whether these factors have changed over time (Arbaugh, 2014; Author A, 2016c, 2016e).

Related to this point, there is a strong divergence in terms of statistical skills and expertise amongst DWs. While all DWs have an understanding of basic descriptives (e.g., Mean, percentage), only four indicated to be proficient in correlation analysis and linear regressions analysis. Three DWs were familiar and confident with some of the advanced statistical methods (e.g., cluster analysis, factor analysis, logistical regressions, longitudinal analysis, machine learning, multi-level analysis, path analysis, predictive analytics, structural equation modelling), and one DW was proficient in all these techniques. Nine out of ten DWs were comfortable and proficient with qualitative analysis (e.g., content analysis, interviews, discourse analysis), although with different levels of insights and application in practice. Perhaps surprisingly, DWs were not working together to share good practice, research questions, and statistical expertise and syntaxes. Several senior managers indicated that DWs should work both on a descriptive, narrative level as well as with key experts within the Faculty to make sense of the more complex relations using advanced statistical analysis.

There was a strong divergence in terms of receiving and reacting to bespoke requests from Faculties. Several senior managers indicated a lack of responsiveness to queries by the Faculty due to lack of resources available to DWs, availability, and expertise. At the same time, some Faculties received excellent service from DWs and senior managers were approaching these DWs whom were not “their” DW for specific requests. Stakeholder 1 indicated that DWs needed to help Faculties with data understanding and concrete action plans (i.e., what does this mean, what can I do, how does this help with Annual Quality Review (AQR))? There is a blurring of boundaries between what DW are doing and other analytics providers/dashboards. Stakeholder 3 noted that it was not helpful just to provide statistics. For this Faculty, there was a need for some clear, coherent narratives to be able to provide interpretation of the data and what is behind it. Many of the bi-annual reports were rather overwhelming, and need to be more focused on what was relevant for the respective

Faculty. In particular, there were conflicting definitions and ways in which the various variables were used in different reports and datasets. In other words, there was a need for consistency. In its present form, several stakeholders questioned the quality and usefulness of the bi-annual reports. Given the increased availability of (near) real-time data and insight, our analysis indicated that for some Faculties the DW reports had relatively limited value, working on data that was 3-12 months old using just descriptives rather than advanced, deep insight of underlying trends, cause and effect, and longitudinal developments (Author A, 2016b; Gasevic et al., 2014).

Finally, the ambassador role of DW could be improved. Depending on the specific DW and Faculty it was an isolated client-service provision rather than a partnership between the Faculty and DW. As indicated by Stakeholder 4, with a true partnership this should lead to quality enhancements. Nearly all stakeholders indicated that DWs should be working in teams rather than working individually with one DW per Faculty. It would be useful to link one academic with one stats person, and making sure that pedagogical expertise is present. Similarly, all stakeholders indicated a desire to have a clear role division between the DW team and other providers of analytics and insight (e.g., Strategy and Information Office, Statistics Unit, Real Time Dashboards). Stakeholder 1 indicated that the DW team should focus on sound understandings of evidence and which types of interventions might work for a particular Faculty. DWs should provide clear, coherent reviews on what changes have been made by module teams and Faculties that have led to successful actions and interventions. Finally, there needs to be a stronger link to AQR that leads to action planning.

RQ3 New ways of working for data wrangling

➔ Insert Figure 2 about here

Based upon the interviews, document analysis, and evidence of good practice of strategic use of learning analytics (Ferguson et al., 2016; Miller & Mork, 2013), as indicated in Figure 2 it

is essential that DWs have a mix of three skills: ambassador, pedagogical understanding, statistical skills and databases. As ambassador, DWs provide a bridge between central units and Faculties and pro-actively work together to understand the key learning and teaching concerns of Faculties. Pedagogical understanding is needed to provide appropriate, evidence-based advice and to support Faculties in improving the learning experiences and learning outcomes of their students. By benchmarking and sharing lessons learned of what works within and across Faculties, effective pedagogical support will help to translate data insights into educational innovations and cost-savings. Finally, advanced statistical skills and a robust understanding of datasets that are available at the OU are essential to provide effective, in-depth and meaningful bi-annual reporting. These advanced statistical skills are also needed to respond quickly and robustly to bespoke requests from Faculties. As highlighted both in research ([Buckingham Shum et al., 2013](#)) and practice both outside ([Ferguson et al., 2016](#); [Ferguson & Clow, 2017](#)) and within the OU, not every data analyst will have each of these three broad skills and therefore it is essential to form expertise teams which have the right mix.

