A framework for strategic planning of data analytics in the educational sector

A thesis submitted to Middlesex University in partial fulfilment of the requirements for the degree of Master of Science by Research (MRes)

Ariadni Tsiakara

M00762069

School of Science and Technology
Middlesex University London

January 2021

Abstract

The field of big data and data analysis is not a new one. Big data systems have been investigated with respect to the volume of the data and how it is stored, the data velocity and how it is subject to change, variety of data to be analysed and data veracity referring to integrity and quality. Higher Education Institutions (HEIs) have a significant range of data sources across their operations and increasingly invest in collecting, analysing and reporting on their data in order to improve their efficiency. Data analytics and Business Intelligence (BI) are two terms that are increasingly popular over the past few years in the relevant literature with emphasis on their impact in the education sector. There is a significant volume of literature discussing the benefits of data analytics in higher education and even more papers discussing specific case studies of institutions resorting on BI by deploying various data analytics practices.

Nevertheless, there is a lack of an integrated framework that supports HEIs in using learning analytics both at strategic and operational level. This research study was driven by the need to offer a point of reference for universities wishing to make good use of the plethora of data they can access. Increasingly institutions need to become 'smart universities' by supporting their decisions with findings from the analysis of their operations. The Business Intelligence strategies of many universities seems to focus mostly on identifying how to collect data but fail to address the most important issue that is how to analyse the data, what to do with the findings and how to create the means for a scalable use of learning analytics at institutional level.

The scope of this research is to investigate the different factors that affect the successful deployment of data analytics in educational contexts focusing both on strategic and operational aspects of academia. The research study attempts to identify those elements necessary for introducing data analytics practices across an institution. The main contribution of the research is a framework that models the data collection, analysis and visualisation in higher education. The specific contribution to the field comes in the form of generic guidelines for strategic planning of HEI data analytics projects, combined with specific guidelines for staff involved in the deployment of data analytics to support certain institutional operations.

The research is based on a mixed method approach that combines grounded theory in the form of extensive literature review, state-of-the-art investigation and case study analysis, as well as a combination of qualitative and quantitative data collection.

The study commences with an extensive literature review that identifies the key factors affecting the use of learning analytics. Then the research collected more information from an analysis of a wide range of case studies showing how learning analytics are used across HEIs. The primary data collection concluded with a series of focus groups and interviews assessing the role of learning analytics in universities. Next, the research focused on a synthesis of guidelines for using learning analytics both at strategic and operational levels, leading to the production of generic and specific guidelines intended for different university stakeholders. The proposed framework was revised twice to create an integrated point of reference for HEIs that offers support across institutions in scalable and applicable way that can accommodate the varying needs met at different HEIs. The proposed framework was

evaluated by the same participants in the earlier focus groups and interviews, providing a qualitative approach in evaluating the contributions made during this research study.

The research resulted in the creation of an integrated framework that offers HEIs a reference for setting up a learning analytics strategy, adapting institutional policies and revising operations across faculties and departments. The proposed C.A.V. framework consists of three phases including Collect, Analysis and Visualisation. The framework determines the key features of data sources and resulting dashboards but also a list of functions for the data collection, analysis and visualisation stages.

At strategic level, the C.A.V. framework enables institutions to assess their learning analytics maturity, determine the learning analytics stages that they are involved in, identify the different learning analytics themes and use a checklist as a reference point for their learning analytics deployment.

Finally, the framework ensures that institutional operations can become more effective by determining how learning analytics provide added value across different operations, while assessing the impact of learning analytics on stakeholders. The framework also supports the adoption of learning analytics processes, the planning of dashboard contents and identifying factors affecting the implementation of learning analytics.

Acknowledgments

Throughout the writing of this dissertation I have received a great deal of support and assistance.

I would first like to thank my supervisors, Professor Xiaohong Gao and Professor Franco Raimondi, whose guidance when needed was valuable to choose the right direction and successfully complete my dissertation.

In addition, I would like to thank my husband for his wise counsel and sympathetic ear. He was always there for me. Finally, I could not have completed this dissertation without his support, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

This introductory chapter provides the necessary foundations for the research conducted. The chapter begins with the contextualisation of the research and its significance for the educational sector. The chapter proceeds with a detailed explanation of the research scope leading to the main research question. Next, the chapter focuses on the steps followed during the research and clarification of the key activities undertaken. Then a discussion of the key considerations and issues addressed during this research is presented. Finally, the chapter ends with an outline of the main outputs and deliverables.

1.1. Contextualising the research

In the modern information society, data of high volume and tremendous variety are produced at any given time, something that is facilitated by the most recent technological advances. Higher Education Institutes (HEI) are drowning in data derived from many different sources such as ‘student registration information’, ‘performance data’, ‘website usage’, ‘emails’ and ‘social networks’ among others (Universities UK, 2016) (Job, 2018) (Sclater eta l, 2016) (Albalowi and Alhamed, 2017) (Carmichael, 2019) (Ong, 2015). When it comes to reaping the benefits of big data, the main challenge HEIs must face is to identify ways to analyse them and subsequently use them either for understanding domain problems (e.g., causes for dips in recruitment, reasons for inconsistent student progress) or supporting decision making (e.g., investing in alternative marketing approaches, targeting new markets).

Increasingly the term ‘smart university’ is used to describe HEIs demonstrating efficiency in managing their data in ways that help them to improve their operations (i.e., student support), enhance their provision (i.e., learning experience) and enable strategic planning through well-informed decisions (i.e., new or revised courses) (IBM, 2016) (Eassom, 2014). Smart universities are HEIs that are capable to adopt a combination of both predictive and prescriptive analytic tools; therefore, being able to predict what will happen based on real facts and to inspect what will happen which is the best way to respond and react on that (Albalowi and Alhamed, 2017).

For smart HEIs, the production of data from different sources is perceived as an opportunity rather than a burden. Hence, they resort in standard practices that employ data analytics for deducting meaning from operational data (e.g., why certain courses receive better evaluation than others, how the evaluation of a specific course changes over the years). For smart HEIs, this is possible through the application of tools and proven practices for business intelligence, predictive analytics, financial performance, strategy management, and other analytic applications to improve their current and future performance (Shacklock, 2016) (Krawitz et al, 2018) (Brooks and Thayer, 2016) (Universities UK, 2018).

Typically, a smart HEI would apply data analytics technology to improve their evaluation and reporting capabilities across a multitude of institution-wide areas including (i) finance, (ii) human resources, (iii) property management, and (iv) student records (Wong, 2017) (Mazriani, 2018) (Kellen et al, 2013) (Reichley et al, 2018) (IBM, 2019). These are key areas, providing an institution with the necessary information for their operations, and subsequently to make well-informed decisions and justify their actions. This research study aims at providing a framework that can be used by HEIs to plan their data analytics strategy with emphasis on identifying the information required for conducting data analysis for a range of operations in the educational sector, determining the data analysis process that must be performed and providing guidelines for the visualisations that must be produced to enable HEI decision-making

1.2. Research scope and question

This research study provides the means for HEIs to put together a strategic plan for data analytics. The proposed framework attempts to fill in a gap in the education sector in the following ways:

- Identify the sources of information, as well as the type of information required for data analytics in the educational sector.
- Determine the process that must be followed for conducting a thorough data analysis on educational data for a range of HEI operations.
- Produce guidelines for the creation of visualisation of data analysis for certain educational operations.

These three elements of the proposed framework correspond to the three main objectives of this research study. By combining these objectives, the following research question is formed:

“**What are the key elements of a framework supporting strategic planning of data analytics in the educational sector**”.

This research study attempts to answer the above research question, and therefore provide the means for HEIs to identify how they can fully embark onto institution-wide data analytics initiatives that support institutions’ both strategic plans and operations.

1.3. Overview of research activity

Due to the nature of this work, the research study brings together a number of activities that classify the work as a multi-methodological study. More specifically, the study is initially based on grounded theory, a literature review is conducted to ensure that similar works are reviewed and reflected upon. Subsequently, a series of interviews and focus groups are conducted to determine various data analytics aspects from an HEI point of view. The choice of the most appropriate data collection technique is affected from the availability of the necessary experts. For the purposes of this study, subject experts include (i) Middlesex University Tableau experts sharing their experiences in producing dashboards, (ii) Middlesex University academic and professional staff who can share their views on the value of data analytics for their respective roles, and (iii) experts from other institutions to allow a benchmarking exercise that covers more than one institution. The study started in January 2020 and is significantly affected by the COVID-19 pandemic with respect to the ability to conduct face-to-face data collection sessions. The study plan shifted towards virtual meetings to ensure access to expertise and knowledge needed for requirements elicitation.

The research process involves the following activities:

(1) Conducting a literature review on the use of data analytics in HEIs.

(2) Interviewing key staff for identifying data analytics requirements per role.
(3) Establishing data analytics with the use of focus groups.
(4) Creating a draft of the proposed framework.
(5) Producing the necessary guidelines.
(6) Evaluating the guidelines with a selection of staff.
(7) Revising the proposed framework.

1.4. Research considerations

This work is highly beneficiary for HEIs, as it provides a framework that can be used both as a self-evaluation tool, as well as a reference for good practice in producing educational data analytics. The following concerns are associated with the successful completion of the research study:

- Ability to access institutional information on data analytics strategy, as this is likely to include sensitive data that may not be widely available.
- Participation of a critical mass across academic and professional staff in order to accurately map data analytics requirements.
- Modelling a significantly wide range of operations and creating the corresponding dashboards to enable a thorough evaluation of the proposed framework.

The research study involves a number of key roles in data collection. These include senior academics from different institutions, professional and academic staff from Middlesex University, as well as data analytics experts from a strategic partner of the Computer Science Department and member of its industrial board (Info-Lab's Data School).

1.5. Research deliverables

This research has a very practical focus in the sense that it can help institutions with the strategic planning of their data analytics processes. The research combines aspects of strategic management in the role of data analytics in HEIs, as well as operational issues associated with the deployment of data collection, analysis and visualisation practices across different areas in modern institutions.

The study contributes to the field in a number of ways via the following deliverables:

- A literature review on the role of data analytics in higher education.
- A State-of-the-Art review of data analytics practices and techniques in HEIs.
- A comparative analysis of data analytics applications and tools in the field of education.
- Reflection summaries from data collection involving experts participating in focus groups and interviews.
- A framework for setting strategic plans for the deployment and exploitation in data analytics in HEIs.
- A set of guidelines on how to use data analytics in HEIs including step-by-step processes to be followed by management, administrative and academic staff.

The work carried out during this research study is discussed in eight chapters, which are illustrated below. The first chapter describes the research context, scope, as well as the various activities carried out and the deliverables associated with the stages of the research process. The second chapter provides a literature review that consists of four sections, feeding in the first version of the C.A.V. framework, which is the main contribution of this research. The third chapter includes an investigation of almost thirty case studies of Learning Analytics (LA) used in Higher Education Institutions (HEIs) and a critique of the most common software applications used for learning analytics. The findings from the case study analysis also plays an important role in shaping the first version of the framework.

The fourth chapter explains the research stance and describes the research process. The next two chapters discuss the contribution of the research study including the collection of primary data (chapter 5) and provision of institutional and user guidelines for learning analytics. The sixth chapter describes the evaluation process followed, leading to the final version of the C.A.V. framework before the conclusions that are presented in the eight chapter.

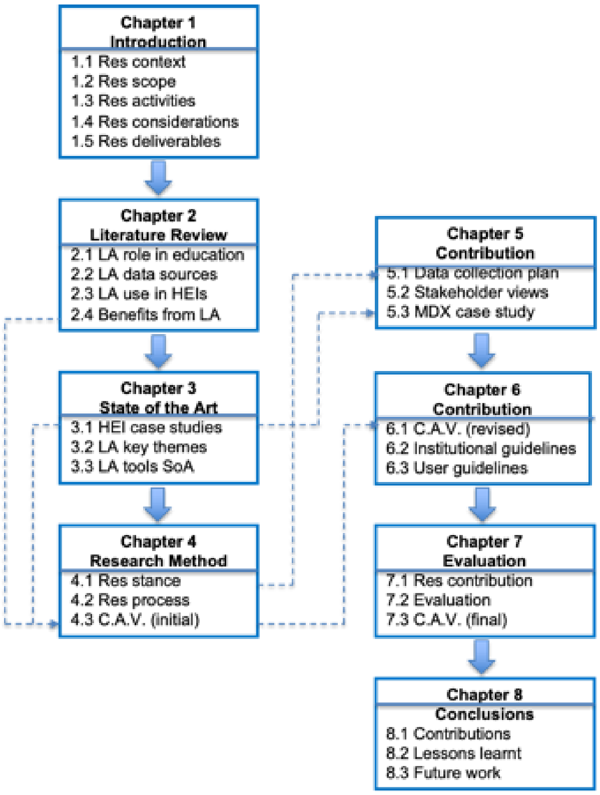

Figure 1-1: Structure of thesis chapters

The following chapter provides the foundation for this research study, as it includes a literature review, a state-of-the-art report in data analytics for the education sector and a review of institutional case studies.

1.6. Summary

This first chapter provided an overview of the thesis and the structure of its chapters. The next chapter focuses on the literature review conducted as part of this research study.

Chapter 2 – Literature review

The second chapter of the thesis discusses the different literature review areas, including the role of learning analytics in education, the range of learning analytic data sources, as well as how learning analytics are used in higher education institutions and any associated benefits.

2.1. The role of data analytics in higher education

Universities UK (UUK), is the representative organisation for the UK's universities, with a mission to "promote a successful and diverse higher education sector." According to UUK's publication on 'Analytics in Higher Education' the vast amounts of data collected by universities are underutilised, making it necessary to reap the benefits of learning analytics "effective implementation of appropriate technology and techniques" (Universities UK, 2016). UUK suggests that universities should focus on personalising the learning experience and target resources where "they can be most effectively employed". This requires a clear plan on how to make better use of educational data, which is a key objective of this research study. The UUK case for better use of analytics in higher education (Universities UK, 2016) advocates the need for institutional transformations, where change must be endogenous, driven and owned by institutions themselves and, ultimately, by the individual owners and end-users of these powerful sets of data". This research study contributes the means for such change, as it offers a strategic planning framework for making better use of data analytics in the educational sector.

Job (2018) suggests and efficient way of applying big data analytics in higher education (Job, 2018). The author suggests a framework that involves extensive data acquisition of structure, semi-structure and unstructured data, leading through predictive analysis to a number of useful outputs such as quality assurance, performance evaluation and improvement mechanisms, progression and regression rates, teaching evaluation and learning resources evaluation (Job, 2018). This framework is very useful in summarising the role of data analytical tools in predictive analysis, as well as identifying the association between the source of big data in education (e.g., student registration data, assessment performance and student admission results) and outputs of data analytics. This study follows a similar approach in the sense that it identifies the necessary sources for HEI data, as well as the different outputs that can be generated following effective analysis of such data sets. The study extends the suggested framework significantly as it focuses on providing specific guidance and detailed description of the instruments needed to produce the data analysis outputs.

In 2016, Sclater et al (2016) produced a very useful report on UK and international practice for 'Learning Analytics in Higher Education'. In their report, the authors conclude that "learning analytics has the potential to transform the way we measure impact and outcomes in learning environments – enabling providers to develop new ways of achieving excellence in teaching and learning, and providing students with new information to make the best choices about their education" (Sclater et al, 2016). Sclater et al (2016) produced their report for JISC, explaining how learning analytics works, covering a range of HEI case studies across the globe. This research study bases its contributions on the JISC's learning analytics architecture as illustrated below.

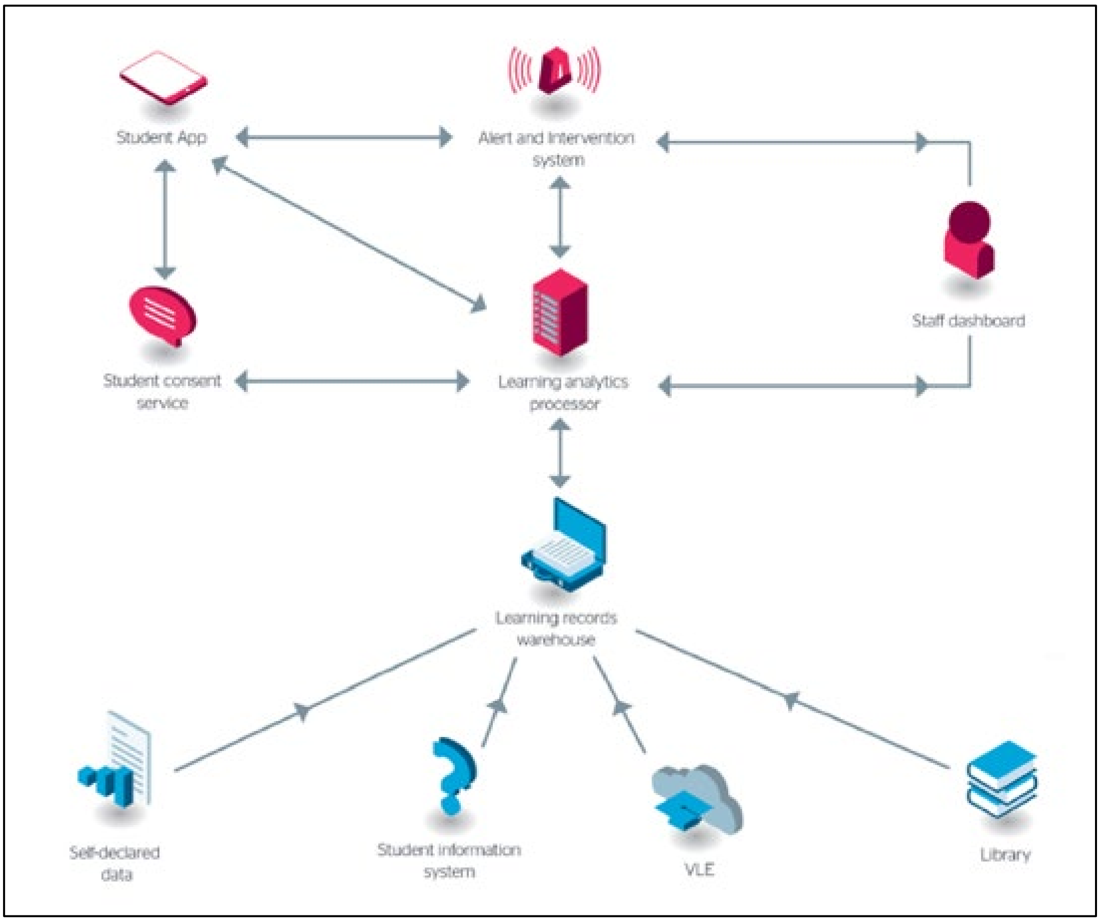

Figure 2-1: JISC's Learning Analytics Architecture (Sclater et al, 2016)

The JISC learning analytics architecture consists of a learning records warehouse (where data from libraries, student systems and virtual learning environments are gathered) and a learning analytics processor that produces staff dashboards, while ensuring student consent, alert and intervention systems and student applications operate in alignment. This study assumes that the basics of this architecture are in place in the majority of the HEIs investigated in the literature and extends the work by providing specific guidance on the design and population of staff dashboards.

2.2. Sources of learning analytics data

A significant volume of literature discusses how different sources of learning data are used for determining the data types and variables used for analytics model in HEIs. Alblawi and Alhamed (2017) propose nine data points including the following (Alblawi and Alhamed, 2017):

- Student personal information
- Student performance statistics
- Student engagement metrics
- Student online learning engagement
- Past student achievement
- Student social network activity
- Student extracurricular activity

- Student health background
- Student financial background

This study appreciates how such lists of student data sources can vary significantly depending on the contributors'' objectives, research objectives and methodological stance. There is also the issue of ethics and conducting analysis on personal information such as financial and health records. This research study attempts to provide guidance on how to assess whether different data sources should be taken under consideration in the production of dashboards in HEIs.

Christine Carmichael (2019) discusses a new approach in dealing with the generation of data analysis reports in the educational sector, where individuals across an institution are empowered to generate the reports, they need with the use of sophisticated data visualisation tools such as the Tableau dashboard generating platform.

There are clear benefits with such practice (assuming that all stakeholders are aware of ethical, security and other concerns) including the following (Carmichael, 2019):

- Enabling Self-Reliance
- Speed at Every Phase
- Flexible & Secure Configurations
- Visual Understanding

This research study aims at supporting these benefits by providing specific guidelines on how to generate learning analytics dashboards.

The study is based on assessing several case studies and how they use dashboards of different design to meet their institutional needs for learning analytics. JISC provides a very useful instrument in assessing how institutions utilise their data. The InfoNet BI Maturity conceptual framework provides a "good way of describing an area of project, so that practitioners can communicate with one another, describe progress in project, and identify shared goals and problems". The framework consists of six stages, and institutions with the highest level of BI maturity demonstrate that "systems are used for evidence-based decision-making and for predictions, models and assessment of future options" (Ong, 2015). The scope of this research study is to help institutions to assess to what extent they have in place appropriate plans for the deployment of data analytics. Ong (2015) discusses in such a case study, the use of student engagement for producing useful dashboards. The author discusses the default engagement measurement index consisting "of the average scoring of student's engagement on predetermined events based on the impact and decay factors", where the impact factor "refers to the level of importance of each engagement event" and the decay factor "refers to the value that does not affect scoring after a certain period of time". This research study involves experts in the field to provide an accurate reflection on the most appropriate indicators that can be used for assessing an institution's performance.

2.3. Using data analytics in Higher Education Institutions

Another report following a "ten-month inquiry co-chaired by Lord Norton and Sarah Porter" discusses the potential of data analytics in Higher Education. The Higher Education

Commission report identifies four key motivations for learning analytics as follows (Shacklock, 2016):

- Increasing retention
- Providing better feedback to students
- Capturing attendance data
- Enhancing teaching and learning

As a result, from the report the Commission "recommends that institutions should be encouraged to use the information from learning analytics systems to identify and foster excellent teaching within their institutions" (Shacklock, 2016). More specifically a list of twelve recommendations provides guidance on how HEIs should deal with data analytics. This research study is ideally aligned with most of these recommendations and in particular the first three as shown in the table below.

Commission's recommendations	Current study contributions
1. HESA, JISC and Universities UK should work together to develop a sector-wide strategy for excellent and innovative data management. This strategy will support and enable sharing and collaboration between institutions.	• Contributes in strategic planning for data analytics. • Suggests ways to identify, collect, analyse, visualise and report appropriate data sets.
2. HESA should take responsibility for rationalising the data collection process across the sector, working in partnership with others.	• Provides guidelines on data collection and analysis.
3. All HEIs should consider introducing an appropriate learning analytics system to improve student support / performance at their institution. Any such decision should be fully informed by an analysis of the benefits, limitations and risks attached.	• Maps strategic plans to operational objectives through specific learning analytic activities. • Determines appropriate actions needed to produce the most suitable data analytics for each aspect of educational activity.

Table 2-1: Alignment of the research study to the Higher Education Commission report on Data Analytics

Krawitz et al (2018) in their McKinsey report on the transformation of HEIs with the use of advanced analytics recognise that it is a difficult task to achieve. They identify a number of key obstacles that form the main challenge for Universities that seek transformation through via data analytics. First, they argue that quite often the key driver for such change is not an objective for innovative management for educational data but the need for compliance to the sector standards. Secondly, data analytics ownership is quite often positioned in silos making it difficult to achieve institution-wide transformation. This isolation effect is also witnessed in the third obstacle that is lack of data sharing practices across different units of a University. Finally, lack of talent may lead to failure of creating a data analytics strategy that is aligned to initiatives that lead to true institutional transformation. The deliverables of this study seem to be closely related to the five action steps proposed by the report as follows (Krawitz et al 2018):

- Articulate an analytics mandate that goes beyond compliance (this study enables strategic plans to be formed with data analytics being a key driver for institutional transformation).
- Establish a central analytics team with direct reporting lines to executive leaders (this study provides guidelines on how to align data analytics tasks to different HEI operations).
- Win analytics buy-in from the front line and create a culture of data-driven decision making (this study provides guidance on formulating a data analytics strategy leading to key benefits for a HEI).
- Strengthen in-house analytical capabilities (this study offers a detailed guide on a range of data analytics practices).
- Deploy a test-and-learn approach (this study provides guidance on how to conduct data analytics for specific operations that can be further extended to cover other aspects across a department or faculty).

Brooks and Thayer (2016) discuss how despite the fact that educational analytics are these days a priority for most institutions, the extent to which analytics is used varies significantly across different departments and units. Typical uses of analytics include planning, admissions, enrolment and progression. This study attempts to identify the different type of support required for deploying and effectively using educational analytics across a range of activities and associated organisation units. A wide variety of examples on how data analytics can effectively visualise patterns and trends in higher education are represented in the Universities UK report (Universities UK, 2018).

2.4. Institutional benefits from the use of learning analytics

Based on a thorough investigation of several higher education learning analytics case studies Wong identifies a number of common benefits for HEIs as follows (Wong, 2017): (i) improving student retention, (ii) increasing cost-effectiveness, (iii) understanding student learning behaviours, (iv) providing personalised assistance to students, and (v) offering timely feedback and interventions. Similarly, IBM Campus Technology report, identifies the following typical uses of learning analytics (IBM, 2016):

- Targeting student scholarships.
- Improving admissions Return on investment.
- Identifying students at risk.
- Tracking attendance.
- Evaluating curriculum.
- Identifying investors.
- Making operational savings.

IBM's Campus Technology is a monthly publication "focusing exclusively on the use of technology across all areas of higher education" (IBM, 2016). Eassom (2019) also discusses the IBM's report and focuses on the fact that "Chancellors, can maximise facility use and enhance decision making around current and future usage, scheduling, leasing and capital projects". More specifically the author explains that "by combining class-leading analytics and optimisation tools with state-of-the-art facilities management software, IBM has addressed that requirement with innovative smarter campus solutions that can help visualise shifts and changes in patterns of occupancy and demand and drive effective space utilisation into all areas of facilities and asset management".

According to Mazriani analytics can increase student success in several ways (Mazriani, 2018). Operations supported by analytics in academic institutions include aligning academic programmes to student needs, optimising the delivery of curriculum, performing enrolment forecasts, budgeting and resource analysis, as well as having a clear and accurate view on student performance.

Kellen et al also provide some suggestions on how to apply big data in higher education (Kellen et al, 2013). They offer a similar list of operations that are affected by data analytics, as follows: (i) giving students real-time feedback on their involvement with the university, (ii) providing richer student retention and graduation analysis, (iii) integrating data from public social media and analyse for student involvement and success, (iv) helping students with learning, (v) detecting when students are failing to make sufficient progress, (vi) giving students intelligent course recommendations, (vii) improving classroom scheduling and building utilisation, (viii) improving tuition revenue forecasting and the pinpointing of financial aid.

This research study is using the literature review to establish its grounded theory, as it investigates the role of data analytics to support the range of operations identified in the different papers. The study identifies a list of operations that can be supported by data analytics, and prioritises how common these operations are between the universities that deploy learning analytics. The research study is also concerned with the feasibility to support certain operations at institutional level, as well as the effectiveness, and impact of data analytics for each of the identified operations.

An interesting view is expressed by Reichley et al (2018), where institutions are evaluated with respect to their maturity in Business Intelligence and the way they use data analytics. This study attempts to support institutions in performing self-assessment on how ready and perhaps mature they are to deploy analytics at strategic and operational levels.

From the literature review a total number of 41 elements are identified as possible areas where data analytics can be used effectively in a Higher Education institution. These elements are currently reviewed in order to produce a final ranking according to:

- How common they are
- Feasibility of using these elements to transform institutional strategy
- Effective use of analytics for institutional operations

These elements provide the foundations for the guidelines constructed in the later chapters of this thesis. The following figure shows how these elements are present across 27 case studies that were investigated in the State of Art of learning analytics presented in the next chapter.

	Element	University of South Florida	Wollongong University	Rio Salado Community College	University of Edinburgh	Purdue University	University of Maryland	New York Institute of Technology,	California State University	Marist College	Edith Cowan University	University of New England	UK Open University	Nottingham Trent University	Open Universities Australia	University of Oklahoma	University of Central Florida	Hillsborough Community College	University of Alabama	Des Moines Area Community College	Carnegie Mellon University	MIT	University of San Diego	University of Otago	University of Bedfordshire	Bridgwater College	University of Derby	Lancaster University
1	Student Performance	X	X	X	X	X	X	X	X		X	X					X	X			X				X	X	X	
2	Outliers/Issues for Introversion		X																									
3	Student Potential		X									X									X							
4	Attrition		X			X														X								X
5	Instruction techniques		X																									
6	Assessment techniques		X																		X							
7	Curricula evaluation		X		X								X		X						X			X				
8	Engagement	X		X	X	X	X					X				X	X	X			X	X	X		X	X	X	X
9	Achievement	X		X	X	X						X			X		X	X			X						X	
10	Social Network				X		X					X			X						X							
11	Finances			X	X			X					X			X	X			X		X	X		X	X		
12	Timetables		X					X																				
13	Administrative Data			X			X	X	X		X	X			X	X		X		X		X	X	X	X	X	X	
14	Admissions Data				X			X				X	X									X	X	X				
15	Research Data				X			X																X				
16	Planned Work							X																				
17	Staff Data							X																X				
18	Course Data				X			X													X							
19	Environmental Data		X					X													X			X				
20	Alumni Data					X		X				X		X														
21	Historical Data					X		X				X				X		X								X		
22	Facilities Data							X													X							
23	Scholarships			X									X															
24	At-Risk Students	X		X		X							X												X			
25	Attendance				X	X	X		X			X	X				X	X			X				X		X	X
26	Identify Investors					X							X															
27	Recruitment													X		X												
28	Student needs				X									X			X									X		
29	Marketing													X														
30	Retention					X					X			X														
31	Students performance feedback		X									X					X											
32	retention / graduation		X	X		X			X		X	X					X		X						X		X	
33	Involvement		X																									
34	Learning Support	X	X		X	X			X		X						X				X					X		
35	Progression		X						X		X						X		X									
36	Course recommendations		X		X																							
37	Tuition revenue		X																									
38	Faculty Performance	X								X												X	X	X				
39	Degree Completion	X		X						X	X	X													X			
40	Student Management									X					X			X			X				X		X	
41	Learning Analytics	X			X				X	X		X					X				X	X	X				X	X
42	Business Analytics				X					X					X							X	X	X	X			

Table 2-2: Data Analytics Elements for HEIs

2.5. Summary

In this second chapter, the literature review conducted was presented in detail. The next chapter provides a detailed state of the art in the field with emphasis on a series of case studies showcasing the use of learning analytics in higher education institutions.

Chapter 3 – State of the Art report

The third chapter of the thesis discusses how learning analytics are used in Higher Education Institutions and identifies the key themes of learning analytics in universities. The chapter also provides a state-of-the-art report on tools used for learning analytics.

3.1. University Case Studies on the use of data analytics

A total of twenty-seven (27) case studies are discussed in the relevant literature with emphasis on their data analytics practices. In this section a selection of case studies is discussed and a comparative analysis aims at (i) identifying the objectives of data analytics practices, (ii) describing the data collection techniques used and (iii) explaining the data analysis results including any dashboard visualisations and reporting techniques.

The following table provides a summary of the case studies and the key objectives of each learning analytic initiative that was reviewed.

#	University	Objectives
1	University of South Florida https://www.usf.edu/	• Improving student graduation dates • Improving retention rates
2	University of Wollongong https://www.uow.edu.au/	• Analysing online student discussion forums • Assessing student engagement • Identifying students who are isolated from the main discussion
3	Rio Salado Community College https://www.riosalado.edu/	• Identifying students at risk • Avoiding students failing
4	University of Edinburgh https://www.ed.ac.uk/	• Improving course design • Improving attainment • Improving the student experience
5	Purdue University https://www.purdue.edu/	• Enhancing student success at a course level • Increasing overall retention and graduation rates • Identifying potential problems as early as the second week in the semester
6	University of Maryland https://www.umd.edu/	• Addressing low-level graduate support and high drop-out rates • Using predictions to support students • Identifying effective teaching practices with a view to enhancing future provision
7	New York Institute of Technology https://www.nyit.edu/	• Making early interventions with at-risk students
8	California State University https://www.calstate.edu/	• Assessing student success based on demographic data and VLE use
9	Marist College https://www.marist.edu/	• Introducing predictive models to help at-risk students
10	Edith Cowan University https://www.ecu.edu.au/	• Enhancing student retention, success and good ratings • Providing proactive support to retain students

11	University of New England https://www.une.edu.au/	• Identifying students who are struggling, so they can be offered timely support • Developing a dynamic, systematic and automated process that would capture the learning well-being status of all students
12	UK Open University https://www.open.ac.uk/	• Enhancing student success
13	Nottingham Trent University https://www.ntu.ac.uk/	• Enhancing retention • Improving attainment • Increasing a sense of belonging • Improving student support
14	Open Universities Australia https://www.open.edu.au/	• Personalising study experiences
15	University of Oklahoma https://www.ou.edu/	• Achieving successful recruitment of a talented student • Driving better business intelligence • Establishing digital citizenship for students, faculty, alumni and partners
16	University of Central Florida https://www.ucf.edu/	• Improving school progress using student success metrics
17	Hillsborough (FL) Community College https://www.hccfl.edu/	• Identifying previously enrolled students who need to complete only 25% or less of their graduation • Identifying students who may be closer to another degree than their declared program of study
18	University of Alabama https://www.ua.edu/	• Determining the correlation between a student asking for an official transcript and then leaving the university
19	Des Moines Area Community College https://www.dmacc.edu/	• Assessing attrition among new students
20	Carnegie Mellon University https://www.cmu.edu/	• Predicting students' future learning behaviour • Discovering or improving domain models • Studying the effects of different kinds of pedagogical support • Advancing scientific knowledge about learning and learners
21	Massachusetts Institute of Technology https://www.mit.edu/	• Modernising the systems that widely support MIT's administrative services and educational enterprise • Enabling the MIT community to serve themselves and do things for themselves by giving them better data/better access to data
22	University of San Diego https://www.sandiego.edu/	• Providing better customer experiences • Creating a digital revolution by making a number of new mobile apps available for their students
23	University of Otago https://www.otago.ac.nz/	• Assessing institutional performance and progress in order to predict future performance

24	University of Bedfordshire https://www.beds.ac.uk/	• Improving the student experience • Identifying at risk students
25	Bridgwater College https://www.bridgewater.edu/	• Ensuring that every learner has the best possible opportunity to be successful and to gain a qualification
26	University of Derby https://www.derby.ac.uk/	• Developing an excellent student experience by better understanding learners and their diverse needs
27	Lancaster University https://www.lancaster.ac.uk/	• Providing an interactive transcript that shows students their progress

Table 3-1: HEI Case Studies

The case studies were collected as a representative sample of international institutions covering the UK, as well as the US and Australia's higher education sector. In each case the analysis focused on:

- Identifying the institutional learning analytics objectives.
- Determining the data collection approaches adopted.
- Specifying the analysis results.

There is a wide range of data collection methods used in the above case studies. The plethora of data sources show the challenge ahead for institutions that wish to become 'smart' universities by utilising their data sources. During the analysis of the case studies the following data sources were encountered.

The University of South Florida collects its data from its Student Information System (SIS) and the Learning Management System (LMS), while the University of Wollongong collects data from its online student discussion forums from the Moodle platform. The data sources for the Rio Salado Community College include student grades, financial aid status, the frequency of interactions with online courses, Virtual Learning Environment (VLE) materials including discussion boards and student scholarships.

The University of Edinburgh uses a wide range of systems including (i) the Learning Analytics Report Card (LARC), (ii) a Virtual learning Environment (VLE), (iii) its Massive Open Online Courses (MOOC), (iv) Online Video Annotations for Learning (OVAL), (v) Multimodal Self-Regulated Learning (SRL) and (vi) Flipped Classrooms / Learning dashboard. Purdue University has developed a predictive algorithm based on student performance, effort, prior academic history and individual characteristics, attendance, VLE use, together with grade information held in the VLE gradebook. At the University of Maryland data are collected from VLE activity such as forum usage, to identify usage patterns, VLE log files in order to explore activity at a more fine-grained level and the relationship between final grades and specific types of VLE activity. New York Institute of Technology collects data from previous students, including key risk factors such as grades, admission application data, registration/placement test data, student surveys and financial data.

The California State University collects multiple demographic variables of its students such as students' current effort (VLE), individual characteristic variables, motivation and learning style. Similarly, Marist College collects demographic details such gender and age, and aptitude data such as high school scores, as well as VLE usage.

The data sources used at Edith Cowan University include grades, entrance scores, language skills, the university's Enterprise Information Management system, demographic information and student progress information. For the University of New England sources of data include a self-reported information about happiness from students via emoticons, extensive use of social media including Facebook, Twitter and Flickr, e-motion student input, class attendance, previous study history, prior test results, assignment submissions, VLE access patterns and previous scores.

At the UK Open University data are collected for students who withdraw from their programme, VLE use, student tutor notes and e-library data. At Nottingham Trent University, actual engagement data comes from four separate systems: (i) the VLE, (ii) the card access database, (iii) the assessment submission system and (iv) the library system. For Open Universities Australia data sources include a Customer Relationship Management (CRM) system, student profiles (e.g. location, socio-demographic factors, prior knowledge etc.), the VLE and the curriculum profile.

University of Oklahoma uses past data of students and data sources from faculties, departments, the registrar's office, financial aid, the university's MOOC and student engagement. The systems used for data collection at the University of Central Florida (UCF) include a Student Information System (SIS), a Learning Management System (LMS), the VLE and financial information. Similarly, Hillsborough (FL) Community College uses the university's Student Information System (SIS), the Learning Management System (LMS), the VLE and students' historical data.

University of Alabama uses student requests for official transcripts, while Des Moines Area Community College (DMACC) uses financial data such as students paying full tuition. For Carnegie Mellon University data come from the VLE, participation in discussion forums, practice tests scores and models used for classifying student activity from basic behaviour data. At MIT data come from academic departments, the registrar's office, financial aid, use of its MOOC and student engagement. Similar data sources are used at the University of San Diego, while the University of Otago collects (i) curriculum data, (ii) administrative data, (iii) department data, (iv) teaching and learning data, (v) student data and (vi) research data.

The University of Bedfordshire uses a range of sources including a student information system, VLE use, attendance records, swipe and proximity cards, the university Management Information System (MIS), as well as finance and HR systems. At the Bridgwater College data sources include data from previous years, self-assessment judgements, GCSE results, socio-economic group and working status, as well as a ProMonitor system and student logs.

The University of Derby uses its student information system, module data from the VLE (Blackboard), assessment data and submission information from Turnitin, Talis and other library systems, as well as attendance data together with swipe and proximity cards. Finally, Lancaster University uses its University Student Information (LUSI), attendance and Submission records, the VLE (Moodle), and a range of other library systems such as ALMA, Primo, EzProxy, Shibboleth and Aleph Archives.

These case studies were selected as they represent institutions with well-organised learning analytics policies and strategies. These case studies have structured learning analytics initiatives and clear procedures that exploit the use of dashboards. There is a wide selection of case studies including UK-based universities, as well as US universities and colleges. One of the key selection criteria was the ability to find sufficient details about learning analytics policies and their impact on operations. It was also important to include case studies that were discussed in peer-reviewed publications and supported by research studies. The next section focuses on identifying the key themes of learning analytics in Higher Education Institutions.

Full details are provided in Appendix A.

3.2. Key themes in the use of data analytics in academia

From the analysis of the case studies, a number of elements were identified as very important for the field of data analytics in higher education. These elements should be used at the core of learning analytics, driving the design of educational dashboards. The 42 elements identified in the previous chapter could be organised into three main themes.

The first theme, which appears to consist of the most popular elements can be described as the **decision-making theme**. These elements provide very useful insights on educational data that are helpful to institutions for a number of key decisions. The elements under this theme are as follows (list includes the number of appearances in the case studies and the reference number of the element based on the order it was found when conducting the review):

- (01) Student Performance – 16 appearances
- (08) Engagement – 16 appearances
- (13) Administrative Data – 15 appearances
- (25) Attendance – 12 appearances
- (09) Achievement – 10 appearances
- (11) Finances – 10 appearances
- (31) Retention / Graduation – 10 appearances
- (40) Learning Analytics – 10 appearances
- (33) Learning Support – 9 appearances

These elements are clearly associated with some key decisions required in HEIs, such as assessing student performance in order to plan interventions, or using student attendance to monitor engagement. Achievement, retention and learning support are areas that affect the way institutions make decisions about their delivery.

The second theme, includes elements with average number of appearances that can be described as belonging to a **statistics theme**. The elements belonging to the statistics theme seem to be used in some case studies, focusing on providing useful insights on certain operations and activities, such as admissions and identification of students at risk. These are likely to be used as a reference for key roles but not necessarily provide the back-end support for decision-making. The elements under the second theme are:

- (14) Admissions Data – 7 appearances

- (41) Business Analytics – 7 appearances
- (07) Curricula evaluation – 6 appearances
- (21) Historical Data – 6 appearances
- (38) Degree Completion – 6 appearances
- (39) Student Management – 6 appearances
- (10) Social Network – 5 appearances
- (24) At-Risk Students – 5 appearances
- (34) Progression – 5 appearances
- (37) Faculty Performance – 5 appearances
- (04) Attrition – 4 appearances
- (19) Environmental Data – 4 appearances
- (20) Alumni Data – 4 appearances
- (28) Student needs – 4 appearances
- (03) Student Potential – 3 appearances
- (15) Research Data – 3 appearances
- (18) Course Data – 3 appearances
- (30) Student performance feedback – 3 appearances

The third theme, consists of elements that are not that common focusing on specific operations, therefore it is described as an **operational theme**. These elements are not too common but can be really useful for specific dashboards if required by certain stakeholder groups.

The elements under the second theme are:

- (06) Assessment techniques – 2 appearances
- (12) Timetables – 2 appearances
- (17) Staff Data – 2 appearances
- (23) Scholarships – 2 appearances
- (26) Identify Investors – 2 appearances
- (27) Recruitment – 2 appearances
- (35) Course recommendations – 2 appearances
- (02) Outliers/Issues for Introversion – 1 appearance
- (05) Instruction techniques – 1 appearance
- (16) Planned Work – 1 appearance
- (22) Facilities Data – 1 appearance
- (29) Marketing – 1 appearance
- (32) Involvement – 1 appearance
- (36) Tuition revenue – 1 appearance

These 41 key elements are used in the C.A.V. framework in two ways, as they are included in the institutional guidelines for the design of learning analytics dashboards, as well as the list of functions that are part of the data collection, analysis and visualisation processes. The next section focuses on the different tools used for educational data analytics.

3.3. State of the Art in Educational Data Analytics

There are several platforms offering Business Intelligence (BI) solutions for modern organisations. These are not necessarily suitable for academic institutions but quite often organisational units may find a suitable selection of tools and functions in some of these software applications. This section provides a description of some of the most common BI solutions and an overview of the features provided by each vendor. The platform of choice for Middlesex University is Tableau, which is the main reason why the guidelines provided are customised for the Tableau software. In the following pages Tableau is compared against other alternatives. The scope of this research is not to provide platform-specific guidance on visualisation of data analytics in educational context but to offer generic guidance that can be adapted accordingly with respect to the BI solution that is deployed in a HEI.

There is no preferred tool for this research study, and the range provided is indicative of the options that can be used for learning analytics. The chosen system for Middlesex University is Tableau. This platform was used for analysis of institutional data and provides access to the university dashboards.

Application	Vendor	Description	URL
Alteryx	alteryx	Blends analytics from a range of sources to simplify workflows as well as provide a wealth of BI insights	https://www.alteryx.com
Birst	birst	A cloud-based service in which multiples instances of the BI software share a common data backend	https://www.birst.com/
Domo	DOMO	A cloud-based platform that offers business intelligence tools tailored to various industries (such as financial services, health care, manufacturing and education) and roles (including CEOs, sales, BI professionals and IT workers.	https://www.domo.com
Dundas BI	Dundas BI by Dundas Data Visualization	Mostly used for creating dashboards and scorecards, but can also do standard and ad-hoc reporting	https://www.dundas.com
Einstein Analytics	salesforce einstein analytics	An attempt to improve BI with AI	https://salesforce.com
Google Data Studio	Data Studio	Google Data Studio is a free dashboarding and data visualization tool that automatically integrates with most other Google applications, such as Google Analytics, Google Ads, and Google BigQuery. Thanks to its integration with other	https://datastudio.google.com

		Google services, Data Studio is great for those who need to analyse their Google data. For instance, marketers can build dashboards for their Google Ads and Analytics data to better understand customer conversion and retention.	
IBM Cognos	IBM COGNOS	IBM Cognos is a business intelligence platform that features built-in AI tools to reveal insights hidden in data and explain them in plain English. Cognos also has automated data preparation tools to automatically cleanse and aggregate data sources, which allows for quickly integrating and experimenting with data sources for analysis.	https://www.ibm.com/
KNIME	Open for Innovation KNIME	KNIME is a data analytics platform that supports data integration, processing, visualization, and reporting. It plugs in machine learning and data mining libraries with minimal or no programming requirements. KNIME is great for data scientists who need to integrate and process data for machine learning and other statistical models but don't necessarily have strong programming skills. The graphical interface allows for point-and-click analysis and modelling.	https://www.knime.com/
Microsoft Power BI	Microsoft	Microsoft Power BI is a business intelligence platform with support for dozens of data sources. It allows users to create and share reports, visualizations, and dashboards. Users can combine a group of dashboards and reports into a Power BI app for simple distribution. Power BI also allows users to build automated machine learning	https://powerbi.microsoft.com/en-gb/

		models and integrates with Azure Machine Learning.	
Microsoft Excel		Sisense is a data analytics platform aimed at helping both technical developers and business analysts process and visualize all of their business data. It boasts a large collection of drag-and-drop tools and provides interactive dashboards for collaboration.	https://www.microsoft.com/en-us/microsoft-365/excel
Oracle BI		Oracle Business Intelligence is a unique platform that enables customers to uncover new insights and make faster, more informed business decisions by offering agile visual analytics and self-service discovery together with best-in-class enterprise analytics.	https://www.oracle.com/business-analytics/business-intelligence/
Python		Python is an object-oriented scripting language which is easy to read, write and maintain. It was developed to supports both functional and structured programming methods. Python is easy to learn as it is very similar to JavaScript, Ruby, and PHP. Also, Python has very good machine learning libraries viz. Scikitlearn, Theano, Tensorflow and Keras. Another important feature of Python is that it can be assembled on any platform like SQL server, a MongoDB database or JSON. Python can also handle text data very well	https://www.python.org/
Qlik		Qlik provides a self-service data analytics and business intelligence platform that supports both cloud and on-premises deployment. The tool boasts strong support for data exploration and discovery by technical and nontechnical users alike. Qlik supports many types of charts	https://www.qlik.com/us/

		that users can customize with both embedded SQL and drag-and-drop modules.	
R		R is the leading analytics tool in the industry and widely used for statistics and data modelling. It can easily manipulate your data and present in different ways. It has exceeded SAS in many ways like capacity of data, performance and outcome. R compiles and runs on a wide variety of platforms viz -UNIX, Windows and MacOS.	https://www.r-project.org/
Rapid Miner		RapidMiner provides all the technology users need to integrate, clean, and transform data before they run predictive analytics and statistical models. Users can perform nearly all of this through a simple graphical interface. RapidMiner can also be extended using R and Python scripts, and numerous third-party plugins are available through the company's marketplace. However, the product is heavily optimized for its graphical interface so that analysts can prepare data and run models on their own.	https://rapidminer.com/
SAS BI		SAS Business Intelligence provides a suite of applications for self-service analytics. It has many built-in collaboration features, such as the ability to push reports to mobile applications. While SAS Business Intelligence is a comprehensive and flexible platform, it can be more expensive than some of its competitors. Larger enterprises may find it worth the price due to its versatility.	https://www.sas.com/
SAP BI		SAP BusinessObjects provides a suite of business intelligence applications for data	https://www.sap.com/

		discovery, analysis, and reporting. The tools are aimed at less technical business users, but they're also capable of performing complex analysis. BusinessObjects integrates with Microsoft Office products, allowing business analysts to quickly go back and forth between applications such as Excel and BusinessObjects reports. It also allows for self-service predictive analytics.	
Sisense		Sisense is a data analytics platform aimed at helping both technical developers and business analysts process and visualize all of their business data. It boasts a large collection of drag-and-drop tools and provides interactive dashboards for collaboration.	https://www.sisense.com/
Splunk		A "guided analytics platform" capable of providing enterprise-grade business intelligence and data analytics	https://www.splunk.com
SPSS		IBM SPSS is short for Statistical Package for the Social Sciences, and it's used by various kinds of researchers for complex statistical data analysis.	https://www.ibm.com/products/spss-statistics
Tableau		Tableau is a software that connects any data source be it corporate Data Warehouse, Microsoft Excel or web-based data, and creates data visualizations, maps, dashboards etc. with real-time updates presenting on web as well. It allows the access to download the file in different formats. Tableau's Big Data capabilities makes them important and one can analyse and visualize data better than any other data visualization software in the market	https://www.tableau.com/

Table 3-2: Business Intelligence / Data Analytics vendor solutions

3.4. Summary

This chapter provided an overview of a wide range of learning analytics case studies in Higher Education Institutions with emphasis on the different key themes, as well as a review of popular software applications used in learning analytics. The next chapter discusses the research method adopted in this research study.

Chapter 4 – Research Method

This chapter discusses the research method followed in this study. The chapter begins with reflections on the research stance of the researcher and the drivers for the range of techniques used for data collection, analysis and discussion. The chapter continues with a detailed description of the key research activities of the study leading to the creation of a solid foundation for the research contribution. This is in the form of the proposed framework for strategic planning of data analytics in educational contexts. The framework and its benefits for HEIs are discussed in the last section of the chapter, but are also revisited in later chapters following the evaluation phase of this research study.

4.1. Research Stance

The nature of this research study is such that requires a combination of research methods in order to answer the main research question. The study is first based on grounded theory, as it requires a literature review to understand the problem domain and the key factors affecting HEIs in the way they deploy and use data analytics. Furthermore, the study needs a solid foundation to make a number of hypotheses in the way data analytics are incorporated in institutional strategic planning. This leads to the use of case study analysis in order to acquire an in depth understanding of the specific aspects of the problem. Middlesex University was the chosen case study due to the fact that there is access to its use of data analytics at strategic and operational level.

Once the ground theory and case study analysis provide the necessary foundation for the research hypothesis (i.e., if HEIs align strategic plans and specific guidelines for deployment of data analytics they are likely to achieve increased effectiveness at operational level) further methods are needed to reach a concrete answer to the study's research question (i.e., what the key elements of a framework are, supporting strategic planning of data analytics in the educational sector). The study used three different methods as follows:

- A comparative analysis of University case studies to determine the current state of the art in the use of data analytics in higher education.
- A series of interviews with key stakeholders in the deployment of data analytics in HEIs to identify and understand factors affecting different operational roles.
- A number of focus groups aiming at clarifying how HEI strategic planning is affected by a range of data analytics factors.

The first method follows more the quantitative paradigm, as it identifies a number of factors affecting the use of data analytics aiming to determine the ones that are used more frequently. The second and the third method fall under the umbrella of qualitative analysis, aiming to assess the impact of the different factors at strategic and operational level in HEIs.

One of the main dilemmas at the beginning of this research study was to determine the best way for collecting primary data but also how to evaluate the research outputs. It appears that a significant proportion of the work carried out in the field is based on quantitative analysis of survey responses. Typically, academics create dashboards that are deployed in a series of pilots and then students are asked to provide their views on a number of questions. This has resulted in numerous papers presenting student views on the usefulness of the provided dashboards and the usability of their designs. This was never intended to be the scope of this study, as the contribution made is more towards establishing a strategic framework that also supported academic operations.

Therefore, it was important to establish a grounded theory in the form of the original framework using an extensive literature review. This included an initial literature review of approximately 30-40 papers to determine the key elements of the C.A.V. framework and the way they are interrelated in five main components (i.e., data sources, data collection, data analysis, data visualisation and dashboard design). Then a detailed analysis of 27 case studies provided sufficient evidence that the original framework incorporated all aspects that were covered in similar studies and action research (i.e., the implementation of dashboards based on research hypotheses and the evaluation of the learning analytics solution provided).

The next step was to conduct a number of interviews and a focus group in order to obtain the views of different stakeholder groups from three universities representing UK, Greece and Cyprus, as well as views from industry. A focus group was conducted to obtain the perspective of Middlesex University teaching staff. These views helped to shape up the C.A.V. framework and asses its applicability across different institutional layers (i.e., covering the needs of both teaching staff, administrators and senior managers), as well as its ability to support strategic and operational aspects of learning analytics.

In order to ensure that the C.A.V. framework covered all aspects of the learning analytics and educational data mining sectors, a second round of literature review was conducted. This was performed as a review of literature reviews (note: perhaps this can be described as a meta-literature review or systemic literature review). This review included around 60-70 papers covering studies presented at more than 500 papers over the past few years. This approach led to the final version of the C.A.V. framework including the detailed institutional and user guidelines, as well as a full list of elements necessary for the four dashboard types supported by the framework. Finally, the same stakeholders conducted for the original interviews were asked to provide their views on the C.A.V. framework components and its ability to support their institutional and individual learning analytics requirements. The evaluation of the C.A.V. framework was concluded with a final review of another part of the relevant literature focusing on evaluation approaches in learning analytics, based on around 20 papers.

Following such a multi-methodological approach was deemed necessary for the nature of the work covered in this research study. For example, Kokoc and Kara (2021) adopt a multiple-study investigation that contributes to the generalisability of their framework by highlighting "the importance of metacognitive and behavioural factors for the impact of learning analytics dashboards on learner performance". Similarly, Sindhiphak et al (2018) used the Delphi method "to query a panel of 19 experts in the fields of art education, educational technologists and artists to gather their input".

4.2. Research Process

The research process followed for this research study consists of three phases and overall, nine research activities. Some of these activities are conducted simultaneously, while others are organised in hierarchical order. The core research contribution of the research study is in the form of the Collect-Analyse-Visualise (C.A.V.) framework.

The nine research activities are identified as follows:

(1) Conducting a literature review on the use of data analytics in HEIs.
(2) Conducting a comparative analysis of HEI case studies in data analytics practices.
(3) Conducting a state-of-the-art study of the functionalities offered by data analytics platforms
(4) Introducing the C.A.V. framework (first draft).
(5) Interviewing key staff for identifying data analytics requirements per role and associated operations.
(6) Running focus group sessions to assess the impact of data analytics factors on HEI strategic planning.
(7) Revising the C.A.V. framework (second draft).
(8) Evaluating the framework with key stakeholders.
(9) Finalising the C.A.V. framework (final version).

An illustration of how each of the nine activities are associated and organised is shown below.