With the continued changes in OU strategies (e.g., maximizing student retention, Students First, Apprenticeships), changes within Faculties and schools, the increased availability of real-time dashboards, and changes in requirements for quality enhancements, the focus of DW should be flexible enough to meet the changing landscape of the OU. This is conceptually visualized by the blue box, which shapes will change over time depending on these four “external” factors (from a DW perspective).

A core team of Data Wranglers (CT) is formed which will design and implement all core statistical analyses for all Faculties and units for both the bi-annual reporting and bespoke requests. Working towards one generic syntax linking across datasets, this will lead to standardization of practice that is robust, reliable, and most importantly replicable. In

addition, the CT will work towards longitudinal rather than cross sectional perspectives which will allow DWs to provide insights of data over time, rather than at one single point in time, as illustrated in Figure 3.

One advantage of this CT approach is that it should save time for all DWs in the long run as experts across the three key DW skills will provide consistent analyses and high quality reporting. By working together in teams with the Faculties, each question or bespoke request will be added to the syntax and analyses will be completed not only for the respective Faculty but also across the institution. This should help to drive DWs by what the Faculties and OU want and provide comparison, lessons-learned, and benchmarking. The syntaxes will be updated on a bi-weekly basis and will be shared with other units (e.g., LTI, Strategy and Information Office). One potential risk would be to ensure sufficient capacity to include specific requests and contexts per Faculty, and how to ensure continued change.

Each DW will work in a team for a respective Faculty, whereby each team will be appropriately balanced in terms of the three skills of ambassadorship, pedagogical understanding, and advanced statistics. In addition to a coherent skills provision to each Faculty, by working in teams there will be more consistent delivery that is less dependent on individual academic time. Of course there will be some duplication of effort as teams of DW will need to attend meetings together, but this will anyway be necessary given the Faculty restructure. Another risk is the increased workload for CT members.

The “other” DWs will take the lead in terms of ambassadorship and pedagogical understanding for the respective Faculty. They will focus on gathering and formatting bespoke requests from the Faculty. Bespoke requests will be recorded by the DW on the monitoring system set up on SharePoint. These requests will be discussed, prioritized, and analysed by the CT. The monitoring system will allow a record of bespoke requests to be maintained, the work allocated to persons with the right skills and checked to ensure it is

delivered on time. The other DWs will provide in-depth conceptual analysis (e.g., literature review of key concepts, best practice, advice for action) and ensure that bespoke reporting is effective and efficient. In addition, the other DWs will lead the conceptual analysis and reporting on specific key topics from/across four Faculties (e.g., Accessibility success, Assessment and Feedback, Co-concurrency of study, Informal to formal, Learning design, Student demographics impacting retention, Student satisfaction, Widening participation success). For example, based on accessibility success scores in terms of retention and satisfaction by the CT across the different schools and Faculties, a DW could identify which specific good practices are used in Faculties with good accessibility success, and how these good practice can be translated to other schools and Faculties.

➔ Insert Figure 3 about here

Discussion

Many educational institutions and distance learning organizations in particular are increasingly using learner and learning data of their students to predict which students need additional support, and how business improvements can be made based upon Big Data principles (Drachsler & Greller, 2016; Ferguson et al., 2016; Miller & Mork, 2013; Papamitsiou & Economides, 2016; Tempelaar et al., 2015). While many organizations have recently started to use learning analytics (Ferguson et al., 2016; Ferguson & Clow, 2017), the Open University UK (OU) has been using large data for nearly two decades to improve the students' experience (Ashby, 2004; Calvert, 2014). As one of the first institutions, the OU instigated a Data Wrangling initiative in 2012, whereby a dedicated academic per Faculty provided learning analytics expertise and data insight to allow Faculties to make strategic, pedagogical, and sense-making decisions.

By sharing our findings from our embedded case-study nested with the OU, we hope to inspire other institutions to think strategically about how learning analytics sense-making and insights can be embedded in their own organisation. This embedded case-study provided an in-depth review of lessons learned of 5 years of data wrangling. Using semi-structured interviews with key stakeholders (10 senior managers/associate deans, 3 professional support staff) and ten Data Wranglers (DWs) and document analyses of previous reports, a clear mismatch was identified in terms of resources, expertise, and skills that can effectively address key needs from Faculties.