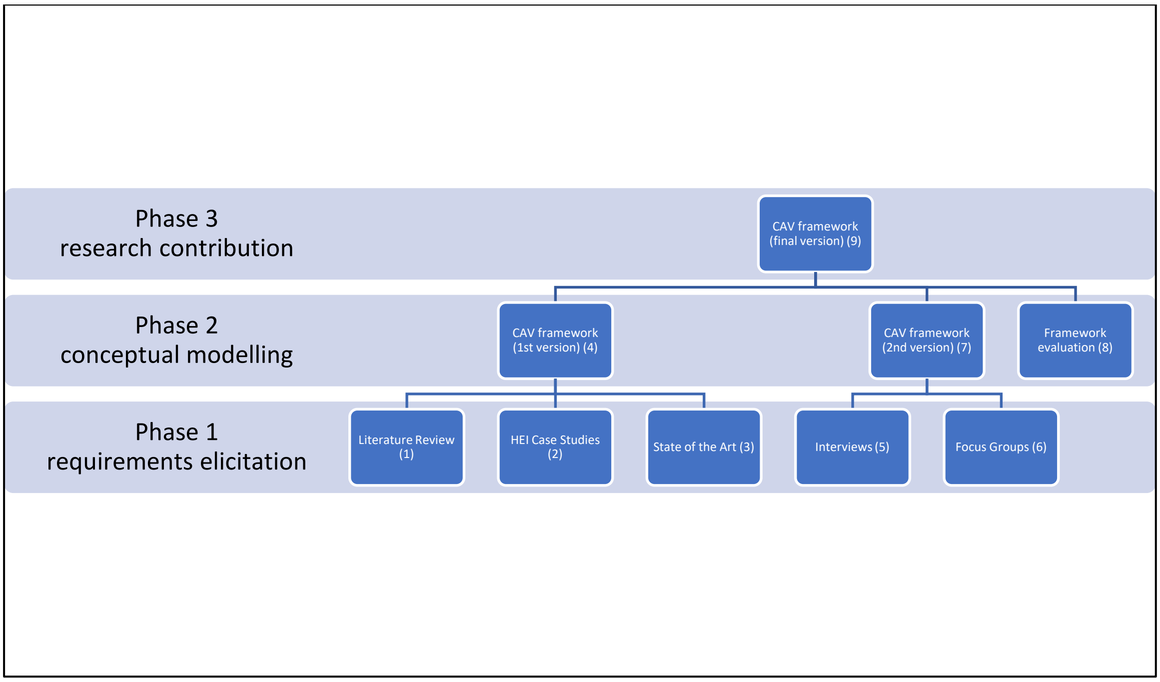

Figure 4-1: Research Process

The nine research activities are organised in three phases as follows (also shown in the figure illustrating the research process):

- Phase 1 – focusing on requirements elicitation and containing the five data collection activities (activities 1, 2, 3, 5 and 6).
- Phase 2 – focusing on conceptual modelling and containing the creation of the two draft versions of the C.A.V. framework and the evaluation of the framework (activities 4, 7 and 8).
- Phase 3 – focusing on the creation of the final version of the C.A.V. framework.

The main risk associated with this research process is that it requires access to focus groups and interviews with key stakeholders. This was eventually possible for this research study but ideally more individuals could be approached for establishing a bigger sample of responses. The number of individuals who contributed to the collection of the primary data and the framework evaluation is perceived to be adequate for the scope of the MRes.

Another weakness of the study relates to the fact that the proposed framework cannot be deployed at institutional level in order to assess its outcomes. This was addressed by evaluating the impact of such a framework with different stakeholders, who represent senior management, administration and teaching tasks.

4.3. Proposed Framework

This research study has a key contribution in the field of data analytics. More specifically it aims at introducing a framework for strategic planning of data analytics in the educational sector. The Collect-Analyse-Visualise (C.A.V.) framework provides an integrated solution for HEIs as it provides both (i) generic guidelines for strategic planning of data analytics in a HEI and (ii) specific guidelines on how to deploy data analytics for different HEI operations.

The framework consists of three parts as illustrated in the following figure. These are:

- Framework funnel (focusing on strategic planning)
- Framework process hierarchy (focusing on operational support)
- Framework circle matrix (focusing on strategic planning)

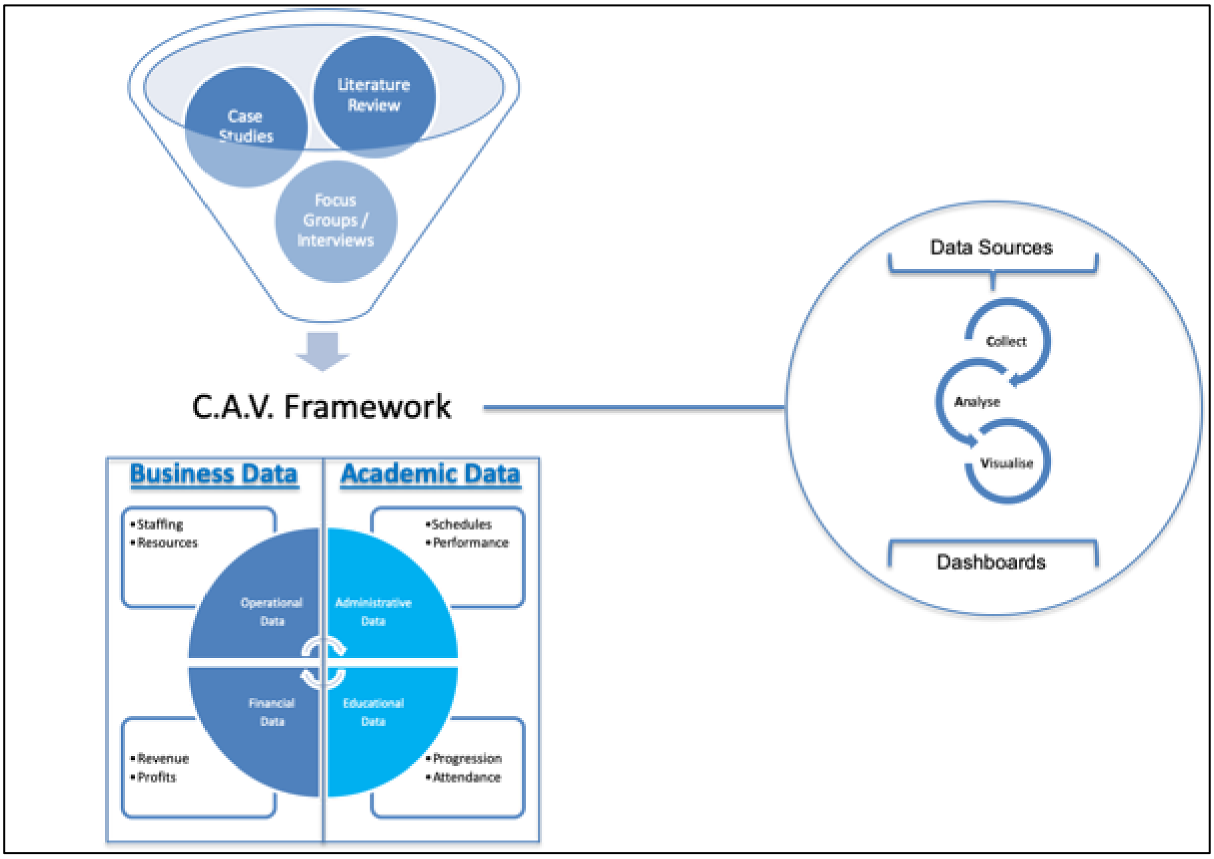

Figure 4-2: The C.A.V. Framework

The framework funnel determines how HEIs can identify those factors that are likely to affect their ability to conduct data analytics and deploy them at institutional level. The framework provides guidance on how to identify those elements of data analytics that will drive the data collection, analysis and visualisation.

The framework process hierarchy consists of five tasks and provides the guidelines for deploying data analytics at operational level. The first task is concerned with identifying the sources of HEI data to be used for the data analytics project. This first task triggers an iterative process consisting of (i) data collection, (ii) data analysis and (iii) data visualisation. Finally, the process concludes with the reporting of the data analytics process in the form of the generated dashboards.

The framework circle matrix returns the focus of the framework on strategic planning as it determines how HEIs can organise their dashboards based on the support and information they provide across the institution. More specifically the dashboard focus is classified as follows:

- Business data – including (i) operational data (such as staffing and resources) and (ii) financial data (such as revenue and profits).
- Academic data – including (i) administrative data (such as schedules and performance) and (ii) educational data (such as progression and attendance).

The C.A.V. framework was reviewed twice during the research study. The first review was conducted after the systematic literature review to ensure that its components cover all aspects of learning analytics that were identified in the literature and the relevant case studies. The second review took place at the end of the study by getting on board views of stakeholders who participated in interviews and focus groups.

4.4. Summary

In this chapter, the method followed in this research study was described including the research stance, as well as the research process followed. The next chapter explains how this research study contributed in the field of learning analytics in higher education.

Chapter 5 – Scope of Analytics in Education

In this chapter, a review of current practices in learning analytics is provided. This is the primary data collection phase of the research study, enabling to understand how learning analytics are used in certain institutions. The chapter presents the views of different stakeholders on learning analytics and explain the way learning analytics are used at Middlesex University.

5.1. Review of Current Practice

In its earlier stages, this research study was based on two sources of secondary data. First, the study was based on an initial literature review that provided the necessary understanding of the learning an analytics field. The review of the literature enabled the researcher to assess the role of data analytics in higher education, as well as determine the sources of learning analytics data. Furthermore, the literature review provided the foundations for discussing different uses of data analytics in Higher Education Institutions and indicated some of the most common benefits for universities from the use of learning analytics. Next, the study was concentrated on a detailed analysis of several case studies of learning analytics use in educational institutions. Each case study was discussed with respect to the institutional objectives for exploiting data analytics, as well as the way the institutions collected and analysed data.

As a result of the work on secondary data, the preliminary conceptual framework for this research study was formed. The framework served as a reference point for the next stages of the research work, demonstrating the research stance and how it would help to create the research contributions. It was necessary to reflect whether the analysis of the secondary data was in line with the way learning analytics are planned, implemented and used in higher education. This led to the collection of primary data with the use of focus groups and interviews. The nature of this research and the focus on institutional use of data analytics meant that the most suitable instruments for data collection would be qualitative rather than quantitative. Collecting rich input from the reflections of individuals who are aware of how data analytics affect their roles was deemed far more appropriate rather than using questionnaires. A statistical analysis of questionnaires could provide some insights on the views of staff or students for the role of data analytics but would not offer the necessary contextualisation.

A combination of focus groups and interviews offered the necessary variety on data collection, meaning that both facilitated discussions and one-to-one exchanges were used to investigate the different views of the participants. A focus group involving teaching staff in the Computer Science Department at Middlesex University enabled the researcher to facilitate discussions between individuals with different teaching roles. The group included the teaching team involved in three undergraduate modules, covering levels 4, 5 and 6. This meant that the participants could reflect on issues associated with the first, second and third year of undergraduate studies. A module leader, an associate lecturer and several Graduate Academic Assistants (GAAs) were involved in the group, while the group included both full time members and hourly paid lecturers (i.e., temporary staff). The group also included recent graduates who could provide student views on the role of data analytics.

Next, a series of interviews was arranged to collect different views from various institutions. The interviews were structured around the three sections of the C.A.V. framework but allowed further discussion and investigation of practices taking place in the participants' organisations. The ten questions are included in Appendix B and each question included a clarification statement, helping participants to contextualise the question.

Evaluating the use of analytics in higher education

There is significant volume of published works in the field of data analytics for higher education. Prior to the development of the questionnaire for the collection of primary data, it was necessary to assess how other studies approached the same issue. The focus of the work was on determining whether the means used for evaluating the use of data analytics in higher education were appropriate and resulted in accurate findings.

Alexandron et al (2018) tackled a very interesting issue associated with learning analytics results, as they consider whether results from similar studies represent "truthful and honest learning activity". More specifically, they attempted to "evaluate the robustness of learning analytics results when the data contain a considerable number of fake learners". The study attempted coming up with ways to determine the progress of different learner types with emphasis on fake learners. The study concentrates on the notion of 'time spent on resources', which is used to distinguish between reading time (time that the users spent on explanatory pages), watching time (time spent on videos) and homework time (time spent in pages that contain homework items). This approach was very useful for this study as it triggered the investigation to find the best criteria for determining the moist suitable data sets, as well as ensuring the data analysis is adapted for different domains.

This was a critical step for the data collection process, as it was necessary to associate performance indicators with sources of data collection in higher education. For example, typical metrics for measuring student performance include (Alexandron et al, 2018):

- Grade – referring to the "total points earned in the course".
- Ability – using "student's skill in an Item-Response Theory (IRT) model. This metric was used based on the assumption that "IRT ability scores are known to be independent of the problem sets each student tried to solve".
- Weekly Improvement – this metric is interpreted as "the slope of the regression line fitted to the weekly IRT ability measures".
- Proportion Correct on First Attempt (CFA) – this metric is calculated as "the proportion of items, among the items that the student attempted, that were answered correctly on the first attempt".
- Mean Time to First Attempt (TTF) – this metric is based on "the average time it took a student between seeing the item, and making the first attempt".
- Mean Time on Task (TOT) – this is calculated as "the average time the student spent on an item".

In her 'learning Analytics' book chapter, Jesse Welsh explains that "by leveraging the vast amounts of data available, learning analytics offers several meaningful benefits to learners, teachers, and researchers". Welsh (2020) also argues that "through the use of analytics, educational institutions can restructure learning design processes”. In the chapter that is included in the ‘The Students' Guide to Learning Design and Research’ (Kimmons and Caskurlu, 2020), Welsh identifies three primary limitations and criticisms of learning analytics as (i) “data quality concerns”, (ii) “ethical concerns about the ownership and appropriateness of the collection of large amounts of learner data”, and (iii) “the fear of an automated educational system and its effect on student learning”. Based on these concerns, this research study focused on investigating how institutions prepare their data collection process, as well as designing any analysis tasks before the production of their dashboards.

There are several researchers discussing the use of learning analytics in evaluating various types of learning activities. One such study by Groba et al (2014) advocates that the use of learning analytics “reduces significantly the assessment time and helps teachers to understand the learning process of the students”. The authors present a conceptual architecture demonstrating “the relations between the educational data, the data processing algorithms, and the actors involved in learning-teaching processes”. This architecture is illustrated in the following figure.

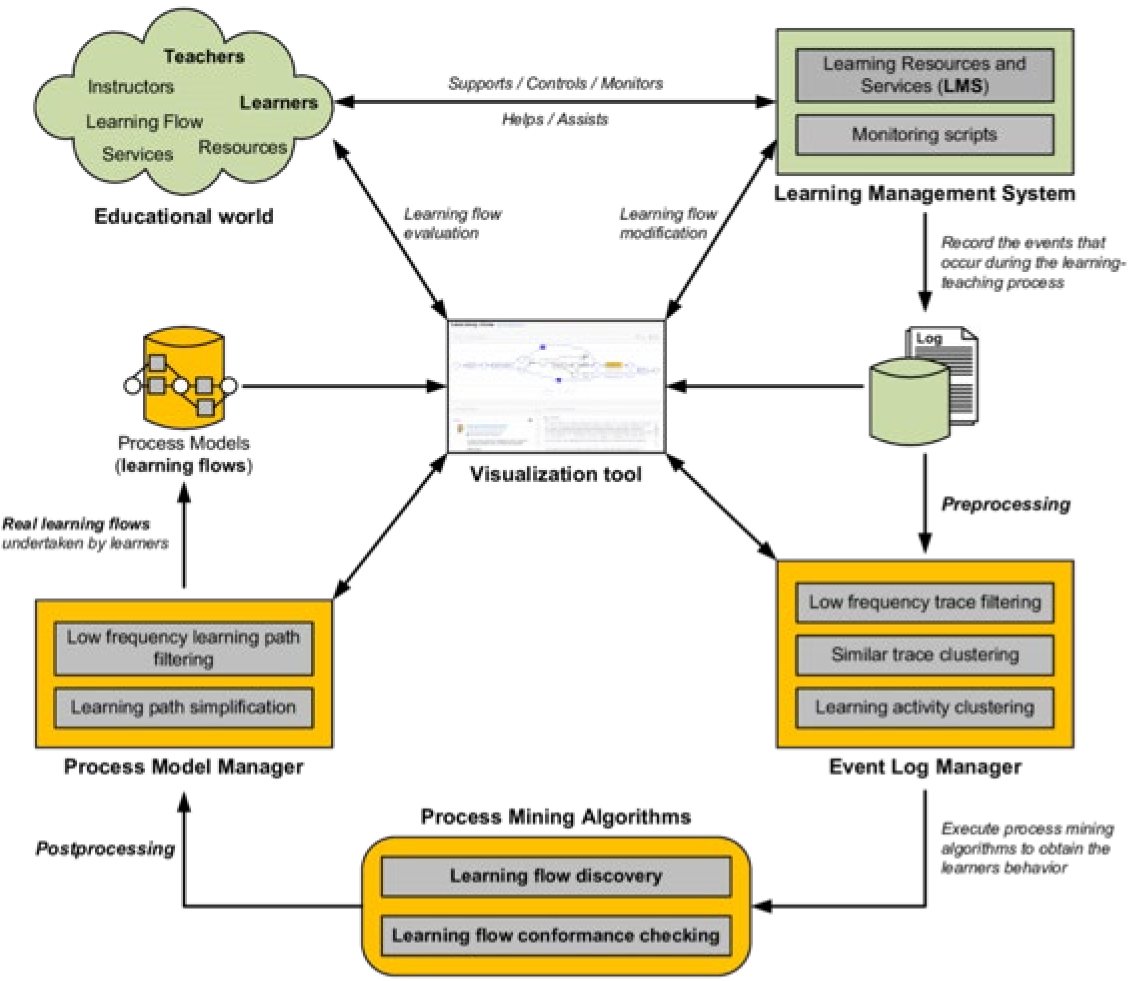

Figure 5-1: Architecture for learning process discovery in self-regulated learning (as cited in Groba et al, 2014).

The key components of the architecture presented by Groba et al (2014) include:

- Educational world – this contains participants of any learning activities that are supported by learning analytics.
- Personal Learning Environments – these are defined as “providing the means to undertake the learning paths designed by the teachers”. The definition is sufficiently generic to cover any virtual learning environment or learning space that enables the collection and analysis of educational data.
- Event log manager – according to the authors creating educational logs involves “strategies for clustering activities based on teacher’s criteria, filtering traces that are not relevant from the point of view of the student’s assessment, and clustering traces when several traces are exactly the same”. For the scope of this research study maintaining logs of learning activity is a process that may span from operational activities (e.g., assessment) to strategic ones (e.g., evaluation of recruitment catchment areas).
- Process mining algorithms – this is described as the core of the architecture. The selection of the algorithms depends on the primary aims for using data analytics in educational contexts.
- Process model manager – according to the authors, “discovery algorithms are developed with the aim of guaranteeing the completeness of the discovered learning processes that describe the learning paths of a course”. Again, as this research study attempts to cover the wider spectrum of educational processes, process modelling varies significantly between strategic and operational aspects of a university.
- Graphical User Interface – this is described as “enabling teachers to understand the students’ behaviour through the visualisation of the learning paths followed by them during a course”. This of course can be extended to any type of visualisation that can support any stakeholder of a higher education institution.

Following from the more generic perspective provided by Groba et al (2014), Martin and Ndoye (2016) focused on reviewing “different categories of online assessments and identify data sets that can be collected and analysed for each of them”. The authors focused solely on assessment data and their visualisations. They have identified a lengthy process that consisted from several steps, covering both formative and summative assessment. These steps include (i) setting learning goals, (ii) providing information (e.g. knowledge, skills and attitudes, (iii) providing students with examples, models and criteria, (iv) supporting student demonstrations, (v) collecting data, (vi) analysing data, (vii) providing feedback, (viii) supporting student revision, (ix) bridging learning gaps, (x) enabling student submission of summative projects, (xi) collecting data, (xii) analysing data, (xiii) providing feedback and (xiv) enabling students to receive data.

A key contribution from Martin and Ndoye (2016) is illustrated in the following figure, where a wide range of assessment types is associated with the necessary learning analytics techniques. Furthermore, each learning analytics technique is mapped to specific data measures required for performing the necessary analysis. As shown in the following figure, data measures include scores, time spent, access frequency, interaction, quality of contribution, quality of reflection, writing skills and quality of evidence.

Types of Assessment	Learning Analytics Techniques	Data Measures
Comprehension type assessment	Quantitative Analysis • Descriptive Statistics • Item analysis	• Current Score • Time Spent • Frequency of access
Discussion board	Social Network Analysis • Interactions between student and facilitator • Interaction among students	Interaction measures • Frequency of Posts • Length of Posts • Themes
	Qualitative Analysis • Discourse Analysis • Conversation Analysis	Quality of Posts • Use of concept and theories • Common patterns • Repeating events • Key phrases
Reflection focused assessments	Qualitative Analysis • Content Analysis, • Concept Mapping • Document Analysis	Quality of Reflection • Rationale • Multiple Perspective • Supporting theories or frameworks • Common patterns • Repeating events • Key phrases • Writing Skills • Grammatical Errors • Typos • Coherence of Ideas Quality of Evidence • Analysis of Artifacts • Type of Artifacts • Justification of Artifacts
Project based assessment	Quantitative Analysis Observation	• Current Score • Time Spent • Frequency of access

Figure 5-2: Different types of Assessments and Learning Analytics (Martin and Ndoye, 2016)

This research study is not limited to the learning analytics associated with assessment; instead, it is intended to provide guidelines across all aspects of the learning process. Nevertheless, the work of Martin and NDoye (2016) serves as a good reference point on how to align different data measures to certain learning analytics techniques and the underlying learning activities.

Student Evaluation and Learning Analytics (StELA) is a project deployed by The University of Limerick (https://www.ul.ie). The project's perspective of learning analytics "involves the collection of educational data, such as grades, survey responses, or number of accesses to online resources from various learning environments to better inform how students learn and engage in their course or programme" (as described at https://www.ul.ie/quality/stela). The project's four key aims are as follows:

- Encouraging student engagement – "to encourage student engagement".
- Building credibility – "to build credibility with staff in the university's student evaluation mechanisms".
- Using existing datasets – "to explore how existing datasets can be used to provide feedback on the student experience".
- Co-creating policy – "to co-create policy on the use of data and on student evaluation".

The StELA project involved a wide range of data collection instruments (as described in detail at https://www.ul.ie/quality/stela), which helped to plan the primary data collection for this research study. Initially, student focus groups were used "to analyse existing approaches and potential improvements top student feedback across a range of disciplines". Subsequently, staff focus groups were utilised to "analyse existing approaches and potential improvements to student feedback across a range of disciplines and learning contexts". A staff survey was used to "explore staff attitudes to the use of educational data and how the use of existing datasets can be used to enhance student learning". Finally, the project focused on student surveys and establishing a feedback mechanism policy, "providing an operational framework to enable a coordinated approach for the implementation of surveys and other student feedback or evaluation mechanisms".

The key stages of the StELA project are illustrated in the following figure, including (i) data extraction, (ii) information transformation, (iii) intervening actions and (iv) review.

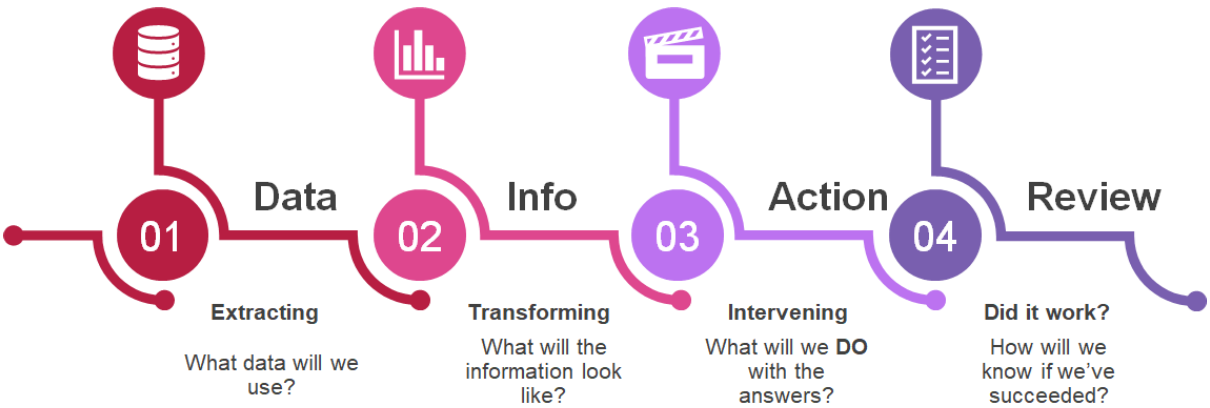

Figure 5-3: The Student Evaluation and Learning Analytics Project (StELA) © https://www.ul.ie/quality/stela

Ali et al (2012) discuss their qualitative evaluation of a learning analytics tool they developed for their institution. Their work focused on evaluating the features of the learning analytics platform as they evolved over two versions. They used the same set of questions three years apart, classified under the following categories:

- Perceived usefulness of the tool for improving course content and instruction.
- Perceived value of the tool's graphical user interface.
- General perception of the tool.

This was a rather useful classification for questions to be used in qualitative analysis of learning analytics. This research study used a similar approach in classifying the questions for the collection of primary data and the evaluation of the C.A.V. framework. The classification used for the questions used in the focus group and the interviews was based on the three core areas of the C.A.V. framework rather than using generic categories such as tool functionalities, usefulness, usability and interface features.

Hilliger et al (2021) discuss the use of a dashboard to "visualise student self-reports of time-on-task regarding subject activities". The authors presented a series of questions used for the evaluation of the dashboard, associated with certain supporting data. The evaluation of the learning analytics dashboard was based on certain dimensions, including data, awareness, reflection and impact. The approach presented by the authors provided a good reference point for this research study while conducting the interviews with different stakeholders of institutional learning analytics dashboards.

The following table summarises how different published works were used to support the decision to use certain questions in the interviews and focus group of this research study. As shown in the table, each of the ten questions is classified in one of the three processes supported by the C.A.V. framework. For each question, a number of support studies are used, justifying the rationale for including the specific question in the data collection. The next sub-section discusses the structure of the questionnaire and how it is formed by ten questions organised under three key dimensions.

#	Question	Process	Support studies
1	How do we determine the necessary data sources for the intended analysis and visualisation?	Collection	Alexandron et al, 2018 Martin and NDoye, 2016
2	How do we assess whether sufficient volume of data is available for the intended analysis?		StELA, 2021 Ali et al, 2012
3	How do we prepare the collected data for the analysis that must be performed?		Welsh, 2020 Groba et al, 2014
4	How do we ensure that data analysis techniques are adapted to suit the specific domain?	Analysis	Alexandron et al, 2018 Ali et al, 2012
5	How do we differentiate between the analysis of strategic and operational data?		Groba et al, 2014 Hilliger et al, 2021
6	How do we confirm the analysis tasks required to produce concrete findings that can be visualised?		Welsh, 2020 Martin and NDoye, 2016
7	How do we select appropriate visualisations for the analysis conducted and the data sets available?	Visualisation	StELA, 2021 Hilliger et al, 2021
8	How do we align the various dashboard components to different decisions that must be made?		Groba et al, 2014 Martin and NDoye, 2016
9	How do we evaluate whether the dashboard contents meet the analysis needs?		Martin and NDoye, 2016 Ali et al, 2012
10	How do we design the dashboard in order to maximise its usability by the intended users?		StELA, 2021 Ali et al, 2012

Table 5-1: Justification of primary data collection questions

<u>Structure of the data collection instrument</u>

Prior to the focus group and the interviews conducted, invited participants were provided with a briefing and the questions used during data collection. Emphasis was given to explain that the focus of the research study was to provide sufficient evidence supporting the value of the C.A.V. framework as a tool for educational analytics. Participants were informed how the C.A.V. framework acted like a bridge between institutional data sources and the design of dashboards used for the representation of (i) operational, (ii) financial, (iii) administrative and (iv) educational data. The participants were also informed about the framework's three phases, namely:

- Data collection
- Data analysis

- Data visualisation

It was also explained to participants that the interview was meant to serve as an evaluation of the key aspects associated with each of the C.A.V. phases. This was the reason for structuring the questionnaire according to the framework’s three sections. The open-ended questions leading to further discussion would enable the researcher and participants to consider administration tasks of educational analytics. The aim would be to find out how the entire process is taking place in participants’ institutions.

The first group of questions focus on **data collection**, aiming to determine how data are collected in the participant’s institution. Three questions are included in the first section of the questionnaire, with the first one considering how the necessary sources of data are identified in higher education. In particular the criteria that are commonly used to determine whether certain sources are useful or not should be discussed. The second question focused on how institutions could establish when sufficient volume of data was reached for the intended analysis to be conducted. Emphasis was given on establishing whether certain thresholds, or benchmarks were used with regards to number of records required for certain types of data analysis. The final question attempts to investigate whether participants’ institutions had their own ‘Extract’, ‘Transform’ ‘Load’ (ETL) processes in place to ensure that the data set is suitable for analysis and visualisation.

The second group of questions focus on **data analysis**, also included three questions. The first question focused on what is required to ensure that data analysis techniques are adapted to suit the specific domain. The question was intended as a prompt to determine changes required in the way data are analysed when working with different business domains, organisation departments, or even data sets. The next question focuses on the different analysis requirements when working with strategic data as opposed to operational data. More specifically, the question is intended to investigate whether the analytical process differs when performed with learning analytics such as student progress and programme evaluation. Finally, the next question attempts to determine the means used in different institutions for confirming which analysis tasks are required to produce concrete findings that can be visualised. In other words, how to decide which calculations must be performed on certain data sets in order to produce effective visualisations.

The third group of questions focus on **data visualisation**. This final part of the questionnaire includes four questions and begins with a prompt on how to select appropriate visualisations for the analysis conducted and the data sets available. This question attempts to identify those criteria should be used to identify the most appropriate visualisation techniques for the data sets available and the analysis required. The next question is concerned with the alignment of the various dashboard components to the different decisions that must be made. Emphasis was given on determining the plans required so certain dashboard areas are mapped to specific decisions that are based on the provided visualisations). The next question considers the ways used to evaluate whether dashboard contents meet the analysis requirements. More specifically, the focus is on identifying certain criteria used to assess whether the dashboard completely meets the needs of its intended users. The final question investigates dashboard design aspects in order to maximise its usability for the intended users. This may involve heuristics, techniques or guidelines that should be followed when creating the visualisation interface of a dashboard.

The questionnaire provided a simple, yet complete set of questions, suitable for gaining an understanding on data collection, analysis and visualisation issues addressed in different Higher Education Institutions. The next sub-section describes how the primary data for this research study were collected and justifies the planning of the focus group and interview sections.

Collection of primary data

Appendix C includes the list of all interviews that took place, as well as the focus group. The appendix also includes the full transcripts of all sessions, after being anonymised. A wide range of stakeholders was involved in the data collection process as shown in the following list. The list follows the chronological order of the sessions organised during the evaluation phase of the C.A.V. framework. The input collected from the different stakeholders was also used to triangulate the findings collected from the literature review and the case study analysis in order to determine the scope of learning analytics in higher education.

The following sessions were used for the collection of primary data during this research study:

- Focus Group (MDX)
  - G.D. Module Leader
  - A.T. Associate Lecturer
  - K.M. Graduate Academic Assistant
  - B.L. Graduate Academic Assistant
  - F.A. Graduate Academic Assistant
- Interviews
  - I.V. Associate Professor / Programme Leader
  - T.K. Assistant Professor / Dean / Head of Department
  - S.K. Associate Professor / Programme Leader
  - G.D. Professor / Director of Programmes / Programme Leader
  - S.R. Data Analytics developer / Alumni
- Focus Group (NUP)
  - S.C. Head of Department / Member of QA committee
  - P.C. Module Leader
  - K.Z. Module Leader
  - S.E. Module Leader
  - E.K. Module Leader

A focus group was organised with the teaching team of a module delivered both in the second and third year of undergraduate programmes offered in both the Computer Science Debarment and the Business School of Middlesex University. The participating teaching staff are also involved in the teaching of other modules across all levels of the undergraduate programmes. The main module taught by the team is delivered across more than 500 students in multiple campuses and involves the delivery of face to face and online sessions as part of the institution's hybrid delivery during the COVID19 pandemic. The focus group discussed how learning analytics are currently used in the module and whether current practices could be deployed across more modules of the taught programmes. Emphasis was given on the way learning analytics affected instructors in delivering their sessions, as well as the impact learning analytics have on student experience and perhaps individual performance. The aim of the focus group was to assess whether the role of learning analytics is affected by the ability and skills of teaching staff to use dashboards in taught sessions.

An interview was conducted with an associate professor responsible to deliver both undergraduate and post graduate modules in the Department of Informatics and Telematics at the Harokopio University of Athens. The focus of the interview was to investigate the way learning analytics could be introduced at department level across the institution, following current attempts to use data analytics in an ad hoc manner. The session also emphasised the need for integrating learning analytics in the delivery of academic programmes.

The next interview was organised with the Dean and Head of the Department of Informatics and Telematics at Harokopio University of Athens. The Dean explained how the decisions for the deployment of learning analytics are made by certain units, at institutional level. The scope of the interview was to understand the decision-making process followed when deciding how to coordinate learning analytics.

A second focus group was organised to collect the views from stakeholders at the Department of Computer Science at Neapolis University of Paphos in Cyprus. The aim was to compare the views from three different institutions in three different countries. The focus group involved five academics and the Head of Department, and the departmental representative at the University's Quality Assurance committee.

Next another round of interviews was organised. The first interview was with a former student of Harokopio University of Athena, who successfully completed a teaching exchange at Middlesex University. The participant also completed an internship at Middlesex as part of the Erasmus+ exchange programme, followed by one year in a development role in a learning analytics software development firm specialising in sentiment analysis. The interview focused on student and alumni views, combined with the perspective of a learning analytics developer.

The final stages of the data collection involved Middlesex University staff. One of the Directors of Programmes in the Computer Science Department at Middlesex University provided insights on how learning analytics are used at institutional and departmental level. The interview also touched upon strategic issues affecting the deployment of learning analytics across the university.

The breadth of roles covered in the multiple sessions, the variety of techniques used for data collection and the in-depth discussions that have taken place during the interviews and the focus groups, should be sufficient for establishing a good understanding of the role learning analytics have in higher education. Furthermore, the involvement of three institutions in the data collection and the evaluation of the framework, represented by both academic and management staff, strengthened the researcher's confidence in the findings of the research study, as presented in the thesis.

Ethics

The role of ethics in learning analytics is discussed in another section of the thesis. With regards to this research study, it was necessary to acquire research approval by the university committee and in particular the sub-committee responsible to scrutinise research in the field of computer science.

From the early stages of the research study, it was clear that data collection would be of minimal risk. The primary data collected were in the form of personal views of participants in focus groups and interviews. These individuals were selected for their expertise, prior experience and knowledge of the subject. The questionnaire used for both focus groups and interviews consisted of ten questions and no personal data were collected or analysed. The individuals were not required to provide any personal data and any information they shared with regards to learning analytics deployment in their institutions would not be identifiable.

Prior to the data collection, the researcher submitted a formal request for ethical approval to the university's ethics committee. This involved the submission of the following documentation:

- Middlesex University Data Protection Checklist and Declaration for Researchers – this was intended for the contextualisation of the data collection and analysis process in order to ensure that its full compliance with the institutional ethics guidelines.
- Participant Information Sheet (PIS) – this provided the necessary information for participating individuals who should be aware of the research scope and the justification for the method followed and the data collected and analysed.
- Consent Form – this included all the necessary clauses for obtaining consent from participants in order to use the provided information.

The necessary documentation is included in Appendix D, including the ethics approval with the reference number as included in the university's system.

5.2. The view of key stakeholders

Once approval was granted the focus of the work shifted towards collecting the views from the identified stakeholders. In this section a synthesis of the different perspectives is attempted, organised around the three sections used fir grouping the ten questions.

It became evident that although all participants had vast experience in data analytics and good knowledge of learning analytics, their responses were affected by the readiness level of their institutions. At strategic level, certain themes emerged with regards to the adoption of learning analytics across the institution, focusing on the need for a detailed stakeholder analysis, evaluation of the impact of learning analytics on the institution and the factors affecting scaling up the deployment of learning analytics. Governance, compliance to legislation and ethical concerns were amongst the issues that were mentioned in most sessions.

When discussing the role of learning analytics in institutional operations, it became evident that dashboard design should be aligned to stakeholder needs that include both academic and business aspects. It is interesting to note that emphasis was given on the fact that learning analytics definitely have a critical role to play in Higher Education Institutions. However, the ability to determine the value added with the use of dashboards at operational level depends on whether university staff are guided towards deterring which key performance indicators should be aligned to dashboard elements. Furthermore, there was a common theme that dashboard planning, design and deployment seems to be the responsibility of central units, which do not necessary liaise with end users when making decisions with regards to learning analytics implementation.

The following sub-sections provide reflections on the findings from the focus group and interview sessions.

<u>Data collection</u>

The first part of the questions focused on aspects associated with the collection of data. This included three questions.

The first question was phrased as follows:

"How do we determine the necessary data sources for the intended analysis and visualisation?"

In summary of the above:

- Institutions must map their data sources and plan for access, control and governance.
- Policies should determine how data sources can be used and the nature of data held.
- End users should be consulted regularly when assessing how institutional data sources can be used in learning analytics.

The second question was:

"How do we assess whether sufficient volume of data is available for the intended analysis?"

In summary of the above:

- Institutions should have in place certain metrics to assure that learning analytics adhere to well-defined thresholds.
- Policies and plans should document the development of learning analytics, offering specific guidelines for the benchmarks, basslines and thresholds used when dashboards are made available.
- End users should contribute in defining the criteria used to decide whether data sets are sufficient for establishing usable learning analytics.

The third question asked:

“How do we prepare the collected data for the analysis that must be performed?”

In summary of the above:

- Institutions should appoint skilled individuals to specific roles in charge of learning analytics.
- Procedures for data preparedness should be documented and disseminated to different stakeholder groups.
- End users should be made aware of the added value from learning analytics including justification for the collection of certain data sets, explanation of the scope of the analysis performed and clarification for any dashboard designs.

Data analysis

The second set of questions emphasised the data analysis process of learning analytics. This included three questions.

The first question focused on:

“How do we ensure that data analysis techniques are adapted to suit the specific domain?”

In summary of the above:

- Institutions should actively seek their advancement to higher maturity levels, aiming to achieve transformation through learning analytics.
- A stakeholder analysis should drive the deployment of any analysis of educational data.
- End users should be able to determine the impact of using a particular dashboard on their role.

The second question asked:

“How do we differentiate between the analysis of strategic and operational data?”

In summary of the above:

- Institutions should deploy a learning analytics strategy that clearly distinguishes the deployment of dashboards at strategic and operational levels.
- Planning checklists should drive the creation of new dashboards and their deployment at strategic or operational level.
- End users should have access levels to the dashboard repository according to the need for certain dashboard elements.

The third question considered:

“How do we confirm the analysis tasks required to produce concrete findings that can be visualised?”

In summary of the above:

- Institutions should justify the deployment of learning analytics by aligning dashboard features to certain role responsibilities and functions.
- Dashboard contents should be decided by an institutional unit that represents all stakeholder groups and has clarity of stakeholder needs and intended use of learning analytics.
- End users should receive sufficient documentation and support in adopting the use of learning analytics in their role.

<u>Data visualisation</u>

Finally, the third part of the questionnaire was concerned with aspects associated with data visualisation. This part included four questions.

The first question asked:

“How do we select appropriate visualisations for the analysis conducted and the data sets available?”

In summary of the above:

- Institutions should align learning analytic requirements and dashboard features to their strategic and operational plans.
- The governing body of the learning analytics strategy and associated processes should consist of visualisation experts and representatives from different user groups.
- End users should be consulted at dashboard design stage.

The second question focused on:

“How do we align the various dashboard components to different decisions that must be made? “

In summary of the above:

- Institutions should be able to justify the use of dashboards by different committees and roles.
- The learning analytics governing body should provide sufficient documentation for each dashboard.
- End users should have access to a centralised dashboard repository with detailed supporting documentation.

The third question considered:

“How do we evaluate whether the dashboard contents meet the analysis needs?”

In summary of the above:

- Institutions should adopt a learning analytics iterative approach including specific steps for testing and evaluating the impact of the produced analysis.
- The learning analytics governing body should have a clear plan about the evaluation and impact assessment of dashboard designs.
- End users should be involved in assessing the usability and effectiveness of dashboard designs.

The fourth question emphasised on:

"How do we design the dashboard in order to maximise its usability by the intended users?"

In summary of the above:

- Institutions should openly involve stakeholder groups in periodic evaluation of dashboard designs.
- The learning analytics governing body should introduce certain heuristics to be used for the evaluation of dashboard designs.
- End users should be actively involved and consulted in the evaluation of dashboard designs.

<u>Data collection reflections – Collection</u>

During the data collection stage, the focus groups and interviews provided very interesting findings. It was evident that there was a significant difference in the way participants approached the discussion and in particular certain questions, depending on the maturity of learning analytics deployments at their institution.

The MDX focus group identified a wide range of areas that could be used for data collection and learning analytics. It was evident that the institution had learning analytics integrated in its teaching practices, which made both senior and junior staff aware of the opportunities to collect and analyse educational data sets. The teaching team was also very clear about the processes followed in order to prepare the data in order to be in a suitable format for further analysis. It was obvious that there were no specific criteria used to assess whether sufficient data were collected in order to have concrete findings following learning analytics.

The HUA interviews indicated that although the institution has identified areas where educational data could be collected, there was still a limited range of data sources, primarily in the form of student evaluations for individual modules and specific KPIs discussed in annual performance reports at department level. The institution has certain thresholds in order to perceive certain student feedback sessions as worthwhile for further analysis. For example, less than ten responses per module are not considered to be representative sample in order to draw concrete findings. It appears that a centralised QA unit is in charge of driving any developments in this area.

The NUP focus group provided an interesting perspective, as their approach towards the use of educational data changed following the validation of their programmes. Their sources of data also included reports from their learning management system. The institution requires a rather high threshold that exceeds 60% participation in order to consider data samples as representatives. Furthermore, GDPR compliance and other QA practices are involved before permitting any analysis.

From the interview and focus group sessions, it is safe to conclude that:

- Institutions should provide a central repository where all possible sources of educational data being collected across the various departments and units are listed. These sources should be described in detail, and the nature and type of available data should be explained. The available data should be checked for accuracy and compliance to legislation and good practice (e.g., GDPR).
- Institutions should agree certain thresholds, for educational data to be labelled as representative sample. In cases where the threshold is not met a clear statement should follow any attempt for learning analytics, informing the users about the sample size and the institution's policy for learning analytics baselines.
- Institutions should provide sufficient guidelines for data preparation (e.g., ETL) and offer training to all staff involved in data collection and analysis. A validation service would be ideal to ensure that data adhere to appropriate formats. Alternatively, a checklist could be used for assessing the state of data prior to conducting any analysis.

Data collection reflections – Analysis

The MDX team appears to be well informed about the features offered by data analytics applications. The teaching staff demonstrated a good understanding on how to adapt different analysis techniques in order to extract meaning and findings from educational data. However, it was evident that the participating staff were not aware about opportunities for learning analytics at strategic level. It was also clear that staff was exposed to good practice and were experienced in reflecting on how learning analytics were used in educational scenarios that were comparable to their own educational activities.

The HUA team shared their experience on how for the first time certain KPIs were used to drive the analysis of educational data. Across the institution, departments have an identified individual who participates in a QA committee that amongst other issues, deals with analysis of educational data. Although at department level discussions are possible to drive requests for further analysis to be performed, usually such decisions appear to be driven centrally by the QA unit.

AT NUP the Senate together with the scientific and business boards make decisions about the way educational data are analysed. The university has in place a total of five surveys used for collecting important data, which are then analysed for departments to use. To a certain extent this is required by the Ministry of Education, while the Senate is responsible to decide how detailed the analysis is. Certain data relating to QA of academic programmes were introduced after the validation of certain programmes by their UK partner. Nevertheless, there is a gap in analysed data at strategic level (e.g., financial performance, recruitment progress).

From the interview and focus group sessions, it is safe to conclude that:

- Institution should provide detailed descriptions of any analysis performed to educational data with adequate justification for stakeholders who need to use the findings. It is also important to have a clear process for a bottom-up approach in identifying required use of learning analytics. For example, departments should be given the opportunity twice a year to make motions for expanding the learning analytics provision to meet certain needs.
- Institutions should have a clear policy on how operational and strategic level learning analytics are decided and produced. This policy should be open to all staff and discussed in training sessions so all staff are aware and able to use the available resource.
- Institutions should document and disseminate the decision-making process followed in order to authorise learning analytics centrally. Learning analytics should be the responsibility of certain committees and specific roles at department level should be accountable to revise the state of available learning analytics. Ad hoc learning analytics at small scale within departments should be permitted if authorised by the department head and/or appropriate committee.

<u>Data collection reflections – Visualisation</u>

At MDX a central unit is responsible for generating the necessary visualisations, responding to requests from central services. It is apparent that the process followed for the creation of university-wide dashboards was not clear amongst teaching staff. It is assumed that such requests should be discussed initially with the head of department.

HUA's data visualisation initiatives are driven by the central QA unit. Departments set their own KPIs and departmental influence is possible to drive the institution's visualisation initiatives. Further clarity on the process followed is required.

NUP's data visualisations are discussed by the academic and business boards. These committees are central to the institution and their responsibilities include the generation of sufficient information for departments.

From the interview and focus group sessions, it is safe to conclude that:

- Institutions should support both top-down and bottom-up process flows that lead to the creation of data visualisations. Different stakeholders such as academics, recruiters, senior managers and administrators should have access to the process in order to offer their input with regards to the dashboards required in their roles.
- Institutions should introduce a form for requesting data visualisations and dashboard components. This form should include fields such as (i) objective, (ii) appropriate use, (iii) restrictions, (iv) compliance and (v) justification.
- Institutions should also introduce a form similar to the above to revise dashboards and ensure that they remain relevant. The forms should be filled by relevant stakeholders and the revision process should be coordinated by a central unit, involving stakeholders from different departments.
- Institutions should have in place a working group of experts in data visualisation and power users of educational data and learning analytics, responsible for (i) evaluating

dashboard designs, (ii) assessing dashboard effectiveness, (iii) identifying usability issues and dealing with requests for (iv) new dashboards or (v) changes to existing ones.

5.3. The Middlesex University Case Study

The previous sections explained how the primary data collection interviews were designed and the input received from stakeholders with respect to their requires for learning analytics. In particular, stakeholder input focused on issues relating to data collection, analysis and visualisation.

With regards to Middlesex University, end users access the available dashboards in three different ways. Staff can access the full set of dashboards by browsing the different categories, as described at a later chapter. However, this approach can be really time consuming, as there are almost thirty dashboard categories, with most of them containing quite a few dashboards sub-groupings. There is no specific policy that documents the strategy governing how decisions are made in relation to dashboard creation and deployment. As mentioned later on, this is an area where this research study attempts to contribute with the proposed C.A.V. framework.

The second option for staff is to access dashboards recommended for their roles. Again, there is no specific set of guidelines explaining how these dashboards are selected. The following figure illustrates the top six dashboards displayed when the researcher selected the 'for you' recommendation type. The researcher is an associate lecturer and module leader and it appears that the dashboards loading first are intended for these two roles. The dashboards include:

- Applicant programme numbers including students with disabilities – this dashboard helps module leaders and teaching staff to plan accordingly for their taught sessions.
- Programme and module leaders report – this dashboard offers a full record of students enrolled to the different programmes and modules.
- SLA programme evaluation student survey – this dashboard provides an overview of student feedback for the role of Student Learning Assistants and the impact they have on their learning experience.
- Personal tutor allocation list – this dashboard provides a full list of tutees allocated to each academic member of staff with useful information for each student.
- Progression and achievement report – this dashboard provides staff with a breakdown of student results and the necessary statistics for progression and achievement for different modules and programmes.
- Widening participation – this dashboard provides information about the participation of students, classified into young and mature cases (depending on whether they are younger or older than 21 years).

These dashboards are displayed as shown in the figure below.

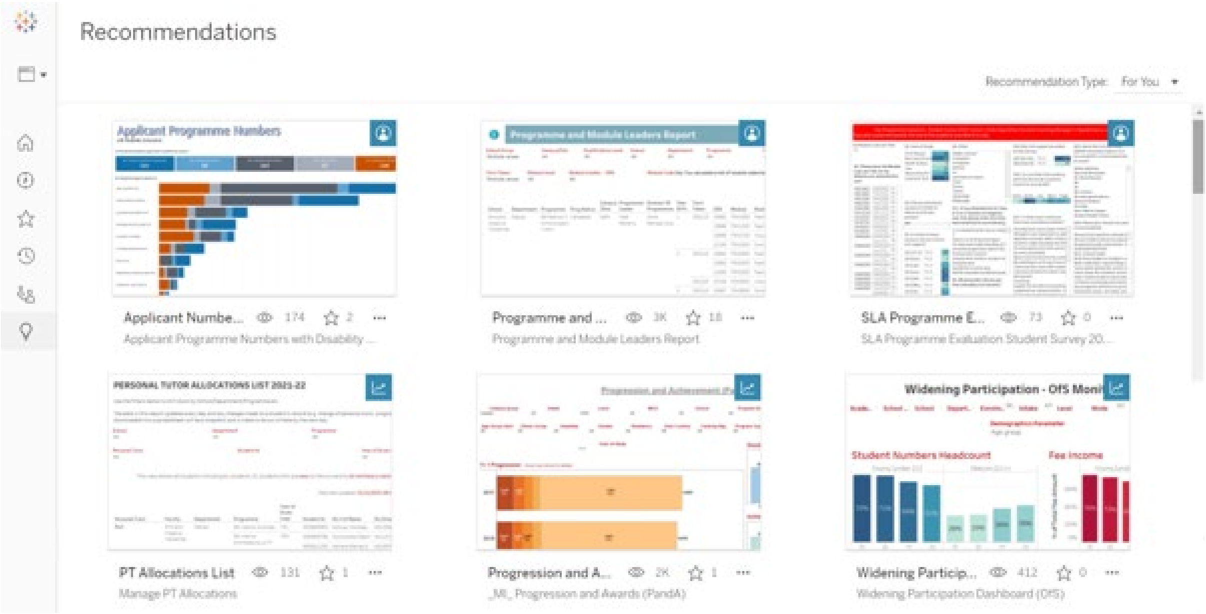

Figure 5-4: Middlesex University dashboards prioritised according to the profile of the thesis author

When selecting the ‘trending’ option for filtering the various dashboards, a slightly different selection is displayed, as shown in the figure below. The dashboards that seem to be trending and are different from the ones in the above list are as follows:

- Student profile report – this dashboard provides a detailed report on attendance and engagement for each student, something that is critical, especially for international students, during the COVID19 pandemic.
- Student modules pass rate – this dashboard offers statistics for different modules and the way students perform by showing the pass rates and the grades achieved overall.
- Report repository – this dashboard includes some research information on learning analytics for student biometrics and its popularity is affected by the frequency of data access by the members of the research group working on the student records.

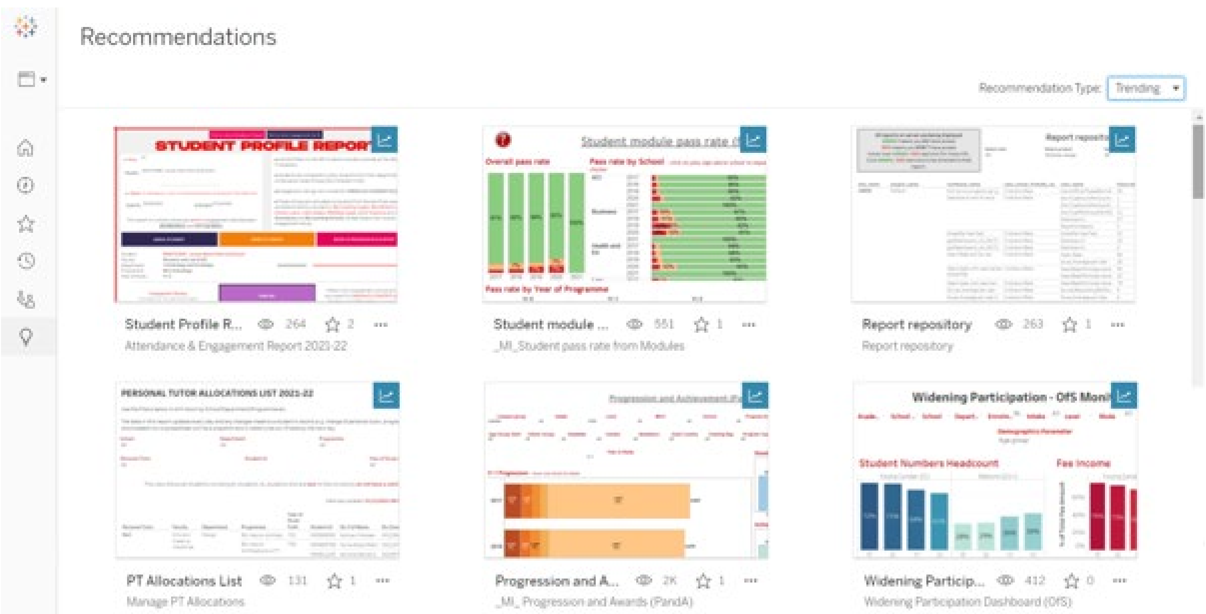

Figure 5-5: Middlesex University dashboards prioritised according to current trends (January 2022)

The following chapter discuses aspects associated with the planning of a learning analytics strategy. More specifically, it attempts a second phase of literature review in order to revise the key components of the C.A.V. framework. The scope of the chapter is to provide more detail in the form of guidelines supporting the deployment of learning analytics at both strategic and operational levels in higher education.

5.4. Summary

In this chapter the different stakeholder views on learning analytics were presented and the Middlesex University case study was presented. The next chapter explains how the proposed C.A.V. framework was revised to include guidelines at strategic and operational level.

Chapter 6 – Planning a Strategy for Data Analytics in Education

This chapter includes the second phase of the literature review that was conducted to ensure that the proposed conceptual framework was in line with similar approaches in the field. The revised C.A.V. framework is described in detail, as well as the production of general institutional guidelines, as well as specific guidelines intended for end users.

6.1. Revised Framework

The C.A.V. framework was presented in a previous section of the thesis, and in particular how it can provide (i) generic guidelines for strategic planning of data analytics in a HEI and (ii) specific guidelines on how to deploy data analytics for different HEI operations. The main elements of the C.A.V. framework were determined from an extensive analysis of case studies describing the use of data analytics in higher education. The findings of the analysis included a number of features necessary for learning analytics in Universities. These findings were strengthened by a review on how learning analytics are used in higher education. This first stage of the literature review was followed by further analysis of the literature after conducting the primary data collection. The rationale was to reflect whether the responses from participants in the focus group and interviews were in line with the predominant views in the relevant literature.

Banoor et al (2019) conducted a systematic literature review to look at "how learning analytics have been used to model learner engagement in online courses and how their engagement has influenced their performances", using Cooper's taxonomy as their research method. The aimed at investigating how learning analytics can be used for modelling students' engagement in online courses, but also for predicting their performance. It is evident that student **engagement** and **performance** are two quite common dimensions used in learning analytics. The following figure illustrates a very useful summary of quality indicators used in learning analytics (Scheffel et al, 2014). These indicators can be used for evaluating the use of learning analytics in educational practice (Banoor et al, 2019).

A shown in the figure, the quality indicators are classified into five categories, namely (i) objectives, (ii) learning support, (iii) learning measures and output, (iv) data aspects and (v) organisational aspects. The indicators classified under objectives include awareness, reflection, motivation and behavioural change, which are in line with the C.A.V. framework's aim to provide the necessary dashboards for raising institutional awareness and reflection. Learning support indicators include perceived usefulness, recommendations, activity classification and detection of student cases (e.g., those at risk). These indicators are addressed by the C.A.V. framework as part of the production of analytics for educational and administrative data. The next set of indicators include comparability, effectiveness, efficiency and helpfulness, which are grouped under learning measures and outputs. These indicators can be classified under either business or academic data dashboards. The next group of quality indicators includes transparency, data standards, data ownership and privacy. These indicators are in line with the C.A.V. framework's funnel used for collecting data from various sources. Finally, the organisational aspects in the literature are represented by a number of quality indicators including availability, implementation, training of educational stakeholders and organisational change. These indicators correspond to the C.A.V. framework's business data that are further classified as organisational and financial data.