Most DWs indicated that their relationship with their respective Faculty was essential. Several DWs had regular meetings with Faculty staff and senior managers, which helped DWs to be a “translator” of the respective Faculty to get to the right question. As indicated by Buckingham Shum et al. (2013), it is essential that educational specialists are able to help senior managers and teachers to make sense of data, and to translate their questions into meaningful and insightful data analysis. At the same time as also indicated several learning analytics researchers (Buckingham Shum et al., 2013; Drachsler & Greller, 2016; Ferguson et al., 2016), not many people within universities have the “holy grail” of advanced statistical skills, pedagogical understanding, and ambassadorship to elicit the “right” questions from teachers and organizations, and the ability to answer these data questions in an appropriate manner.

Given the increased availability of (near) real-time data and insight within the OU, our analysis indicated a mismatch between what Faculties were expecting from DWs and what some DWs were providing. In part this was related to a wider mismatch of skills provision within the organization as a whole (Author A, In Press), while at the same time

recognizing that the holy grail of skills might need to be found by putting an appropriate mix of different people together to work in teams with Faculties.

Our answer to the mismatch between skills, people, and advanced insights is to form a core team of Data Wranglers (CT), which will design and implement all core statistical analyses for all Faculties and units for both the bi-annual reporting and bespoke requests. Working towards one generic syntax linking across datasets, this will lead to standardization of practice that is robust, reliable, and most importantly replicable. In addition, the CT will work towards longitudinal rather than cross sectional perspectives which will allow DWs to provide insights of data over time, rather than at one single point in time. An obvious risk of this standardization of practice might be that the data wrangling might become a bureaucratic rather than organic process. At the same time, by ensuring the same quality and using advanced statistical techniques nested within the expertise of the CT, over time we hope to provide more coherent, robust understanding and insights of Big Data to Faculties and teachers in particular.

Limitations

One obvious limitation is that people participating in interviews might not be able or willing to share their feelings, insights, and perspectives ([Lichtman, 2013](#); [Twining et al., 2017](#)). Similarly, in qualitative research the lens of the researcher is important in terms of interpreting the data and reporting the findings ([Lichtman, 2013](#)). By working with a relatively large number of stakeholders and all DWs, comparing interview notes between two coders who had different organizational roles and perspectives, and triangulating the interview data with document analysis, some of these potential biases might have been mitigated. Given that all stakeholders were in agreement that there was a clear need for data insight and pedagogical understanding from experts like DWs, but at the same time all

stakeholders indicated a mismatch between the intended roles of DWs and practice, a coherent narrative of change was evident.

Conclusions and ways forward

Providing evidence-based insights of “what works” and how learning analytics can help to empower teachers and managers is a key issue that many distance learning institutions are currently struggling with. After five years of Data Wrangling at the OU, our experience indicates that institutions who are starting with learning analytics should bring together data experts, statisticians, pedagogical experts, managers, and teachers in networked teams that work together to address the key challenges and opportunities. Relying on single expert-client relations might be a practical solution in the short-term, but in the longer term organizations need to put into place a clear strategy to bring the best people together, and provide appropriate professional development to ensure that people working across different disciplines and fields of expertise are able to communicate effectively.

The OU has implemented a new structure of Data Wrangling in October 2016. Rather than providing individual reports per Faculty using inconsistent data analyses and visualisations, the new Key Metrics Report now integrated and visualised all data per Faculty, and datasets and results are compared and contrasted longitudinally using R. Furthermore, a Scholarly insight report was generated by all data wranglers ([Author A, 2016f](#)), which provides state-of-the-art and forward looking insights into what drives OU students and staff in terms of learning and learning success. Several key cross-Faculty themes were identified that influenced OU students’ learning experiences, academic performance, and retention. Using Big Data analyses the Scholarly insight report focussed on how the OU designs modules, formative and summative assessments and feedback, helps students from informal to formal learning, and how these learning designs influenced student satisfaction and performance ([Author A, 2016f](#)). Preliminary anecdotal comments via email indicate that

stakeholders seemed positive about the new provision, but we will continue to monitor whether the new structure indeed provides a better match to meet the learning analytics insight needs of staff and Faculties.