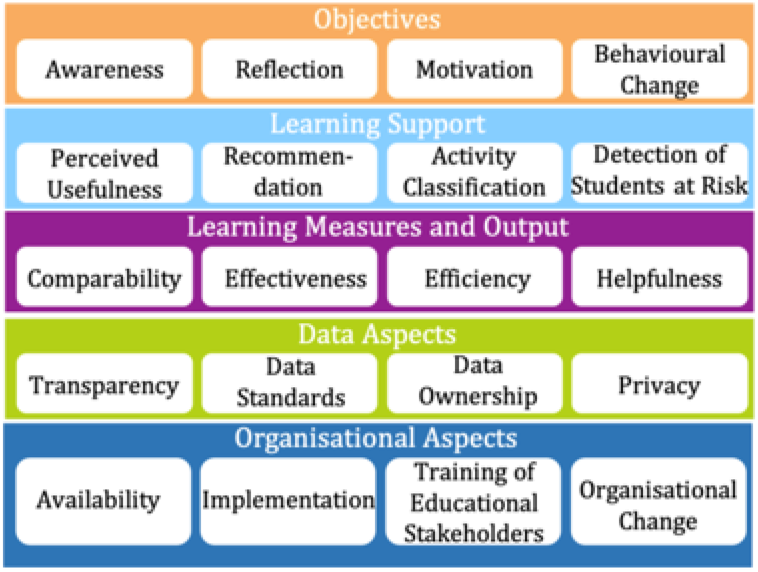

Figure 6-1: Quality indicators for learning analytics (Scheffel et al., 2014)

Quality indicators are integrated in all three key components of the C.A.V. framework, as they are used as data source features, functions of data collection, analysis and visualisation, as well as dashboard features. Sin and Muthu (2015) discuss the application of bug data in education data mining and learning analytics. In their paper they discuss the various techniques that can be used for educational analytics including regression, nearest neighbour, clustering and classification. The authors also discuss applications of big data techniques in learning. The various applications discussed in their paper are retrieved from Mavrikis et al (2013) and include the following:

- Performance prediction – based on analysing student interactions with instructors and peers.
- Attrition risk detection – based on identifying evidence for reduced student engagement.
- Data visualisation – based on the range of visualisation techniques that can be selected according to the dashboard user needs to understand data.
- Intelligent feedback – based on the provision of dashboard visualisations that will indicate to students their assessment and any corrective actions expected from them.
- Course recommendations – based on the selection of suitable modules according to student preferences, activities and competencies.
- Student skill estimation – based on assessing the development of student skills.
- Behaviour detection – based on forecasting of individual behaviour according to 61anticipation in team activities.

Matcha et al (2019) discuss Learning Analytics Dashboards (LAD) through an extensive literature review. The authors suggest that "future research and development should not make any a priori de-sign decisions about representation of data and analytic results in learning analytics systems such as LADs." Their work is focused on self-regulated learning and how is involves iterations of a number of phases including (i) task definition, (ii) goal setting and planning, (iii) enactment of tactics and strategies, and (iv) adaptation (Winne and Hadwin, 1998). According to Winne and Hadwin (1998) there are five components that run recursively throughout the learning cycle including Conditions, Operations, Products, Evaluation, and Standards, forming the COPES model. Matcha et al (2019) use the COPES model introduced by Winne and Hadwin (1998) for the classification of dashboard themes.

Khalil (2017) discusses the role of learning analytics in Massive Open Online Courses (MOOCs). The following figure illustrates a dashboard design that was developed as part of a prototype system for assessing a range of learning data. The areas covered by the dashboard include the following:

- Quiz results, self-assessment, scores and grades
- Downloaded files
- Login frequency
- Forum reading frequency
- Forum posting frequency
- Watched videos

These areas can be used for assessing student performance as part of the C.A.V. educational data but may also provide some useful administrative data sets. As shown in the following figure, the dashboard includes the full range of activities demonstrating student progress and interactions.

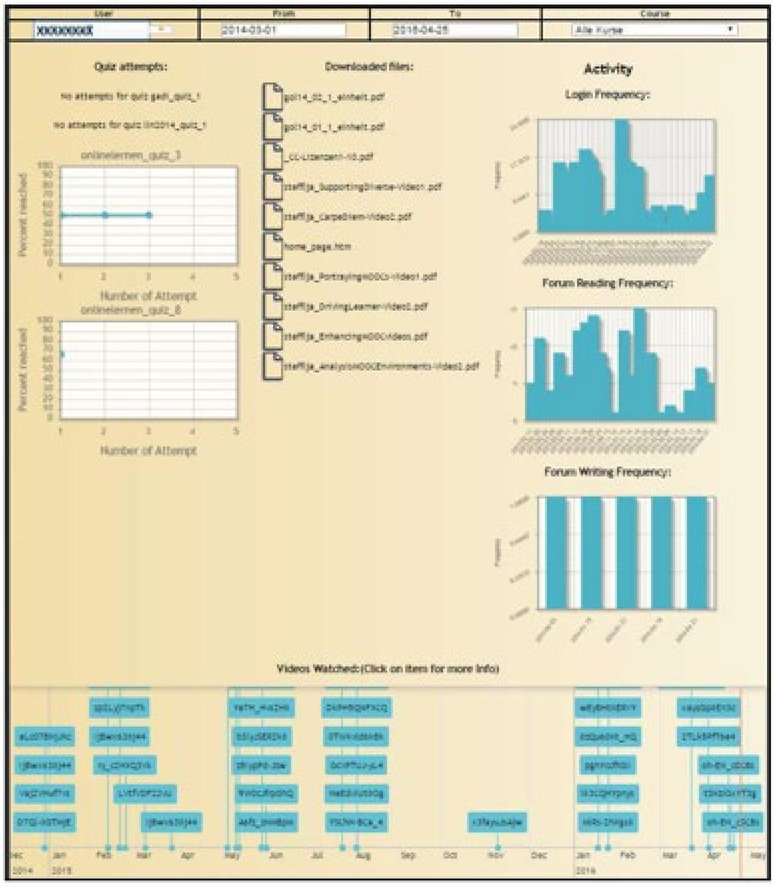

Figure 6-2: Sample dashboard interface (Khalil, 2017)

Chatti et al (2012) describe a reference model for learning analytics based on “four dimensions, namely data and environments (what?), stakeholders (who?), objectives (why?), and methods (how?)”. The four dimensions that are illustrated in the figure below, are described as follows:

- What? “What kind of data does the system gather, manage, and use for the analysis?”
- Who? “Who is targeted by the analysis?”
- Why? “Why does the system analyse the collected data?”
- How? “How does the system perform the analysis of the collected data?”

These dimensions are used to provide certain key elements for the C.A.V. components, including data sources, the main functionalities and design aspects of the produced dashboards.

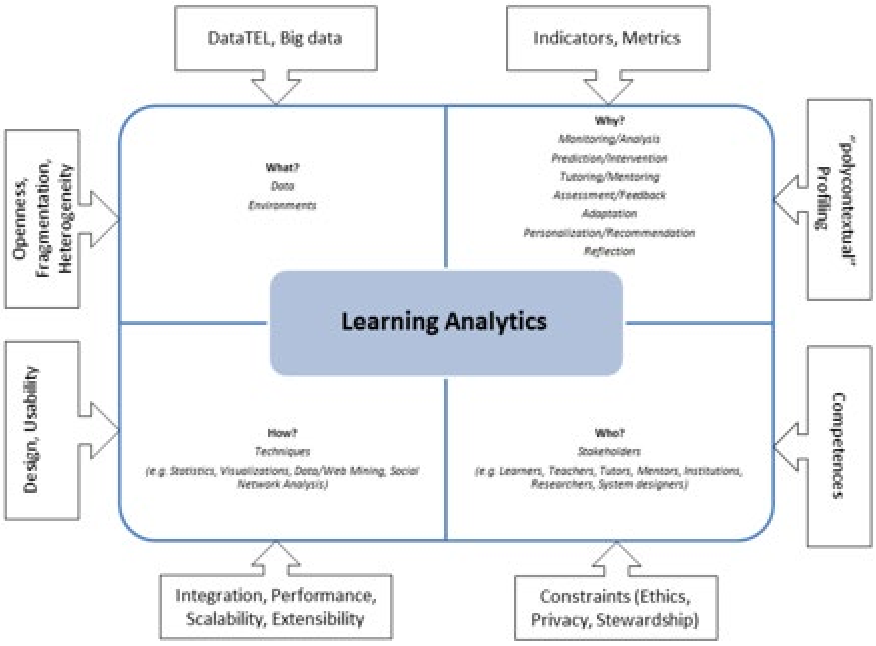

Figure 6-3: Learning analytics reference model (Chatti et al, 2012)

A very interesting review of empirical research in self-regulated learning and learning analytics is provided by Viberg et al (2020). The authors concluded that “there is a critical need to exploit the learning analytics support mechanisms further in order to ultimately use them to foster student self-regulated learning in online learning environments”. It is interesting to note that the authors suggested that scholars should adopt an ethical approach towards learning analytics in a more systematic way. It appears that the most common methods for data analysis adopted in the literature include data distillation for human judgment, prediction, clustering, relationship management and use of models from previous studies. The authors used the Zimmerman self-regulated learning model to organise their literature review (Viberg et al, 2020). The dimensions that are considered to be useful for the purpose of this research study included (i) goals, (ii) time management, (iii) motivation, (iv) planning, (v) self-efficacy and (vi) awareness.

According to Oliva-Cordova et al (2021) the main benefits of applying learning analytics can be summarised in the areas of (i) curriculum and assessment, (ii) pedagogical mediation and (iii) application of digital skills. This is illustrated in the figure below. The authors also identified a wide variety of application areas for learning analytics as illustrated in the word cloud below.

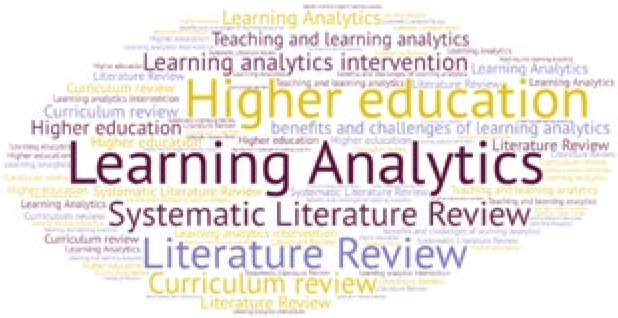

Figure 6-4: Areas of application of learning analytics (Oliva-Cordova et al, 2021)

It is important to determine the main purposes of learning analytics in higher education as discussed in the relevant literature (Oliva-Cordova et al, 2021). For the purpose of this research study, the following purposes were considered as suitable for generating visualisations of financial data:

- Teaching intervention effort
- Tutoring and mentoring effort
- Personalisation support costs

In the literature review conducted by Leitner et al (2017) the concept of 'actionable intelligence' from data mining is discussed as "supporting the teaching and learning and provides ideas for customization, tutoring and intervention within the learning environment" (Campbell and Oblinger, 2007). The analysis process is described as a series of five steps, illustrated in the following figure. The five steps include (i) capturing, (ii) reporting, (iii) predicting, (iv) acting and (v) refining. These five steps help to further decompose the analysis process as described in the C.A.V. framework.

Figure 6-5: The five steps of the analysis process (Campbell and Oblinger, 2007)

The outputs of the analytics process, although they are primarily focused on learning activities, it is beneficial for different stakeholders. According to Romero and Ventura (2013) there are four primary stakeholder groups, namely:

- Learners – "support the learner with adaptive feedback, recommendations, response to his or her needs, for learning performance improvement".
- Educators – "understand students' learning process, reflect on teaching methods and performance, understand social, cognitive and behavioural aspects".

- Researchers – “use the right data mining technique which fits the problem, evaluation of learning effectiveness for different settings”.
- Administrators – “evaluation of institutional resources and their educational offer”.

The four stakeholder groups are used in the C.A.V. framework as they provide different viewpoints in the framework’s visualisation functions.

Papamitsiou and Economides (2014) conducted a systematic literature review of empirical evidence in the fields of learning analytics and educational data mining. The authors highlighted four main areas of research as follows (Papamitsiou and Economides, 2014):
- Pedagogy-oriented issues – including student modelling, prediction of performance, assessment and feedback, reflection and awareness.
- Contextualisation of learning – focusing on “focus on positioning learning within specific conditions and attributes”.
- Networked learning – considering social aspects of learning, including interactions between learners, as well as learners and the content.
- Educational resources handling – organising and recommending educational resources.

One of the classifications of the case studies investigated, was based on the research objectives driving each case study (Papamitsiou and Economides, 2014). The authors listed the different research objectives that included student behaviour modelling, performance prediction, self-awareness increase, retention prediction, assessment improvement, feedback services and recommendations of resources.

In their literature review, Avella et al (2016) revealed that learning analytics “use various methods including visual data analysis techniques, social network analysis, semantic, and educational data mining including prediction, clustering, relationship mining, discovery with models, and separation of data for human judgment to analyse data”. A key outcome of the review was a series of benefits from learning analytics, including (Avella et al, 2016): (i) targeted course offerings, (ii) curriculum development, (iii) student learning outcomes, (iv) behaviour and process, (v) personalised learning, (vi) improved instructor performance, (vii) post-educational employment opportunities, and (viii) enhanced research in the field of education.

So far in this section a wide range of issues associated with learning analytics was described. These issues helped to formulate the main features and functions of the core components of the C.A.V. framework. Banihashem et al (2018) also discussed ethical, educational, and technical issues in the use of learning analytics in education. Their work did not add any new issues to the ones identified so far in this section, however their contribution to the field included the classification of learning analytics benefits according to the different stakeholders. The main stakeholders gaining from learning analytics were identified as:
- Learners
- Teachers
- Institutions
- Researchers
- Course designers

- Parents

This classification contributed to this research study in the sense that the visualisation of learning analytics considered different dashboard views for administration and teaching. For example, the C.A.V. dashboards may focus on enabling decision-making when planning courses (e.g. enrolling students on modules according to their prior achievement or preferences), making adjustments on course content (e.g. determining whether certain topics affect the assessment results and they need further support), supporting teaching delivery (e.g. identifying areas that have a lower progression rate), enhancing learning experiences (e.g. supporting students that appear to have lower performance) or researching areas for staff development (e.g. using student feedback for the professional development of instructors). The C.A.V. frameworks is focused on Higher Education Institutions, and does not support parent requirements in dashboard design in its current form. However, the administration functions of the framework could be revised in future versions to support additional stakeholders, such as parents, members of boards of governors or even external agencies.

An important are that requires more attention relates to the challenges imposed from ethics and privacy of learning analytics in education along with "the lack of attention to theoretical foundations and scope and quality of data" (Banihashem et al, 2018). This is covered in the C.A.V. framework as part of the data ownership and data stewardship dimensions, meaning that the institutional data officer should be involved in the process of determining the data sources used when conducting data collection, as well as the design of dashboards when performing the data analysis and visualisation.

Concluding the discussion on the C.A.V. framework revisions, it is necessary to refer to the taxonomy provided Mangaroska and Giannakos (2018). In their systematic literature review of analytics-driven design to enhance learning. The authors provide a taxonomy of learning analytics for learning design. The taxonomy, which is illustrated in the following figure provides an excellent mind map of concepts associated with learning analytics (i.e., capturing, reporting, predicting, acting and refining), as well as learning design aspects (e.g., finding and handling information, communication, being productive, experimental, interactive or adaptive). The foundations of the taxonomy are fully aligned with the core functions of the C.A.V. framework functions as listed below (Mangaroska and Giannakos, 2018):

- Student retention
- Student assessment
- Personalised learning
- Usefulness of learning analytics tools
- Predictive modelling
- Overall user satisfaction
- Improved orchestration
- Collaboration and interaction
- Teacher's professional development
- Student self-reflection / self-assessment
- Student learning behaviour and engagement
- Design and management of learning scenarios/activities

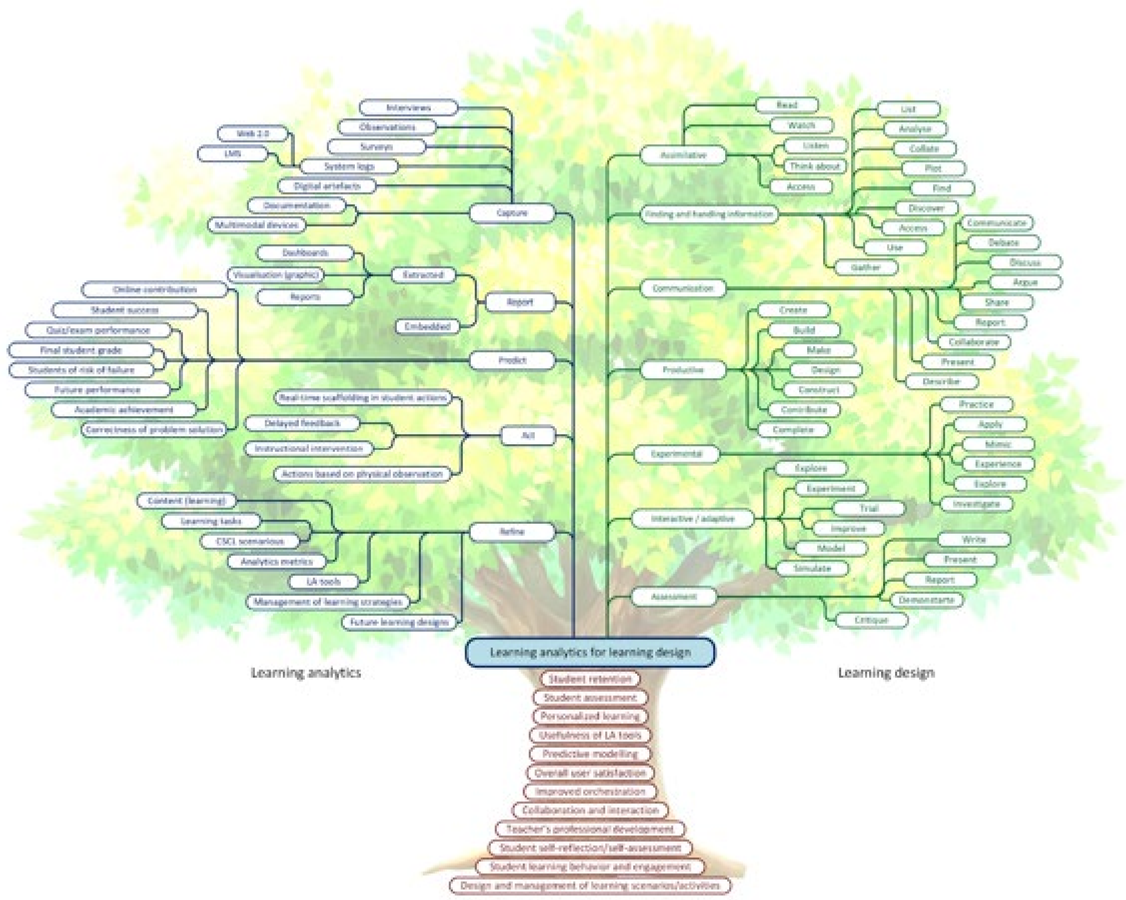

Figure 6-6: The five steps of the analysis process (Mangaroska and Giannakos, 2018)

The C.A.V. framework consists of three key parts. The first part is concerned with the data sources used by institutions performing learning analytics. The following table provides a compilation of the key features that must be considered when selecting data sources at institutional level.

Feature	Description	Source
Awareness	Identifying possible sources of data	Scheffel et al, 2014
Comparability	Being able to assess the richness of data provided	Scheffel et al, 2014
Fragmentation	Retrieving data from various, dispersed sources	Chatti et al, 2012
Heterogeneity	Handling different types of data	Chatti et al, 2012
Openness	Gaining access to the necessary data sources	Chatti et al, 2012
Recommendations of resources	Determining the necessary resources for conducting learning analytics	Papamitsiou and Economides, 2014
Usefulness	Assessing the usefulness of an available data source	Scheffel et al, 2014

Table 6-1: C.A.V. framework – data source features

The core of the C.A.V. framework includes the three components focusing on collection, analysis and visualisation of institutional data. The following three tables summarise the functions that are supported by the framework.

Function	Description	Source
Adaptation	Adapting the collection technique to meet user needs	Chatti et al, 2012
Curriculum development	Determining the performance of learners in different modules	Avella et al, 2016
Data ownership	Controlling who owns and processes the data	Scheffel et al, 2014
Data standards	Working with data that are complete and accurate	Scheffel et al, 2014
Intervention	Enabling tutors to provide feedback based on data sets	Chatti et al, 2012
Mentoring	Selecting data sets that support tutor decision-making	Chatti et al, 2012
Monitoring	Determining the data sets that must be accessible	Chatti et al, 2012
Personalisation	Customising data collection to meet profiling needs	Chatti et al, 2012
Personalised learning	Determining the performance of learners according to different demographics	Avella et al, 2016
Privacy	Ensuring the privacy of personal data	Scheffel et al, 2014
Student learning outcomes	Identifying the performance of learners for different learning outcomes	Avella et al, 2016

Table 6-2: C.A.V. framework – data collection functions

Function	Description	Source
Assessment improvement	Providing evidence of assessment improvements associated with certain learning interventions	Papamitsiou and Economides, 2014
Attrition detection	Detecting students that are at risk of dropping out	Sin and Muthu, 2015
Behaviour detection	Detecting anticipated behaviour based on interactions with others and participation in team activities	Sin and Muthu, 2015
Course recommendation	Identifying appropriate modules and topics aligned to student capabilities, activities and preferences	Sin and Muthu, 2015
Helpfulness	Enabling specific decision making	Scheffel et al, 2014
Performance prediction	Supporting forecasting of assessment results	Sin and Muthu, 2015
Personalisation support costs	Determining the effort (and associated costs) for supporting learning personalisation	Oliva-Cordova et al, 2021
Motivation	Providing the rationale for each analytic task	Scheffel et al, 2014
Self-awareness increase	Providing evidence of learner improvement on learning strengths and weaknesses	Papamitsiou and Economides, 2014
Social networks	Analysing social network interactions	Chatti et al, 2012
Student behaviour modelling	Modelling different behavioural patterns emerging during certain learning activities	Papamitsiou and Economides, 2014
Teaching intervention effort	Determining the effort (and associated costs) for providing teaching interventions	Oliva-Cordova et al, 2021
Transparency	Sharing the full set of calculations performed	Scheffel et al, 2014
Tutoring & mentoring effort	Determining the effort (and associated costs) for providing tutoring and mentoring	Oliva-Cordova et al, 2021
Web browsing	Analysing web navigation patterns	Chatti et al, 2012

Table 6-3: C.A.V. framework – data analysis functions

Function	Description	Source
Administrator view	Demonstrating how institutional resources are used to support learning activities	Leitner et al (2017)
Availability	Providing continuous access to required visualisations	Scheffel et al, 2014
Corporate change	Adapting decision making processes to use dashboards	Scheffel et al, 2014
Educator view	Demonstrating the performance of learning processes	Leitner et al (2017)
Feedback	Offering visual cues of intelligent feedback to students	Sin and Muthu, 2015
Implementation	Creating visualisations as required	Scheffel et al, 2014
Learner view	Demonstrating individual learning progress	Leitner et al (2017)
Performance prediction	Showing predicted performance for individual learners or teams of learners	Papamitsiou and Economides, 2014
Researcher view	Demonstrating the effectiveness of learning activities	Leitner et al (2017)
Retention prediction	Showing predicted rates of student retention	Papamitsiou and Economides, 2014
Skill estimation	Showing the development of individual skillsets	Sin and Muthu, 2015
Statistics	Producing statistical analysis of performance indicators	Chatti et al, 2012
Training provision	Introducing dashboard use as part of certain roles	Scheffel et al, 2014
Visualisation	Selecting the most appropriate data visualisation	Sin and Muthu, 2015

Table 6-4: C.A.V. framework – data visualisation functions

Finally, the C.A.V. framework supports the design and creation of dashboards. The following table lists all the necessary features of institutional dashboards showing the visualisations of learning analytics. As in previous tables, each feature is briefly described and the supporting reference is provided. It is important to note that all functions and features of the framework are supported by papers summing up findings from extensive literature review. The purpose was to ensure that the identified features are a result from as many papers as possible resulting into a list that is widely acceptable.

Feature	Description	Source
Awareness	Demonstrating learner awareness of learning priorities	Viberg et al, 2020
File downloads	Listing the relevant files downloaded by students	Khalil, 2017
Forum usage	Showing forum reading and posting frequency	Khalil, 2017
Goals	Showing ability of learners to achieve certain goals	Viberg et al, 2020
Improved instructor performance	Demonstrating the performance of instructors in association to student evaluations and learner results	Avella et al, 2016
Motivation	Providing evidence for self-motivated learning	Viberg et al, 2020
Quiz results	Showing assessment results, grades and scores	Khalil, 2017
Planning	Providing evidence of self-regulated learning	Viberg et al, 2020
Post-educational employment opportunities	Demonstrating potential employment opportunities according to learner performance	Avella et al, 2016
Reflection	Determining how dashboard content is used	Scheffel et al, 2014
Stakeholder views	Providing views for different users (learners, teachers)	Chatti et al, 2012
Self-efficacy	Providing evidence of learners' confidence	Viberg et al, 2020
Time management	Demonstrating ability to maintain progress according to a specific learning plan	Viberg et al, 2020
Videos watched	Providing the history of watched videos	Khalil, 2017

Table 6-5: C.A.V. framework – dashboard features

The next tables show the different types of data that should be included in the dashboards produced as part of the deployment of the C.A.V. framework at institutional level. Dashboard data are classified primarily either as business or academic. Universities will benefit from dashboards showing their operational performance in a range of areas such as data openness (e.g., how accessible institutional data are from different departments) and data fragmentation (e.g., how data are drawn from different organisational units

- Operational data
  - Awareness of learning priorities – Viberg et al, 2020
  - Corporate change – Scheffel et al., 2014
  - Curriculum development – Avella et al, 2016
  - Ethical clearance – Chatti et al, 2012
  - Extensibility of data sets – Chatti et al, 2012
  - Fragmentation of data sets – Chatti et al, 2012
  - Heterogeneity of data sets – Chatti et al, 2012
  - Improved instructor performance – Avella et al, 2016
  - Integration of data sets – Chatti et al, 2012
  - Openness of data sets – Chatti et al, 2012
  - Personalised learning – Avella et al, 2016
  - Post-educational employment opportunities – Avella et al, 2016
  - Privacy constraints – Chatti et al, 2012
  - Recommendations of resources – Papamitsiou and Economides, 2014
  - Reflection – Scheffel et al., 2014
  - Scalability of data sets – Chatti et al, 2012
  - Stewardship of data – Chatti et al, 2012
  - Student learning outcomes – Avella et al, 2016

Table 6-6: C.A.V. framework – operational data

The C.A.V. framework emphasises the need for financial data visualisation. Increasingly, senior management decisions affecting academic provision are driven by institutional finances. A typical example would be the decision whether to run a programme that seems to have high progression rates and excellent overall assessment results but has a low number of enrolled students. The original version of the C.A.V. framework mentions revenue and profits which are based on the number of enrolled students, fees paid, and costs associated with the delivery of programmes such as staffing and infrastructure costs. Unfortunately, there are almost no resources discussing the role of financial data in educational analytics. However, this appears to be an opportunity for further work as there is a gap in the relevant research covering financial aspects of education that can be part of data visualisations.

- Financial data
  - Teaching intervention effort – Oliva-Cordova et al, 2021
  - Tutoring and mentoring effort – Oliva-Cordova et al, 2021
  - Personalisation support costs – Oliva-Cordova et al, 2021

Table 6-7: C.A.V. framework – financial data

The next two tables provide the data elements that are expected to be part of academic data visualisations. The majority of the publications in the field discuss issues associated with educational data and are concerned with learning activities. Typically learning analytics are concerned with assessment results, communication and interactions. Academic data analytics are also concerned with administrative data such as progression rates, use of academic resources and facts about educational processes.

- Administrative data
  - Awareness – Viberg et al, 2020
  - Attrition detection – Sin and Muthu, 2015
  - Browsing patterns – Chatti et al, 2012
  - Course enrolment (study planning) – Matcha et al., 2019
  - Course recommendation – Sin and Muthu, 2015
  - File downloads – Khalil, 2017
  - Learning design – Matcha et al., 2019
  - Monitoring patterns – Chatti et al, 2012
  - Performance prediction – Papamitsiou and Economides, 2014
  - Planning – Viberg et al, 2020
  - Retention prediction – Papamitsiou and Economides, 2014
  - Skill estimation – Scheffel et al, 2014
  - Social network use – Chatti et al, 2012
  - Student behaviour modelling – Papamitsiou and Economides, 2014
  - Students at risk – Scheffel et al, 2014
  - Time management – Viberg et al, 2020
  - Videos watched – Khalil, 2017

Table 6-8: C.A.V. framework – administrative data

- Educational data
  - Assessment improvement – Papamitsiou and Economides, 2014
  - Behaviour detection – Sin and Muthu, 2015
  - Competency development – Matcha et al, 2019
  - Emotion indicators – Matcha et al, 2019
  - Feedback to students – Scheffel et al, 2014
  - Forum usage – Khalil, 2017
  - Gamification indicators – Matcha et al, 2019
  - Learning difficulty detection – Matcha et al, 2019
  - Motivation – Viberg et al, 2020
  - Performance prediction – Sin and Muthu, 2015
  - Quiz results – Khalil, 2017
  - Self-awareness increase – Papamitsiou and Economides, 2014
  - Self-efficacy – Viberg et al, 2020
  - Social network interactions – Chatti et al, 2012
  - Students' learning progress – Matcha et al, 2019
  - Teamwork progress – Matcha et al, 2019
  - Web navigation history – Chatti et al, 2012

Table 6-9: C.A.V. framework – educational data

6.2. Institutional Guidelines (strategic level)

In the previous sub-section, the C.A.V. framework was revised in the light of a more up-to-date review of published works including literature reviews in the field of learning analytics. The main features and functions of the C.A.V. components were described in more detail. The next two sections of this chapter include the remaining contributions of this research study, namely institutional guidelines at strategic level and user guidelines at operational level. This section describes in detail how Higher Education Institutions (HEIs) can put together their data analytics strategy. The section does not focus on policy making, but instead it suggests four key stages of strategic planning, with a number of associated steps and processes. The C.A.V. framework is integrated in the institutional guidelines that are organised around the following stages:

A. Assessing the institution's maturity in Learning Analytics (LA).
B. Reflecting on the institution's stage of Learning Analytics (LA) development.
C. Identifying the main themes driving the adoption of Learning Analytics (LA) across the institution.
D. Checking the level of completion of the institution's Learning Analytics (LA) plans.

The four stages are illustrated in the following figure, providing a pictorial representation of the approach proposed by this research study in establishing an institutional strategy towards Learning Analytics (LA).

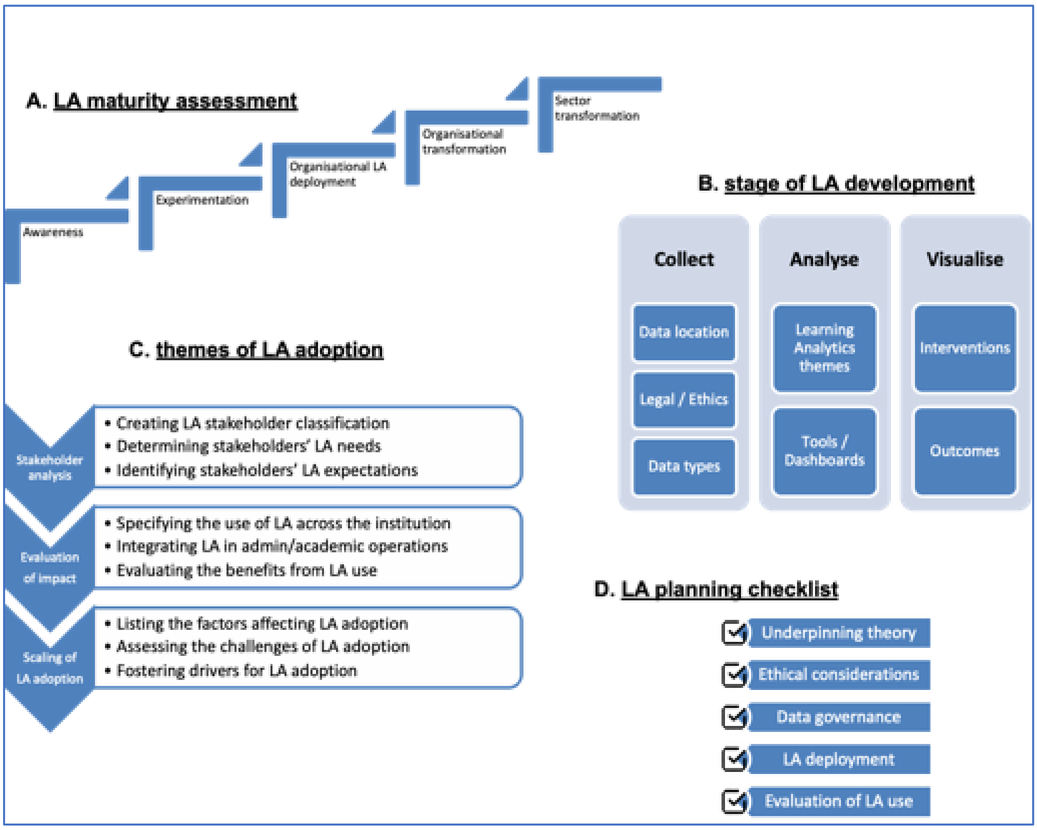

Figure 6-7: Institutional guidelines (strategic level)

This study advocates the need to assess institutional maturity before establishing concrete strategic plans, and definitely before attempting to create a written policy that governs that use of data analytics in a HEI. The importance of learning analytics is advocated by JISC, which adopts the micro-analysis of the field conducted by Papamitsiou and Economides (2014) represented as a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis (Sclater et al, 2016).

Some of the key **strengths** of learning analytics in the education sector include the fact that there are large volumes of available educational data in most institutions and the "ability to use powerful, pre-existing algorithms" to mine data. Being able to produce "multiple visualisations for staff and students", provide "increasingly precise models for adaptation and personalisation of learning", as well as having a "growing insight into learning strategies and behaviours", make learning analytics really attractive for the HEI sector.

Before adopting a learning analytics strategy, institutions must be aware of several **weaknesses**, including "the potential misinterpretation of the data" and "a lack of coherence in the sheer variety of data sources". It is also important to reflect in the "lack of significant results from qualitative research" and the fact that the deployment of learning analytics is likely to involve "overly complex systems and information overload".

Universities need to investigate **opportunities** associated with the deployment of learning analytics such as "using open linked data to help increase compatibility across systems". Universities should also focus one exploiting the use of learning analytics for improving self-reflection, self-awareness and learning through intelligent systems. Furthermore, HEIs should be aiming at "feeding of learning analytics results to other systems to help decision making".

It is also necessary for universities to be aware of learning analytics **threats**, with a key concern being ethical and data privacy issues. Further threats include "over-analysis and the lack of generalisability of the results, possibilities for misclassification of patterns, and contradictory findings".

This micro-analysis should be used by HEIs as a reference point when assessing the feasibility of deploying learning analytics initiatives at institutional level. At the same time HEIs should consider their capacity to support a learning analytics architecture including learning records warehouses and dedicated learning analytics processors as shown in figure 2.1 (Sclater et al, 2016).

The institutional guidelines provided by this research study begin with the need to assess the institution's learning analytics maturity. This was based on JISC's notes from the 6th UK Learning Analytics Network event in Newport (Bailey, 2016), where participants reflected on their institution's activity with regards to the Learning Analytics Sophistication adapted for FE from Siemens, (Siemens et al, 2014). As shown in the following figure, institutions can be positioned across two axes with regards to the maturity of their learning analytics development and whether the impact of learning analytics is limited or entirely integrated in the organisation's operations.

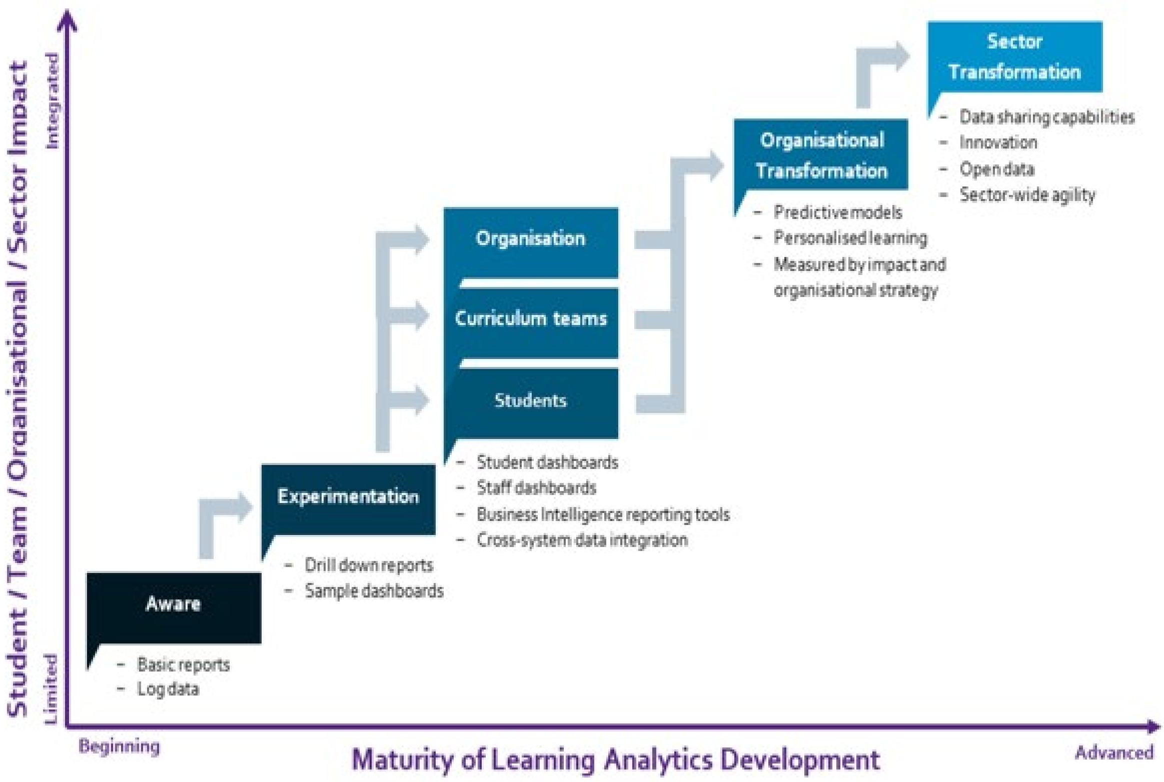

Figure 6-8: Learning analytics sophistication (Bailey, 2016)

As shown in the figure, there are five different maturity stages covering the full spectrum of institutions, from those that are aware of learning analytics, producing basic reports using log data to the ones that are driving sectoral transformation with innovative uses of open data and data sharing capabilities across the organisation. These maturity stages were adopted in the guidelines provided by the research study, as follows:

- Awareness – institutions that use very basic aspects of learning analytics.
- Experimentation – institutions that experiment with simple dashboard use.
- Organisational deployment – institutions that use a wide range of customised dashboards.
- Organisational transformation – institutions that adapt their strategy to exploit predictive models based on learning analytics.
- Sector transformation – institutions that drive innovations in learning analytics.

It is interesting to note that the institutions participating at JISC's learning analytics event were grouped under four categories depending on the focus of their use of analytics, including (i) improving learner performance, (ii) improving institutional performance, (iii) improving learning quality and (iv) defining criteria for institutional strategy.

JIC is driving the sector's developments on learning analytics. JISC conducted a research study resulting in a learning analytics report on the state of play in UK's higher and further education, a number of institutions is (Sclater, 2014). The study was based on a series of structured interviews and provided a guide for the collection of primary data for this research study with the use of focus groups and interviews.

As a result, the second part of the institutional guidelines resulting from this research study suggests a number of stages for learning analytics development. More specifically, institutions should assess their development of learning analytics across seven areas organised under the three core functions of the C.A.V. framework. Under **data collection**, institutions must ensure that they (i) determine the location of all necessary data used for analytics, (ii) have in place the necessary policies and procedures addressing legal aspects of learning analytics and provide an ethical framework for handling data, and (iii) identify the different data types to be used for learning analytics. With regards to **data analysis**, institutions need to (i) determine the learning analytics themes to be used across the university and (ii) select the appropriate tools to be used for learning analytics and decide the contents of the learning analytics dashboards produced. Finally, when considering **data visualisation**, institutions must (i) identify those interventions to institutional activities resulting from the use of learning analytics and (ii) determine the outcomes for different parts of the institution from dashboard findings.

The third part of the institutional guidelines is based on the themes identified by Gasevic et al (2021) from the relevant literature, including (i) stakeholder expectations and needs, (ii) evaluations of impact and (iii) scaling adoption. The research study concluded in the following themes and associated tasks:

- Stakeholder analysis, which involves the creation of the main stakeholder groups, the determination of LA needs for each stakeholder group and the identification of stakeholder expectations for the use of LA in the institution.
- Evaluation of impact, which involves specifications on how LA is used across the institution, the integration of LA in existing operations both academic and administrative, as well as the evaluation of the benefits from using LA across the institution.
- Scaling of LA adoption, which involves the listing of all factors that affect the adoption of LA at institutional level, the assessment of any identified challenges associated with the adoption of LA and fostering drivers towards LA adoption.

The National Forum for the Enhancement of Teaching and Learning in Higher Education provides very useful resources for the use of learning analytics in higher education. Forum notes provide guidelines for students to become aware of the role of LA and how they impact their learning experience. For example, students are informed that “data from their student record may include module registrations, grades, and major”, while some demographic information such as age, gender, and the distance they travel to college may be also collected. LA can be based also on student attendance, and the frequency of using library resources (NFETLHE, 2017). The notes also explain that the data are collected to help the students, while remaining private and safe at all times (NFETLHE, 2017).

The national forum also provides briefing papers with useful insights. For example, the report on students leaving higher education aims “to further inform our understanding of why some students do not progress to the completion of their programmes of study in higher education”, as well as “to determine how best to support students in their transitions into and through higher education” (NFETLHE, 2016a). The briefing paper identified that the main drivers for non-completion are “ill-chosen courses and courses not meeting the expectations of students”. Furthermore, non-completion is also “strongly influenced by the stress associated with factors outside of the course of study such as financial concerns, commuting distance, practical responsibilities or unexpected life events” (NFETLHE, 2016a).

Similarly, the forum insights on students’ experiences of the transition from further education and training (FET) to higher education suggested that “there was an overall sense that the FET experience built the confidence and self-efficacy of students, giving them the skills and self-belief to succeed in higher education” (NFETLHE 2016b). Forum insights also highlight the importance of obtaining student feedback in an effort to improve the quality of teaching and learning (NFETLHE 2013).

Guzman-Valenzuela et al (2021) conducted an extensive review of current work in learning analytics. The authors’ contribution is a clear distinction “between a practice-based and management-oriented community of learning analytics and an academic-oriented community”. The authors conclude that “within both communities, though, it seems that the focus is more on analytics than on learning”. According to Guzman-Valenzuela et al (2021) there are certain critical issues and challenges for learning analytics. These include educational theories, ethical issues, structural factors, research results, data governance, methods and data, as well as issues associated with teachers and students. These are illustrated below.

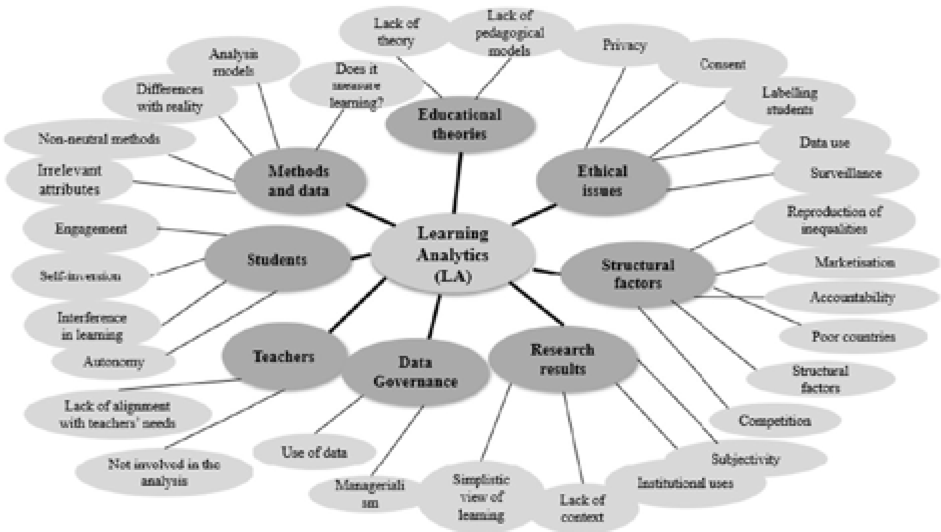

Figure 6-9: Critical issues and challenges for learning analytics (Guzman-Valenzuela et al, 2021)

The critical issues identified by Guzman-Valenzuela et al (2021) were used as the basis for the fourth part of the institutional guidelines provided by this research study. This final part is in the form of a planning checklist for learning analytics. The checklist includes the following elements:

- Underpinning theory – institutions need to establish strong foundations for their policies and shape their own methods, techniques and procedures as part of a theoretical underpinning for their LA strategy.

- Ethical considerations – institutions should ensure their LA practices are in line with current legislation (e.g., GDPR) and a well-documented ethical framework governs the way data is collected and handled.
- Data governance – specific roles such as data controller and members of ethics committee are necessary to ensure that the LA practices adhere to regulations and directives. Any stakeholders associated with LA practices should have clear responsibilities and their role accountability should be well documented.
- LA deployment – the deployment of LA practices should include specific processes, the application of well-defied techniques and the use of appropriate tools (e.g., Tableau software).
- Evaluation of LA use – institutions should have in place the necessary mechanisms for evaluating the impact of LA across the different operations.

The following table shows a proposed approach that consists of three elements (i.e., data, model, and transformation) "designed to ease communication with organisations (adopters of analytics) and assist senior leaders to grasp the benefits and challenges associated with analytics in organizational decision making (Gasevic et al, 2019). As shown in the following table, it is proposed for institutions to have in place certain principles for creative data sourcing, as well as increasing awareness of data limitations, and securing the necessary IT support. The focus is on institutional transformation by developing an institutional LA strategy and associated policies, implementation of LA tools, and developing a decision-making culture based on the use of LA. The institution must have in place appropriate leadership models that value the use of LA and enforce principles for ethical use of LA, while protecting data privacy.

Data	Model	Transformation
- Development of principles for creative data sourcing - Increasing awareness of data limitations - Securing necessary information technology support	- Following question-driven approaches to the applications of machine learning - Informing the use of machine learning by educational research and practice	- Development of Institutional policy and strategy for learning analytics - Establishing effective leadership models to drive and oversee the implementation - Adopting principles for privacy protection and ethical use of analytics - Implementation of learning analytics tools catering the primary stakeholders - Development of analytics-informed decision making culture

Table 6-10: An approach for the systemic adoption of learning analytics in Higher Education Institutions (Gasevic et al, 2019)

The institutional guidelines proposed in this research study in order to ensure the successful use of the C.A.V. framework is in line with the approach suggested by Gasevic et al (2019), aiming to achieve institutional transformation through the use of LA. In conclusion, the institutional guidelines stemming from this research study are as follows:

A. Assessing the institution's maturity in Learning Analytics – aiming to determine the starting point for institutional transformation through the use of LA.

B. Reflecting on the institution’s stage of Learning Analytics development – aiming to provide specific guidelines for data collection, analysis and visualisation.
C. Identifying the main themes driving the adoption of Learning Analytics across the institution – aiming to perform a stakeholder analysis including an assessment of their needs and expectations and the evaluation of LA impact on different stakeholders.
D. Checking the level of completion of the institution’s Learning Analytics plans – aiming to provide a series of check points used for assessing the progress of LA deployment plans.

6.3. User Guidelines (operational level)

The next contribution of this research study includes specific guidelines for the users of learning analytics in a HEI. The user guidelines are organised in five steps, which are illustrated in the following figure. The steps are as follows:

A. Determining added value through Learning Analytics (LA)
B. Assessing the impact of Learning Analytics (LA) on stakeholders
C. Adopting a Learning Analytics (LA) process
D. Planning dashboard contents
E. Identifying factors affecting Learning Analytics (LA) implementation

These guidelines must be followed within the context of institutional policies and initiatives for learning analytics, as there may be different priorities and restrictions associated with the learning analytics requirements of each institution. For example, Ouli and Voutilainen (2019) discuss the processing of student data in universities, emphasising the importance of legislation. More specifically they explain that from a legal perspective, “it is essential to distinguish between the following categories of data based on the personal data protection legislation”: (i) data other than personal data, (ii) ‘ordinary’ personal data, (iii) specific categories of personal data and (iv) unique identification numbers.

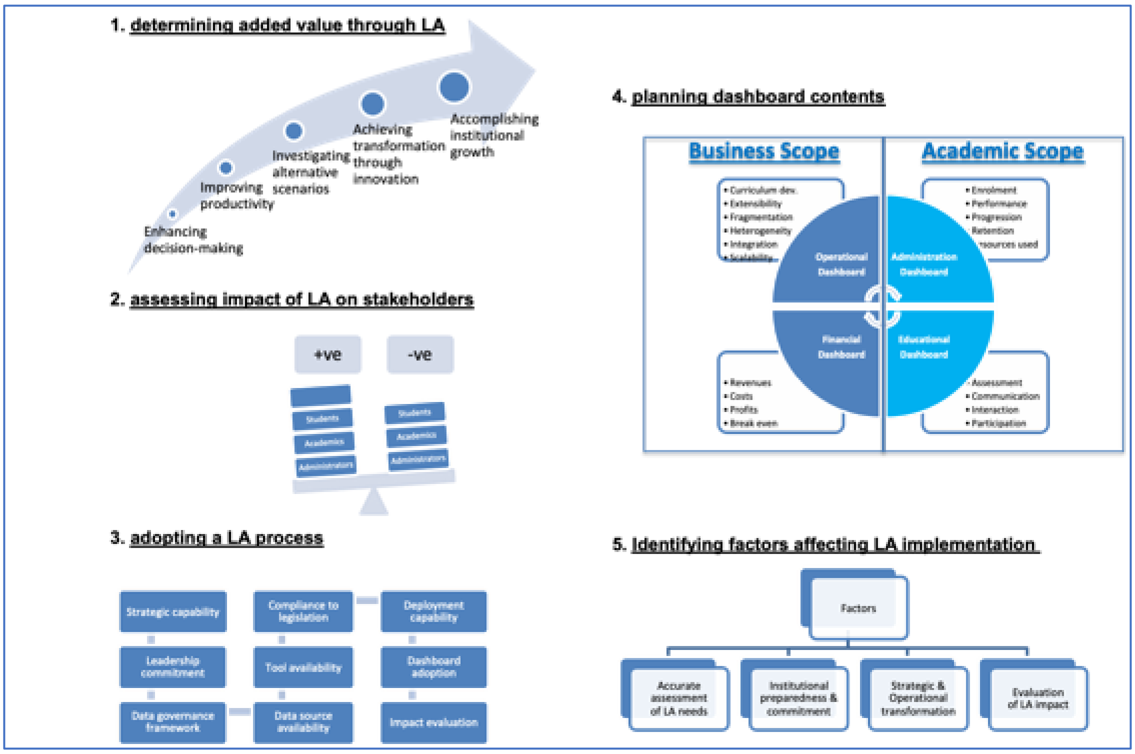

Figure 6-10: User guidelines (operational level)

It is critical to consider the role of EU's General Data Protection Regulation (GDPR) in data analytics at institutional level. In the previous sub-section, the importance of compliance with legislation and governing regulations was emphasised. At operational level, stakeholders must ensure that the way data is collected, handled and analysed is in line with GDPR articles. According to Ouli and Voutilainen (2019) mention article 28 of the GDPR that determines "the contract between a processor and a controller" and article 40 that refers to the codes of conduct that determine the accountability of data controllers. Furthermore, GDPR's article 5 describes the principles that governs the processing of personal data and the concept of legitimate interest that can be used by Universities that need to process students' personal data. Article 22 that refers to automated decision making based on data analytics is also discussed by the authors as important for learning analytics in HEIs (Ouli and Voutilainen, 2019). The proposed operational guidelines consider compliance to legislation as part of the learning analytics process that is adopted by the institution.

Institutions must ensure that their LA stakeholders receive a full awareness course on GDPR and the role of the regulations on learning analytics. The full legal text of the GDPR is available at https://gdpr-info.eu including the full chapter list shown below:

- Chapter 1 – General provisions: articles 1-4.
- Chapter 2 – Principles: articles 5-11.
- Chapter 3 – Rights of the data subject: articles 12-23.
- Chapter 4 – Controller and processor: articles 24-43.
- Chapter 5 – Transfers of personal data to third countries or international organisations: articles 44-50.
- Chapter 6 – Independent supervisory authorities: articles 51-59.

- Chapter 7 – Cooperation and consistency: articles 60-76.
- Chapter 8 – Remedies, liabilities and penalties: articles 77-84.
- Chapter 9 – Provision relating to specific processing situations: articles 85-91.
- Chapter 10 – Delegated acts and implementing acts: articles 92-93.
- Chapter 11 – Final provisions: articles 94-99.

Queiroga et al (2020) argue that “student dropout is considered one of the main problems and has received much attention from the learning analytics research community”. In their paper they provide a typical case study, where learning analytics are used to address such an issue. This is an excellent example of how an institution can put together specific guidelines to address a key issue such as student drop-out. The work of Queiroga et al (2020) a solution “using only students’ interactions with the virtual learning environment and its derivative features for early predict at-risk students”. This research study has collected a wide range of similar case studies, which was discussed in an earlier chapter in order to provide a sufficiently broad set of guidelines to cover all operational aspects of a university.

The following figure illustrates the approach adopted by Queiroga et al (2020) and in particular the use of machine learning algorithms, such as classic Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Logistic Regression (LG), and the meta-algorithm AdaBoost (ADA).

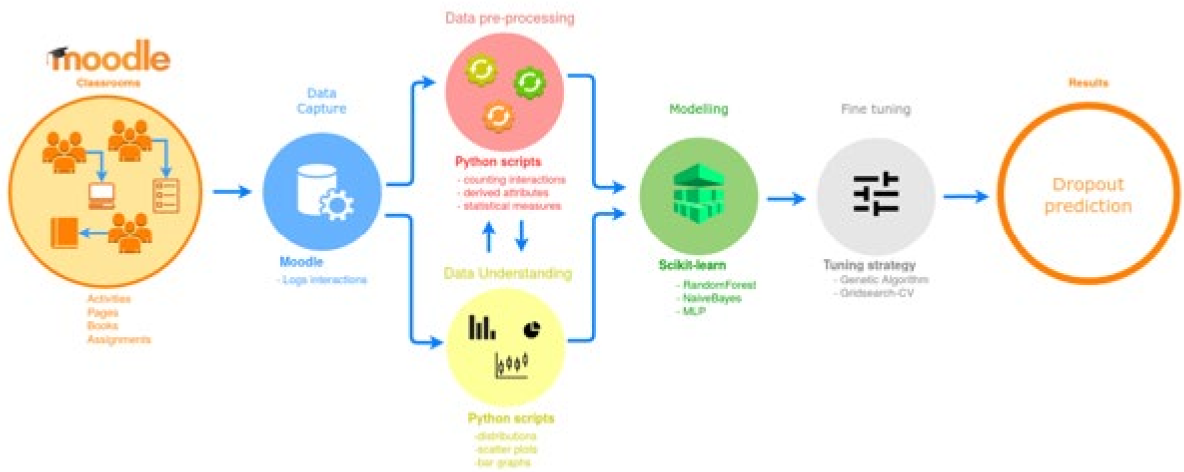

Figure 6-11: A proposed approach to identify students at risk with learning analytics (Queiroga et al, 2020)

The proposed approach collects data from the interactions of students with the Moodle platform, including activities, pages and assignments. The Moodle logs are used to capture the necessary data, which are processed with the use of Python scripts to derive the necessary understanding of student interactions. The different machine learning algorithms are used for data modelling and fine tuning before the drop-out predictions are made. This approach is similar to the one followed in this research study as discussed later in the chapter. The proposed operational guidelines include a step for the adoption of the LA process followed by the design of the dashboard contents. These steps involve a similar approach, where the collection of data from the virtual learning environment are used to create visualisations necessary for decision-making.