Overall, our findings indicate that universities and distance learning institutions in particular have to critically assess how to provide sense-making and pedagogical advice from learning analytics data, as just having off-the-shelf analytics tools, predictive learning analytics engines and visualisations in place will not “automatically” improve our students’ experience and retention.

Author A 2015a (Rienties, Johan, & Jindal-Snape, 2015)

Author A 2016a (Rienties, Boroowa, et al., 2016)

Author A 2016b (Rienties, Cross, & Zdrahal, 2016)

Author A 2016c (Rienties & Toetenel, 2016)

Author A 2016d (Toetenel & Rienties, 2016)

Author A 2016e (Li, Marsh, Rienties, & Whitelock, 2016)

Author A submitted (Rienties, Herodotou, Olney, & Schencks, Submitted)

Author B 2010a (Tingle & Cross, 2010)

Author B 2012 (Cross, Galley, Brasher, & Weller, 2012)

References

- Arbaugh, J. B. (2014). System, scholar, or students? Which most influences online MBA course effectiveness? *Journal of Computer Assisted Learning*, 30(4), 349-362. doi: 10.1111/jcal.12048
- Ashby, A. (2004). Monitoring student retention in the Open University: definition, measurement, interpretation and action. *Open Learning: The Journal of Open, Distance and e-Learning*, 19(1), 65-77. doi: 10.1080/0268051042000177854
- Author A. (2015a). [details removed for peer review].

- Author A. (2016a). [details removed for peer review].
- Author A. (2016b). [details removed for peer review].
- Author A. (2016c). [details removed for peer review].
- Author A. (2016d). [details removed for peer review].
- Author A. (2016e). [details removed for peer review].
- Author A. (2016f). [details removed for peer review].
- Author A. (In Press). [details removed for peer review].
- Author B. (2010a). [details removed for peer review].
- Author B. (2012). [details removed for peer review].
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1-57.
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27-40. doi: doi:10.3316/QRJ0902027
- Buckingham Shum, S., Hawksey, M., Baker, R., Jeffery, N., Behrens, J. T., & Pea, R. (2013). Educational data scientists: a scarce breed *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 278-281). Leuven, Belgium: ACM.
- Calvert, C. (2014). Developing a model and applications for probabilities of student success: a case study of predictive analytics. *Open Learning: The Journal of Open, Distance and e-Learning*, 29(2), 160-173. doi: 10.1080/02680513.2014.931805
- Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6), 683-695. doi: 10.1080/13562517.2013.827653
- Clow, D. (2014). *Data wranglers: human interpreters to help close the feedback loop*. Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge.
- Conole, G. (2012). *Designing for Learning in an Open World*. Dordrecht: Springer.
- Cross, S., Galley, R., Brasher, A., & Weller, M. (2012). Final project report of the OULDI-JISC project: challenge and change in curriculum design process, communities, visualisation and practice. York: JISC.
- Drachsler, H., & Greller, W. (2016). *Privacy and Analytics—it's a DELICATE issue. A Checklist to establish trusted Learning Analytics*. Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh.
- Ferguson, R., Brasher, A., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., . . . Vuorikari, R. (2016). Research evidence of the use of learning analytics; implications for education policy. In R. Vuorikari & J. Castano-Munoz (Eds.), *A European Framework for Action on Learning Analytics*. Luxembourg: Joint Research Centre Science for Policy Report.
- Ferguson, R., & Clow, D. (2017). *Where is the evidence? A call to action for learning analytics*. Paper presented at the Proceedings of the 6th Learning Analytics Knowledge Conference, Vancouver.
- Gasevic, D., Rosé, C., Siemens, G., Wolff, A., & Zdrahal, Z. (2014). *Learning analytics and machine learning*. Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, Indianapolis, Indiana.
- Higher Education Commission. (2016). From bricks to clicks. In X. Shacklock (Ed.). London: Higher Education Commission.
- Hlosta, M., Herrmannova, D., Zdrahal, Z., & Wolff, A. (2015). OU Analyse: analysing at-risk students at The Open University. *Learning Analytics Review*, 1-16.