However, the adoption of such an approach for creating visualisations of data analytics is only a small part of the operational guidelines needed for HEIs. Patwa et al (2018) study “why and how enormous information will benefit teachers, institutes, online course developers and students as a whole”. The authors refer to the work of Tulasi (2013) that summarises the added value by the exploitation of learning analytics, summarised in seven points such as enhancing decision-making, improving institutional productivity, assisting the handling of complex issues, transforming processes, exploring alternative scenarios, and further developing the institution’s brand. This work helped to form the first step of the operational guidelines, which focus on five areas where learning analytics generate added value for the university:

- Enhancing decision-making – institutions should identify the areas where learning analytics can be used to support decision-makers with in-depth analysis and visualisations.
- Improving productivity – the use of visualisations should be integrated in institutional processes leading to better informed decisions that can be made faster.
- Investigating alternative scenarios – learning analytics should be used to test assumptions and investigate the outcomes of alternative decisions.
- Achieving transformation through innovation – the use of dashboards can help introduce automation in education and come up with new way to manage the learning experience (e.g., sending reminders to students, provide semi-automated feedback, and alert the ones that are at risk).
- Accomplishing institutional growth – using learning analytics should enable institutions to identify areas for future growth and better utilise resources in order to achieve better results.

Patwa et al (2018) also provided a learning analytics model that is based on identifying positive and negative aspects at institutional and faculty level. The authors also discussed advantages and disadvantages associated with the use of learning analytics for students and content developers. The use of descriptive statistics for analysing educational data is illustrated below.

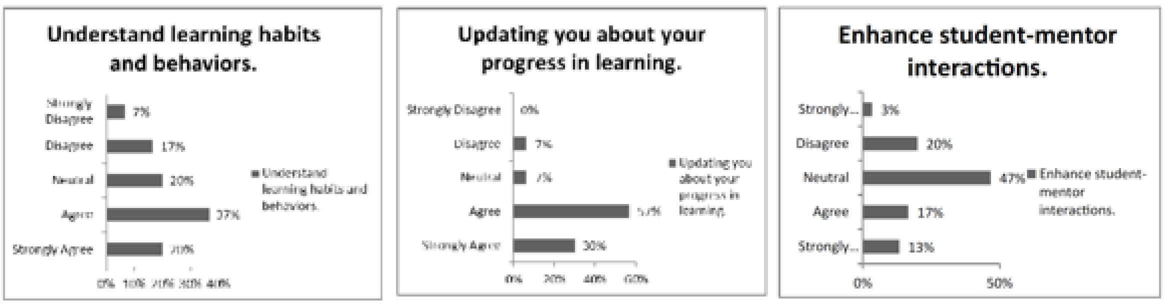

Figure 6-12: Descriptive statistics of educational data (Patwa et al, 2018)

The second step of the operational guidelines provided by this research study focus on the identification of positive and negative aspects of learning analytics for the different stakeholders at institutional level. Institutions should perform a learning analytics needs analysis for their main stakeholders, including students, academics and administrators. Universities may wish to consider how learning analytics are likely to affect content developers and dashboard designers who may be internal staff. The guidelines provided by this research study focus on the impact of learning analytics on three stakeholder types, namely (i) students, (ii) academics and (iii) administrators. For example, students can use learning analytics to check their progress and for self-assessment, while academics can receive feedback on their performance and identify areas for improvement in their modules. Administrators can use analytics for key performance indicators such as progression at programme level and identification of students at risk due to low attendance. It is necessary to keep reflecting on whether other stakeholder groups are in need for learning analytics dashboards. For example, the finance department may wish to have certain dashboards illustrating the viability of programmes, department heads may wish to use dashboards for determining where most of the budget is spent, while research offices could visualise groups, departments or individuals according to the research funding generated.

Viberg et al (2018) discuss an interesting viewpoint of how learning analytics are discussed in the relevant literature. There are so many papers discussing how learning analytics are used to visualise certain aspects of learning. However, there is not much evidence that learning analytics have a significant impact at institutional level. Most publications describe a series of pilots or the way learning analytics are deployed to assist stakeholders by providing visualisations. There is the need for published findings on how learning analytics have achieved a significant impact at institutional level. Ideally, evidence should be published on how learning analytics assist institutions to make improvements at policy or operational level through the exploitation of learning analytics. The authors' suggestions are based on four propositions from the literature, including (Ferguson & Clow, 2017):

- Learning Analytics (LA) improving learning outcomes.
- Learning Analytics (LA) improving learning support and teaching
- Learning Analytics (LA) improving being used widely including deployment at scale.
- Learning Analytics (LA) being used in an ethical way.

The role of learning outcomes as an improvement of learning outcomes is of particular interest to this research study. According to Viberg et al (2018), learning analytics have an impact on **knowledge acquisition**, as students who use learning analytics to assess their performance are likely to work further on improving their results since they are aware of their weaknesses. Furthermore, there is evidence for the impact of learning analytics in **skill development** since learners can determine those skills that they must further enhance. Finally, learning analytics can lead to **cognitive gains**, based on the findings on "how statistical discourse analysis can be used to overcome shortcomings when analysing knowledge processes" (Chiu and Fujita, 2014).

In order to achieve these gains, institutions must integrate data mining and learning analytics in their operations. Romero and Ventura (2020) present the current state of the art in Educational Data Mining (EDM) by reviewing "the main publications, the key milestones, the knowledge discovery cycle, the main educational environments, the specific tools, the free available datasets, the most used methods, the main objectives, and the future trends" in the field. The following figure illustrates how EDM and LA are deployed at institutional level.

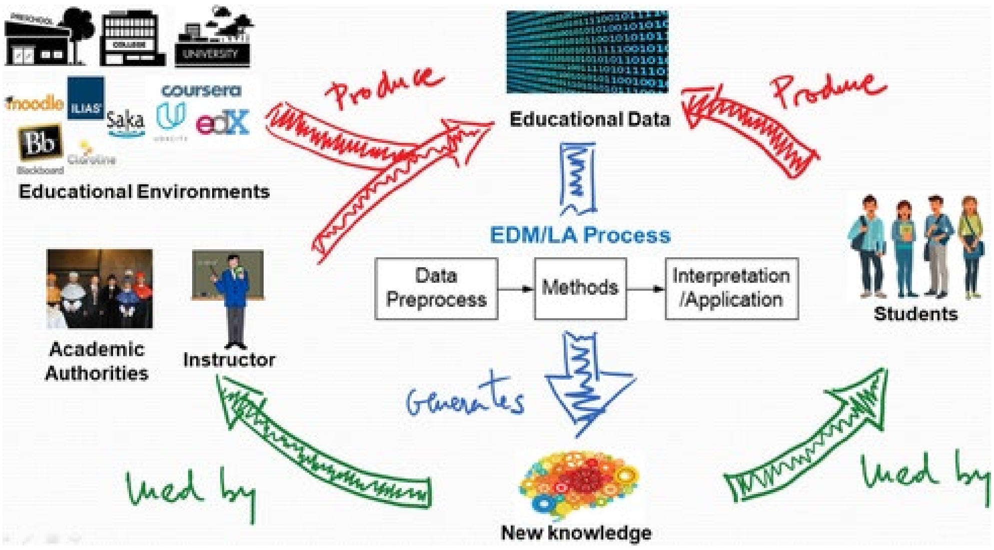

Figure 6-13: Educational Data Mining and Learning Analytics (Romero and Ventura, 2020)

The figure can be used as a reference on how EDM and LA are based on educational data that are produced by educational environments, as well as their stakeholders that may include instructors, academic authorities and learners. The process described in the figure includes the three steps supported by the C.A.V. framework, as it focuses on data pre-processing, which corresponds to the ETL actions on data collected, as well as data analytics. The process also involves the interpretation and application of data on the institution's domain.

Siemens and Baker (2012) present three types of prediction models that are common in EDM and LA, namely (i) classifiers, (ii) regressors, and (iii) latent knowledge estimation. The authors explain that popular classification methods in educational domains include "decision trees, random forest, logistic regression, support vector machines, and increasingly, neural network variants such as recurrent neural networks, long short-term memory networks, and convolutional neural networks". It is important to understand that the method adopted for EDM or LA depends on the capabilities and skillsets of the analysts, but also depends on the analysis that is required by the institution. Siemens and Baker (2012) also explain that inregressors, the predicted variable is a continuous variable, for example a number while in latent knowledge estimation, "a student's knowledge of specific skills and concepts is assessed by their patterns of correctness on those skills".

Joksimovic et al (2019) position learning analytics within "the broader agenda of systems thinking as means of advancing its institutional adoption". The authors discuss how learning analytics are adopted by educational institutions. In particular they consider how learning analytics can be approached as a dynamic system, providing useful insights to a wide range of stakeholders. The following figure illustrates the view of Colvin et al (2016) on how learning analytics are adopted by universities.

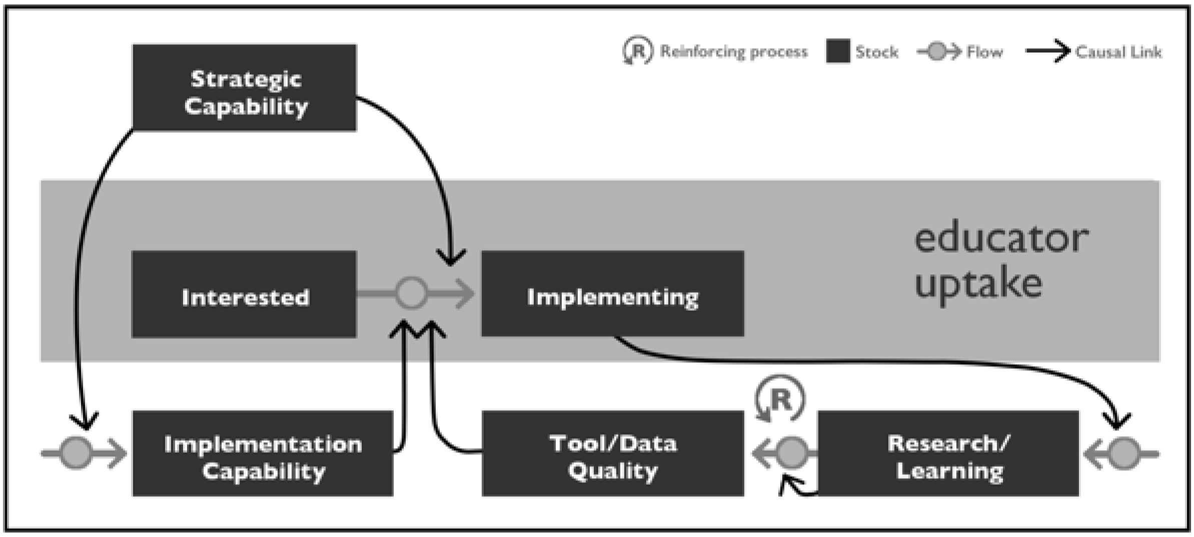

Figure 6-14: University adoption of learning analytics (Colvin et al, 2016))

As shown in the figure, the adoption of learning analytics in a university begins by assessing the strategic capability of the institution, as well as its ability to implement the necessary technology using certain tools and ensuring the quality benchmarks for the necessary data. The initial interest leads to the implementation of the learning analytics platform.

Similarly, the third step of the proposed operational guidelines describe the adoption of learning analytics as a process that consists of the following:

- Assessing the **strategic capability** of the institution to deploy learning analytics.
- Securing the **commitment** of senior management and institutional leadership for the implementation of learning analytics.
- Having in place a framework of **data governance** based on specific responsibilities and accountability associated with different roles (i.e., data controller).
- Ensuring compliance with relevant **legislation** (e.g., GDPR) covering data privacy and ethics.
- Putting in place the necessary technology, such as **tools and platforms** required for learning analytics.
- Providing access to the **required data** for the intended learning analytics.
- Being able to **deploy learning analytics** initiatives to support different operations.
- Integrating the necessary **dashboards** to support institutional operations.
- Evaluating the **impact of dashboards** in supporting decision-making across the institution.

The next step of the learning analytics guidelines includes the planning of the dashboard contents. The original version of the C.A.V. framework distinguished between the different types of data. After conducting the interviews and the focus group session, this part of the C.A.V. framework was revised to include two changes. First, the planning process now distinguishes between a business scope and an academic scope. The former, focuses on how learning analytics support the business aspects of an educational institution (e.g., budgeting, and resource planning). The latter, focuses on the impact of learning analytics on academic aspects such as enhancing the learning experience and improving student progression. The second change was that the C.A.V. framework determines the key characteristics of four different types of dashboards, namely (i) operational, (ii) financial, (iii) administration and (iv) educational.

The planning of each dashboard type requires several elements according to its scope. Ifenthaler and Yau (2020) discuss "a considerable number of learning analytics approaches which utilise effective techniques in supporting study success and students at risk of dropping out". The authors focused on studying success factors that were operationalised as part of learning analytics, the way these factors support study success and how they trigger the necessary learning analytics interventions. According to Ifenthaler and Yau (2020), a wide range of variables has been covered in the literature of operationalised learning analytics in higher education. When attempting to identify attained students it is important to monitor e-portfolio submissions, hits and logins to the assessment resources. When considering the potential for course completion, it is important to consider the number and frequency of posts, as well as the length of the submitted posts. Course completion can be also assessed by using data such as demographics, entry criteria to the course, but also academic ability, financials support, academic goals, technology preparedness, course engagement and motivation, as well as course characteristics.

Learning analytics can be used to assess the pass rate in courses that involve social interactions. In these cases, it is important to analyse social behaviour data, student interactions in forums, as well as attendance and access to online learning materials. It is expected that assessment results are essential for assessing student performance, together with Learning management Systems data. Learning analytics can be also used for improving the design of courses according to a wide range of performance measures such as video use, graded quizzes, access to resources, forum answers posted, and time allocation towards specific study activities.

Prediction of student performance and behaviour is dependent on the analysis of a wide range of data sets such as proportion of sessions attended, total reading time of learning content, login frequency on the virtual learning environment, irregularity of learning intervals, interactions with peers and instructors, usage and satisfaction levels of learning dashboards, learning achievement and level of completion.

The C.A.V. framework was revised in the previous section with detailed data sets identified for the different aspects of the learning analytics involved with institutional operations. In conjunction with the findings from the literature, the four dashboard types have some primary features identified as follows. **Operational dashboards** should be focusing on aspects associated with curriculum development, the ability to extend the dashboard features across different operations, deal with data fragmentation and heterogeneity of data sources, as well as integrating different data sources and support the scalability of analytics to cover bigger data sets and more operations. **Financial dashboards** should focus on analytics associated with the analysis of revenues, costs, profits and break even points at programme level, but also covering the provision of entire departments. **Administration dashboards** should illustrate the results of analysis on student enrolments, cohort performance, as well as overview of progression, retention and use of resources for specific modules or entire programmes. Finally, **educational dashboards** illustrate analysis of student assessment, communication, interaction and participation.

As mentioned earlier in the chapter, the work of Leitner et al (2017) helped to formulate the steps of the analysis process supported by the C.A.V. framework. More specifically, the five steps suggested by Campbell and Oblinger (2007) contributed to the institutional guidelines for learning analytics. Leitner et al (2017) also discussed the importance of data ownership and privacy, and more specifically who owns the data collected and analysed as part of learning analytics. The following categories are critical for proper learning analytics in higher education (Slade and Prinsloo, 2013):

- “The location and interpretation of data”.
- “Informed consent, privacy and the de-identification of data”.
- “The management, classification and storage of data”.

It is also necessary to reflect on the technical aspects associated with learning analytics. This may involve the technology architecture that must be deployed to support the institutional needs for educational data mining. Lauria et al (2017) discuss “the migration of the early detection framework from a single node architecture to a big data, cluster computing architecture using Apache Hadoop and Spark”. This is a proposed solution for a university learning analytics solution for Higher Education Institutions. According to Lauria et al (2017) “a single node architecture can have problems of scale when considering much larger data sets”, while “adding new sources of structured and unstructured data could also lead to an off-scale increase of input data processing that could easily overwhelm the single-node architecture”.

The following figure illustrates the architecture of a learning analytics processor support an early detection framework (Leitner et al, 2017). The figure shows how the processor works with current data (e.g., student academic performance, student demographics, even log data from the learning management system, and gradebook data from assessment records), as well as the ability to extend the system to accommodate future data (e.g., library data, student engagement data and social network data). The outcome of the learning analysis is the identification of those students who are at risk, leading to specific interventions. The cluster computing architecture consists of job scheduling, the hive that deals with ETL tasks, a predictive model using classifiers and a scoring component using different mining techniques such as logistic regression and random forests.

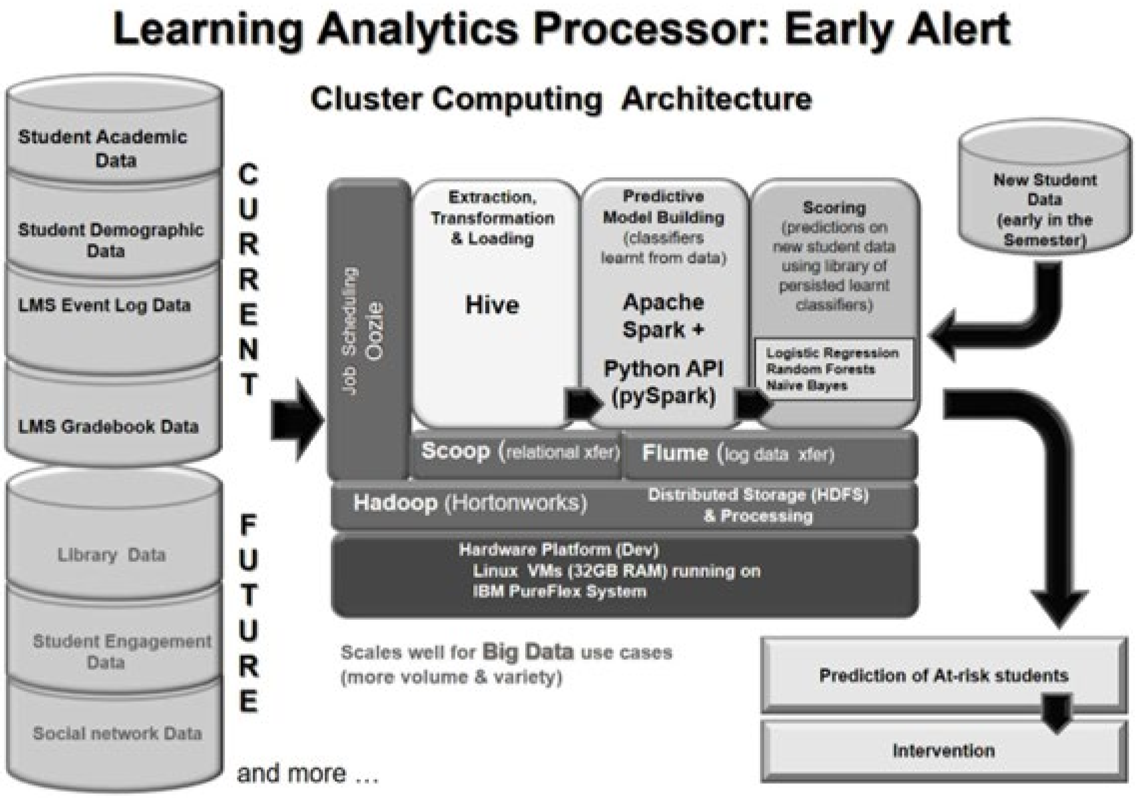

Figure 6-15: Learning analytics processor (Lauria et al, 2017)

Institutions must consider how to deploy similar architectures to support their learning analytics models. Manderveld et al (2017) propose a set-up for a learning analytics architecture and infrastructure. The set-up is based on four steps including (i) determining a number of questions to be used for collecting the necessary data, (ii) creating the necessary functions for collecting the data from the learning record store, (iii) completing the data set and analysing the data using the learning analytics processor and (iv) visualising the results. The following figure illustrates the four layers of the proposed technical architecture.

The four layers are as follows:

- Layer 1 (presentation) – visual presentations of the results of the learning analytics displayed on the dashboard.
- Layer 2 (business) – “the learning analytics processor, which aggregates, organises, analyses and customises data from the learning record store for different users in the presentation layer”.
- Layer 3 (data) – this is the core layer of the entire architecture used “for storing student activities carried out in the various online learning environments used by students.”.
- Layer 4 (input) – this layer involves the connection of various environments to the learning record store.

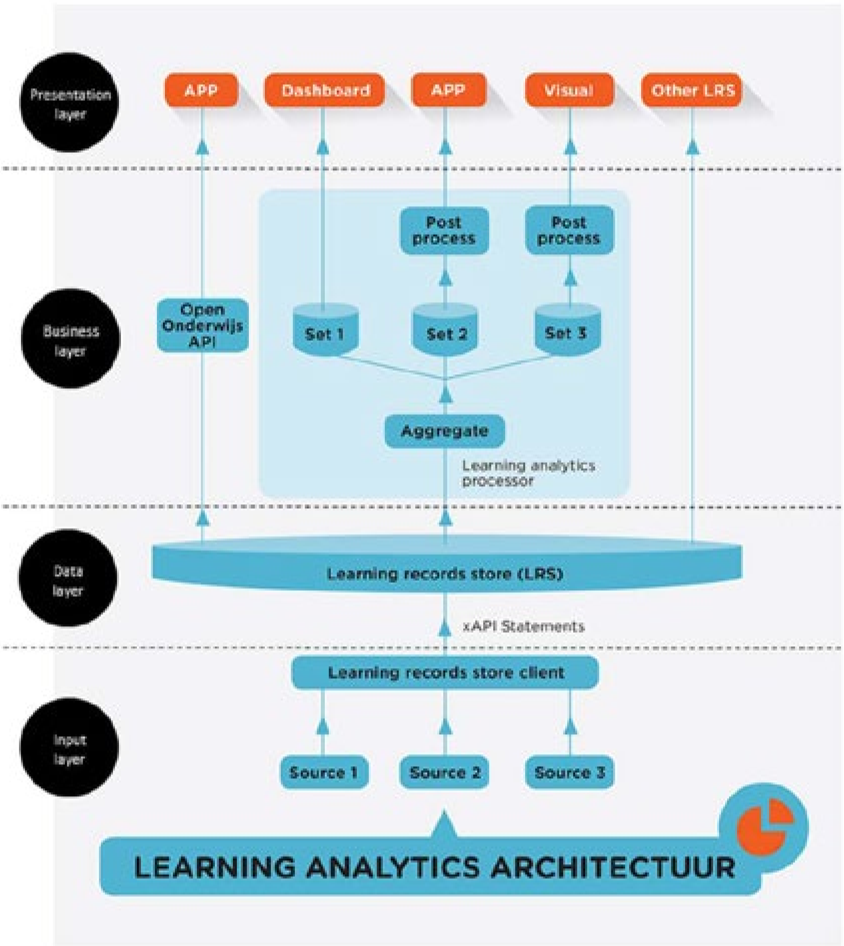

Figure 6-16: Learning analytics technical architecture (Manderveld et al, 2017)

Elias (2017) explain how they "integrated current learning theory, work-related performance data, and learning/ training data from customer care centres to create a data-supported, scenario-based curriculum map". The learning analytics model shown in the following figure, represents the different components of an applied learning design (Elias, 2017). At the centre of the model there are four key elements of a learning design including the different focus of various organisation stakeholders, and the use of technology in various systems, such as Learning Management Systems (LMS) and Customer Relationship Management (CRM). At the core of the model there is also emphasis on the learning analytics needs of different dashboard users (i.e., people) and underpinning theories used to contextualise the use of learning analytics. The learning analytics model follows a path of four processes, namely (i) data gathering, (ii) information processing, (iii) knowledge application and (iv) sharing. These processes are covered by the C.A.V. framework, while the core elements of the model proposed by Elias (2017) can be applied according to the different dashboards designed for an institution.

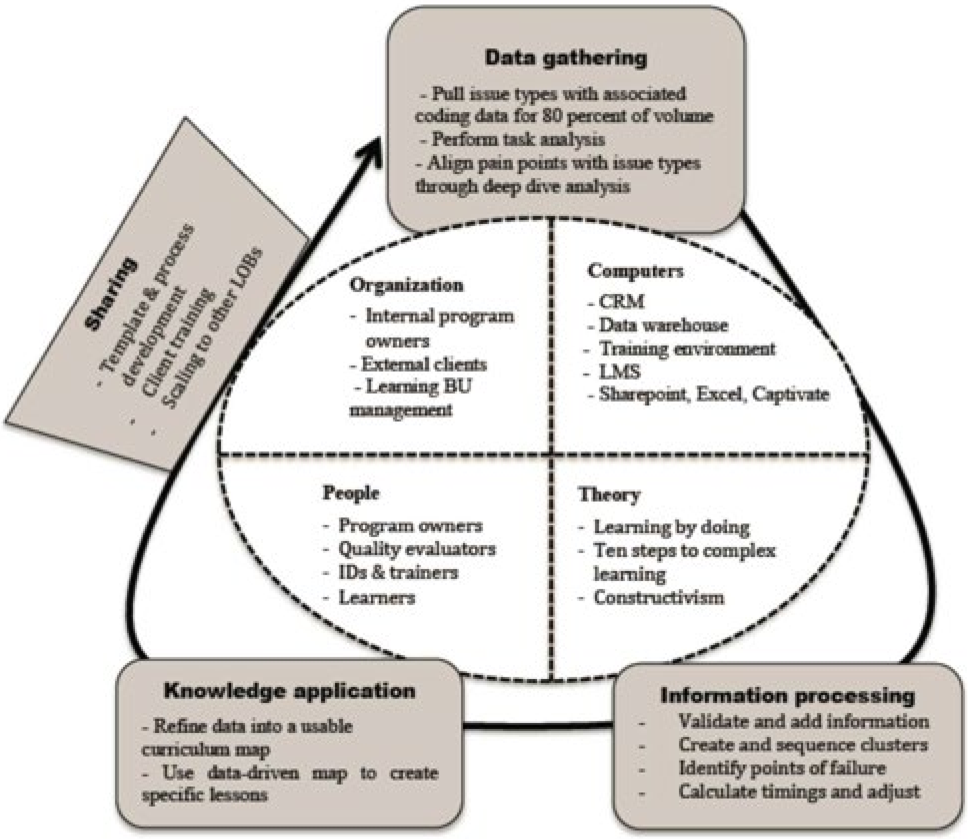

Figure 6-17: Learning an analytics model (Elias, 2017)

Following the discussion of the various models from the literature that affected the formation of the operational guidelines of the C.A.V. framework, the final step of the process focuses on the factors affecting the implementation of learning analytics. Tynan Buckingham Shum (2013) discusses how student success at the Open University is driven by a number of factors such as entry characteristics, academic compatibility, social and academic integration, as well as external factors. The strategic roadmap for Open University focuses on developing institutional capabilities and strengths across ten key areas. These areas are illustrated in the following figure, and are grouped under data availability, analysis and creation of insight, as well as processes that impact student success (Tynan Buckingham Shum, 2013). Availability of data include (i) technology architecture, (ii) data storage and access for analysis, and (iii) data collection. With regards to analysis and creation of insights key areas include (i) data exploration and rapid prototyping, (ii) operational analysis models, and (iii) interpreting results. The areas classified under processes that impact student success, include (i) direct intervention, (ii) information advice and guidance, (iii) continual quality enhancement and (iv) learning design and delivery methods.