- Inkelaar, T., & Simpson, O. (2015). Challenging the 'distance education deficit' through 'motivational emails'. *Open Learning: The Journal of Open, Distance and e-Learning*, 1-12. doi: 10.1080/02680513.2015.1055718
- Li, N., Marsh, V., Rienties, B., & Whitelock, D. (2016). Online learning experiences of new versus continuing learners: a large scale replication study. *Assessment & Evaluation in Higher Education*. doi: 10.1080/02602938.2016.1176989
- Lichtman, M. (2013). *Qualitative research in education: a user's guide* (3 ed.). Thousand Oaks, CA: Sage.
- Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, 9(2), 17-28.
- Miller, H. G., & Mork, P. (2013). From Data to Decisions: A Value Chain for Big Data. *IT Professional*, 15(1), 57-59. doi: 10.1109/MITP.2013.11
- Open University. (2016). The Strategic Plan. Retrieved 19 December 2016, from <https://intranet7.open.ac.uk/collaboration/strategic-plan/SitePages/Home.aspx>
- Papamitsiou, Z., & Economides, A. (2016). Learning Analytics for Smart Learning Environments: A Meta-Analysis of Empirical Research Results from 2009 to 2015. In J. M. Spector, B. B. Lockee, & D. M. Childress (Eds.), *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy* (pp. 1-23). Cham: Springer International Publishing.
- Richardson, J. T. E. (2006). Investigating the relationship between variations in students' perceptions of their academic environment and variations in study behaviour in distance education. *British Journal of Educational Psychology*, 76(4), 867-893. doi: 10.1348/000709905X69690
- Richardson, J. T. E. (2012). The attainment of White and ethnic minority students in distance education. *Assessment & Evaluation in Higher Education*, 37(4), 393-408. doi: 10.1080/02602938.2010.534767
- Rienties, B., Borooa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action Evaluation Framework: a review of evidence-based learning analytics interventions at Open University UK. *Journal of Interactive Media in Education*, 1(2), 1-12. doi: 10.5334/jime.394
- Rienties, B., Cross, S., & Zdrahal, Z. (2016). Implementing a Learning Analytics Intervention and Evaluation Framework: what works? In B. K. Daniel (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (pp. 147-166). Heidelberg: Springer.
- Rienties, B., Herodotou, C., Olney, T., & Schencks, M. (Submitted). Is there a need for professional development training when using learning analytics?: a Technology Acceptance teacher perspective.
- Rienties, B., Johan, N., & Jindal-Snape, D. (2015). Bridge building potential in cross-cultural learning: a mixed method study. *Asia Pacific Education Review*, 16, 37-48. doi: 10.1007/s12564-014-9352-7
- Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: a cross-institutional comparison across 151 modules. *Computers in Human Behavior*, 60, 333-341. doi: 10.1016/j.chb.2016.02.074
- Sharples, M., Adams, A., Ferguson, R., Gaved, M., McAndrew, P., Rienties, B., . . . Whitelock, D. (2014). *Innovating Pedagogy 2014*. Milton Keynes: Open University.
- Sharples, M., de Roock, R., Ferguson, R., Gaved, M., Herodotou, C., Koh, E., . . . Wong, L. H. (2016). *Innovating Pedagogy 2016: Open University Innovation Report 5*. Milton Keynes: The Open University.

- Simpson, O. (2004). The impact on retention of interventions to support distance learning students. *Open Learning: The Journal of Open, Distance and e-Learning*, 19(1), 79-95. doi: 10.1080/0268051042000177863
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior*, 47, 157-167. doi: 10.1016/j.chb.2014.05.038
- Tingle, R., & Cross, S. (2010). *How do 'WP' students differ from others in their engagement with e-learning activities*. Paper presented at the OU Widening Participation Conference, Milton Keynes, UK. https://latestendeavour.files.wordpress.com/2010/06/wp_students_engagment_with_elearning_poster_tinglecross2010.pdf
- Toetenel, L., & Rienties, B. (2016). Learning Design – creative design to visualise learning activities. *Open Learning*, 31(3), 233-244. doi: 10.1080/02680513.2016.1213626
- Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C.-C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education*, 106, A1-A9. doi: 10.1016/j.compedu.2016.12.002
- Universities UK. (2013). The funding environment for universities: an assessment. London: Universities UK.
- Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). *Dataset-driven research for improving recommender systems for learning*. Paper presented at the Proceedings of the 1st International Conference on Learning Analytics and Knowledge, NY.
- Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). *Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment*. Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge, Indianapolis.
- Yin, R. K. (2009). *Case study research: Design and methods* (5 ed.). Thousand Oaks: Sage.

Figure 1 Mature stage of Data Wrangling (Clow, 2014)

Figure 2: Context of Data Wrangling

Figure 3: New structure of Data Wrangling