These ten elements affect the implementation of learning analytics and can be considered as internal factors that may affect the way an institution will deploy learning analytic dashboards.

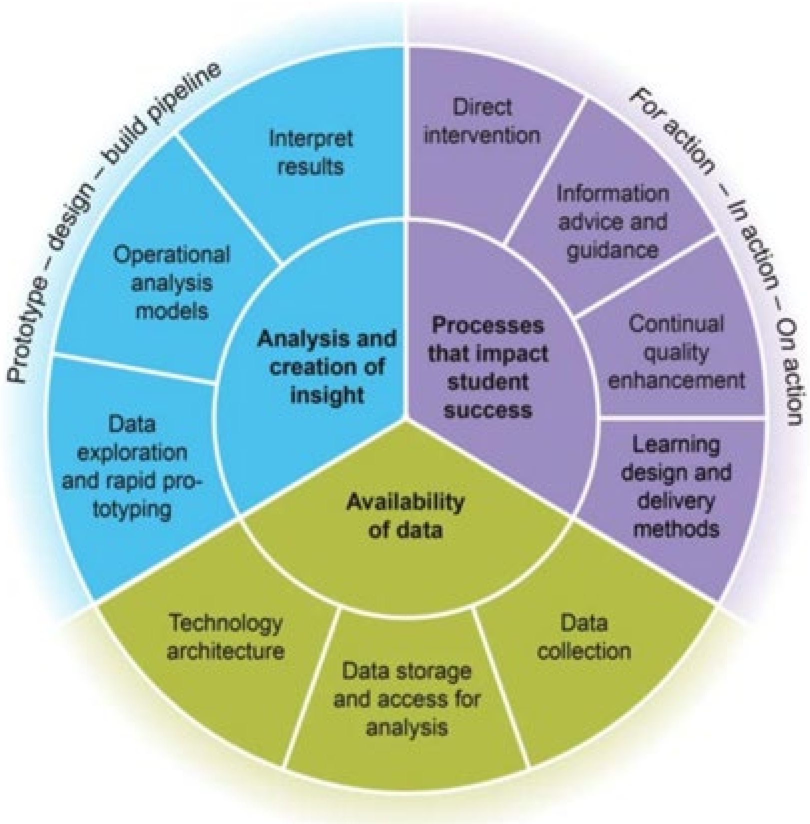

Figure 6-18: Underpinnings of the OU Strategic Analytics Model (Tynan & Buckingham Shum, 2013).

Adejo and Connolly (2017) investigate the roadmap for successful implementation of LA in higher Educational Institutions. The authors discuss that during the roadmap development phase that "involves designing and development of the key strategies to meet the specific requirement of LA development" it is important to consider a number of factors including technological, organisational and human-environmental ones. According to Adejo and Connolly (2017), organisational factors are further decomposed to the ones illustrated below.

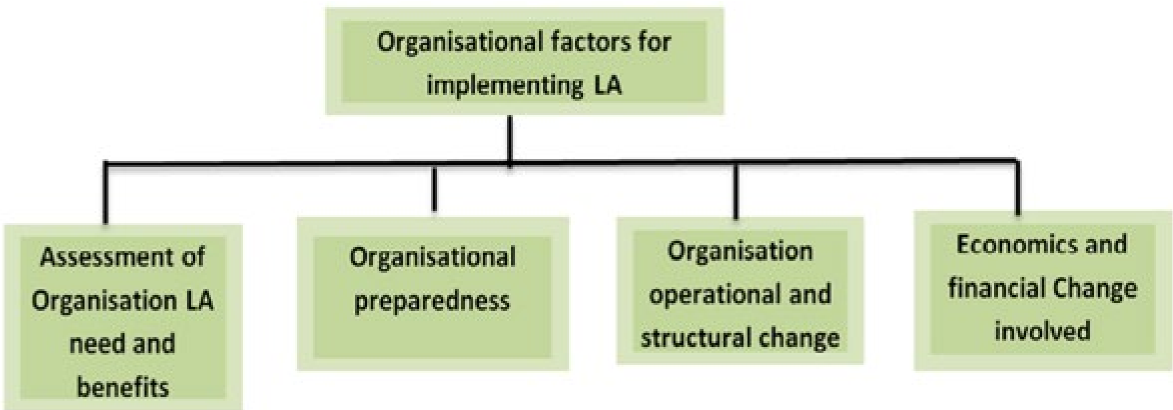

Figure 6-19: Graphical representation of organisational factors for LA implementation Adejo and Connolly (2017)

Organisational factors include the assessment of organisational needs for learning analytics, the organisation's preparedness for learning analytics deployment, as well as issues associated with organisational change. Institutions must consider how the implementation of learning analytics may be affected by their ability to change their infrastructure and operations, as well as integrate any learning analytics implementations to other technologies used across the organisation. It is also necessary to determine whether sufficient support for the implementation of learning analytics is secured, and that appropriate controlling and monitoring mechanisms are in place. These factors are included in the final step of the guidelines provided by this research study, identifying the factors affecting the implementation of learning analytics.

The factors that may affect the implementation of learning analytics in a higher education institution can be classified as follows:

- Ability to provide an accurate assessment of learning analytics needs for the institution.
- Demonstration of institutional preparedness and commitment towards the implementation of learning analytics.
- Capability to achieve strategic and operational transformation.
- Integration of evaluation practices in order to measure the impact of learning analytics across the institution.

In conclusion, the guidelines intended for users of institutional learning analytics in Higher Education are as follows:

1. Determining added value through Learning Analytics (LA) – aiming at institutional areas that are significantly improved after deploying learning analytics.
2. Assessing the impact of Learning Analytics (LA) on stakeholders – aiming at assessing how each stakeholder group benefits from the use of learning analytics.
3. Adopting a Learning Analytics (LA) process – aiming at introducing a learning analytics process that can be used by different university departments without the need for drastic changes.
4. Planning dashboard contents – aiming at identifying dashboard designs using certain components and including specific visualisations.
5. Identifying factors affecting Learning Analytics (LA) implementation – aiming at identifying internal and external factors that affect the ability of institutions to use dashboards effectively.

Examples of dashboards used in Computer Science Department

The following figures illustrate dashboard examples used in Computer Science modules, supporting instructors and students during the learning process. The three dashboards provide findings on student attendance, submissions and assessment combined with engagement.

The first figure illustrates the trend of students' attendance, out of total number of students (73) per week. Different colour is used to distinguish between attendance of labs and lectures. The linear trend model is computed based on the distinct number of students per week. The minimum and maximum reference lines are indicated. There is a peak point for lectures and labs during the week when presentations were held, and a second peak for lab sessions coinciding with the submission deadline for the report.

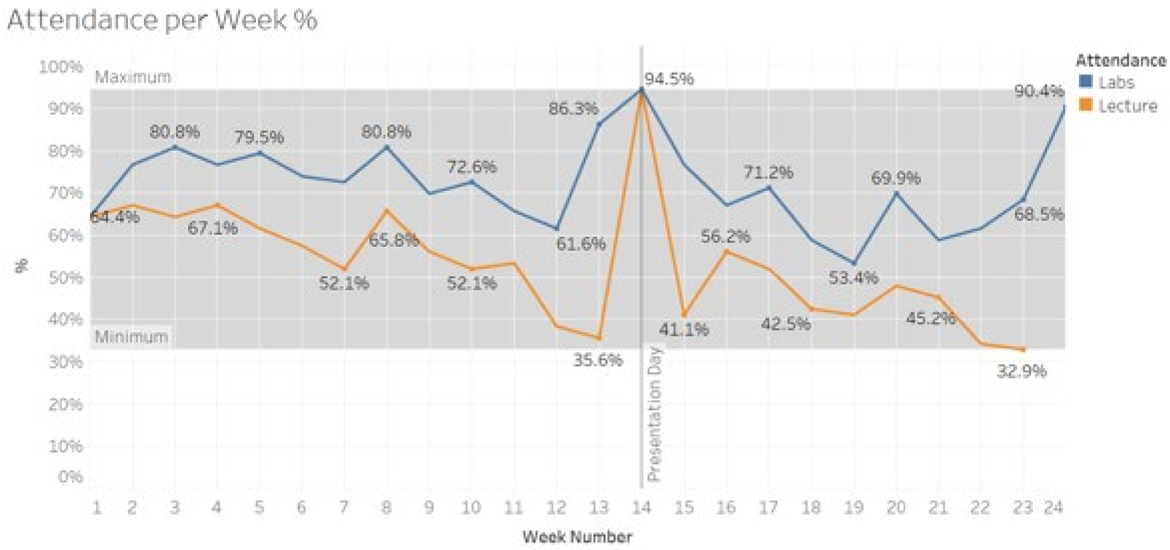

Table 6-11: Attendance dashboard example

The second dashboard illustrates an example for a dashboard showing the summary of GOALs (previously SOBs) observed early, on time or late for different assessment components. The green bars illustrate early submissions, blue lines illustrate submission on the intended weeks, while the red lines illustrate late submissions and therefore delayed observations. Students appeared to fail contributing in social networks, something required as part of the particular module, till after the intended deadline.

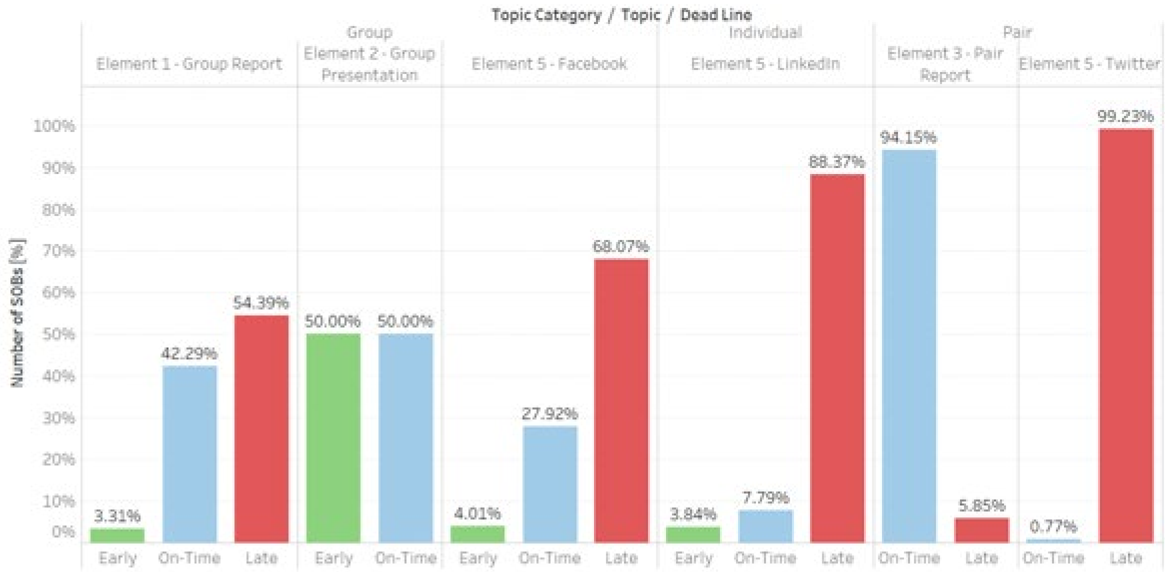

Table 6-12: Student work submission dashboard example

The third dashboard illustrates a bar chart showing the student ranking according to the GOALs observed. The data is filtered using the Group dimension, showing which groups have more group GOALs observed. The pie chart illustrates how many individual and group GOALs are observed for each group member. The highlighted table gives the summary of records broken down by Group, Individual and Report GOAL. Colour codes are used to distinguish between group members but also indicate group members with less GOALs observed, indicating lower contribution and engagement.

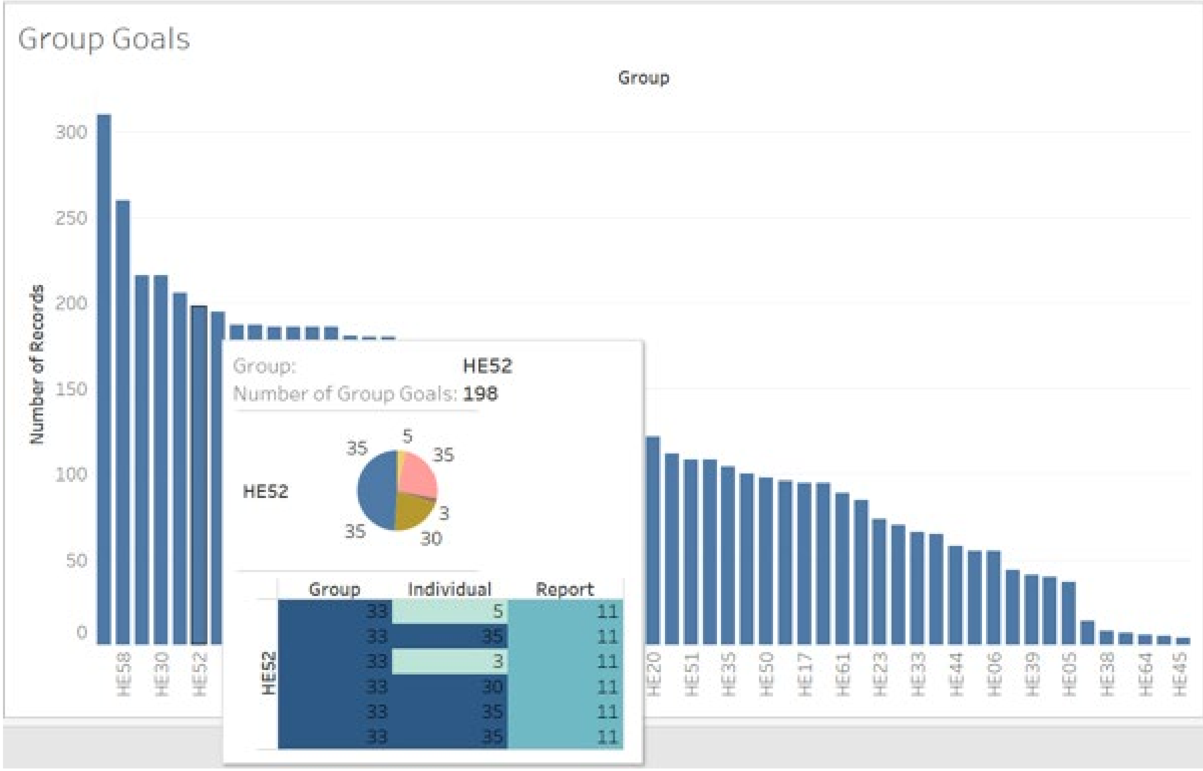

Table 6-13: Student assessment dashboard example

	Group	Individual	Report
HE52	33	5	11
	33	35	11
	33	3	11
	33	30	11
	33	35	11
	33	35	11

The next chapter provides reflections on how this research study relates to different models proposed by similar published works. The chapter also explains the key contributions pf the research study in the field.

6.4. Summary

This chapter discussed the revisions made in the original C.A.V. framework and provided a detailed overview of a wide range of dashboards used at Middlesex University.

Chapter 7 – Reflections on Contribution

This chapter provides a discussion on how this research study contributes in the field. Emphasis is given on how the work was evaluated, resulting in a final version of the proposed framework.

7.1. Research Contribution

Siemens (2012) presents "an integrated and holistic vision for advancing learning analytics as a research discipline and a domain of practices". The author identifies a gap, as "the work of researchers often sits in isolation from that of vendors and of end users or practitioners". Siemens explains that this gap is challenging as "it reflects a broken cycle of communication and interaction between empirical research and how those findings are translated into practice".

The key message from Siemens (2012) is that three areas of development are needed to "drive the adoption of analytics in education":

- New tools and techniques.
- The practitioner experiences.
- The development of analytics researchers.

This is a key contribution for this research study, as it is argued that there is a need for a framework that provides guidance on how to put together a learning analytics strategy at institutional level, and complete it with specific guidelines for integrating learning analytics in HEI operations.

This section discusses the various dashboards provided at Middlesex University that are accessible by staff members. Each dashboard is briefly described, with emphasis on how intended stakeholders can find the necessary information summarised by the available institutional data.

As part of a number of teaching initiatives the author of the thesis has conducted The Information Lab (https://www.theinformationlab.co.uk) in the past, as it provided training on Tableau for staff, researchers and selected students in the Business Information Systems programme. The firm has published a case study with a customer success story based on Middlesex University. The report emphasises how the university sees "a huge improvement in the delivery and clarity of data" by using Tableau. The report explains how the university's planning department "responsible for supporting decision making across the University, found that sharing reports internally was inefficient and time consuming". The improvements from deploying Tableau included (The Information Lab, 2015):

- "Ease of access to data".
- "Accessibility of reports amongst team".
- "Clarity and presentation of results".
- "Improved communication amongst University staff".

The university currently has grouped learning analytics reports under a 'Management Information' category, including a wide range of dashboards, as illustrated in the following figure. Some dashboard categories include several categories, while others are still empty. There are dashboards intended for certain users and have restricted access, while there is capacity for creating ad hoc dashboards. The following figure illustrates the main navigation screen used for accessing these dashboards.

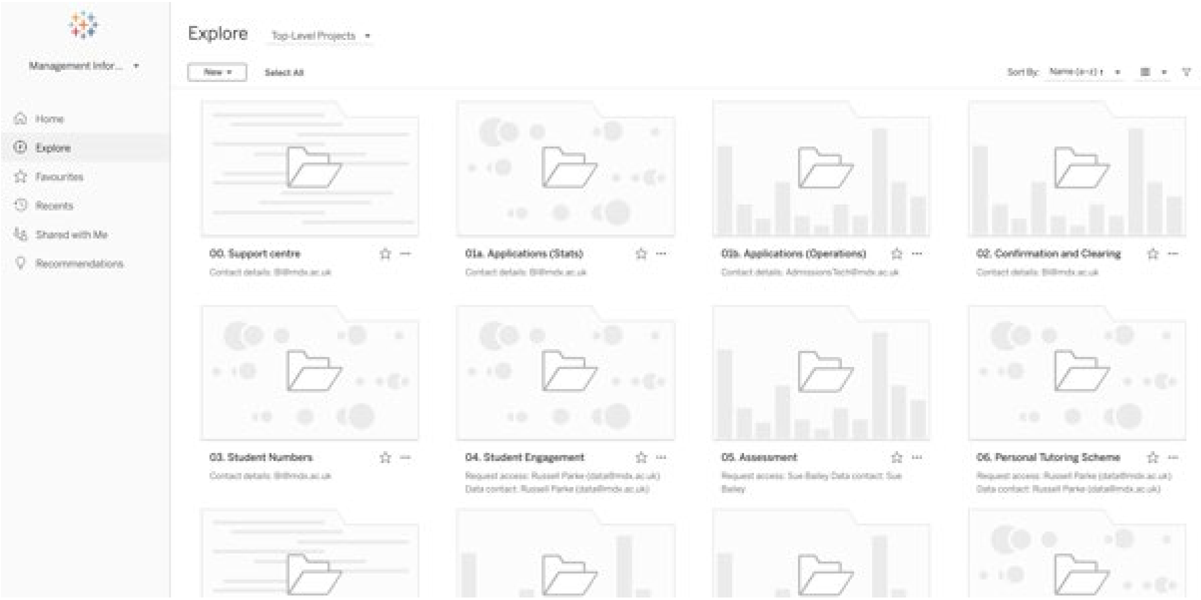

Figure 7-1: Middlesex University management information dashboards

The list of dashboards that had active content in January 2022 is provided below. Each learning analytics section is briefly discussed and its dashboards are illustrated. There was no illustration including readable contents to avoid exposing institutional information, but also to avoid unintentionally displaying personal information for any individuals. The list of dashboards includes:

1. Applications – showing statistics about applications to university programmes.
2. Student numbers – showing student enrolments, and profile details.
3. Student engagement – showing attendance, and participation in certain surveys.
4. Assessment – showing pass rates, assessment statistics and short extension requests.
5. Personal tutoring scheme – showing personal tutor allocations.
6. Progression and awards – showing progression results by faculty and department.
7. Employability – showing destination of university leaver results.
8. Research and Knowledge Transfer – showing external partners and research income.
9. Student surveys – showing student responses to different surveys such as NSS.
10. KPIs – showing annual monitoring and enhancement results.
11. Marketing – showing applicant engagement, educational liaison and outreach.
12. Visa compliance – showing progression report results for tier-4 international students.
13. Learning Support Services – showing operational reports.
14. Programme administration – showing programme and curriculum planning details.
15. CCSS – showing support tickets logged to CCSS.
16. Timetable – showing teaching timetable details.
17. Research student monitoring – showing status of research students.

18. Apprenticeships – showing number of apprentices and associated funding.
19. HR – showing staff turnover, sick days, etc.
20. Student Fees and Finance – showing payments, instalments and sponsorship status.
21. Student records – showing distribution of student credits.
22. External returns – showing aggregate offshore student returns.

MDX Management Information Dashboard – Applications

The applications dashboards (illustrated in the following figure) include (i) acceptances and declines, (ii) application turnaround time, (iii) applications activity by day, (iv) other application statistics.

Application processing status analysis includes a 120-day snapshot of applications including those applications pending tutor decisions, or awaiting other action. The application dashboard also provides an analysis for the median turnaround time for each faculty, as well as processing time by month of application. The analysis also includes the classification of admission decisions by day.

Figure 7-2: Application

MDX Management Information Dashboard – Student numbers

The student-numbers dashboards (illustrated in the following figure) include (i) student enrolment, (ii) student profiles, (iii) commuter students, (iv) widening participation student details.

These dashboards include a wide range of information such as student numbers classified as young (under 21) or mature (over 21) and the corresponding fee income. There is a breakdown of student enrolments per faculty, department, cluster and year of study. Additional information shown includes the map of term-time address, travel times and accommodation type, as well as students who are eligible to enrol or have interrupted in different programmes.

Figure 7-3: Student numbers

<u>MDX Management Information Dashboard – Student engagement</u>

The student engagement dashboards (illustrated in the following figure) include (i) attendance and engagement reports, (ii) pre-arrival surveys, and (iii) student support callers.

Attendance reports show students who did not participate in their course and how many of them managed to pass their modules. Student engagement is shown as attendance percentages for all modules of the programme. Student profile reports include details about the sessions attended, their participation on my Learning (i.e., Moodle) as well as logins to the MDX App, library loans, MDX App tile clicks, card swipes and the use of e-resources. There are also records for students who stated they had issues with online access or money issues, as well as the ones who stated they had access to a laptop, Wi-Fi and data access.

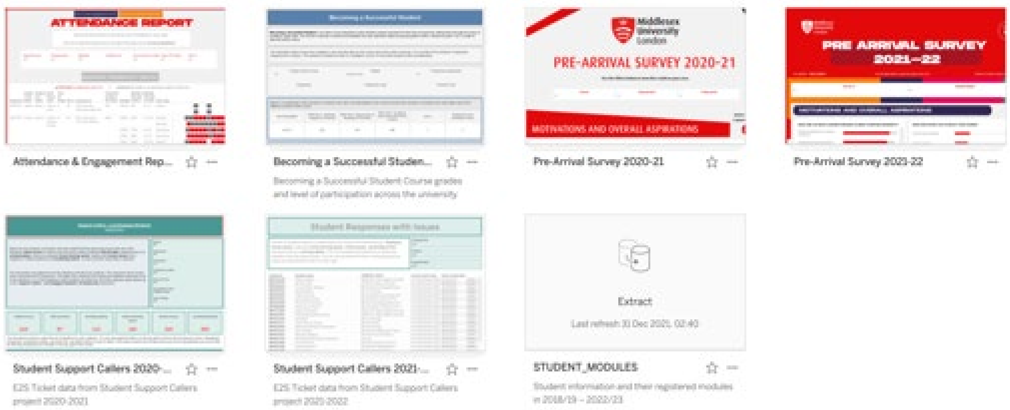

Figure 7-4: Student engagement

<u>MDX Management Information Dashboard – Assessment</u>

The assessment dashboards (illustrated in the following figure) include (i) student pass rates, (ii) assessment data and (iii) short extension report.

Assessment pass rates are calculated by faculty, department, cluster, year of the programme, including percentages for students who passed everything, failed one module or failed all modules. Students are filtered with respect to the grades they achieved, while a report is provided for students who requested short extensions for one or more of their modules.

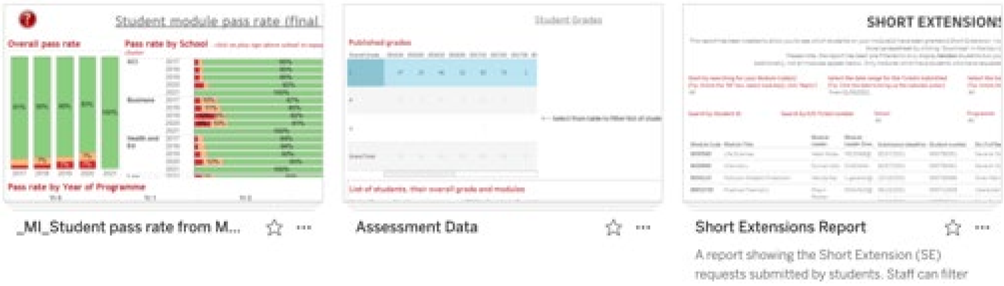

Figure 7-5: Assessment

<u>MDX Management Information Dashboard – Personal tutoring scheme</u>

The personal tutoring dashboards (illustrated in the following figure) include (i) contacting tutees and (ii) managing personal tutor allocations.

The full list of tutees for each member of staff is provided, including details such as faculty, department, programme of study, year of study, and enrolment status. This information is also filtered by programme leader showing the number of tutees per academic member of staff. Personal tutor allocations breakdown shows how many students from each year of study, each personal tutor has.

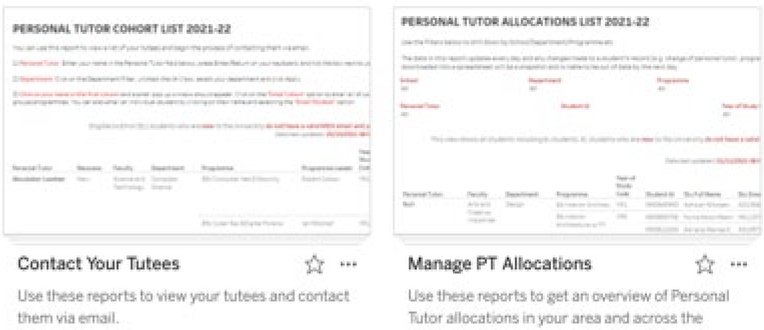

Figure 7-6: Personal tutoring scheme

<u>MDX Management Information Dashboard – Progression and awards</u>

The progression and awards dashboards (illustrated in the following figure) include progression and achievement. The progression percentages of the first year are monitored over the past four years. Achievement is recorded as the percentages of each cohort that achieve different grade classification, as well as the percentages of students achieving good awards.

Figure 7-7: Progression and awards

<u>MDX Management Information Dashboard – Employability</u>

The employability dashboards (illustrated in the following figure) include (i) general information about the Destination of Leavers of Higher Education (DLHE), (ii) DLHE measure diagrams, (iii) DLHE performance indicators, (iv) DLHE results, (v) employed graduates, and (vi) employment trends from graduation.

Employability dashboards include information about DLHE activity values, such as working full time or part time. Due to start work in the next month, engaged in full or part time study, as well as taking time out to travel. Doing something else or being unemployed. Information provided includes percentages for each type of activity mentioned already, as well as average and median salaries. Employability performance indicators are used to rank different faculties and departments, as well as programme. DLHE data also include the number of students occupied with specific roles, such as managers, professional occupations, associate professional and technical occupations, etc. Gender, ethnicity and age filters are applied to view graduates in part- and full-time mode. Sector data also include figures for Middlesex comparators, different regions, as well as the region.

Figure 7-8: Employability

<u>MDX Management Information Dashboard – Research and Knowledge Transfer</u>

The research and knowledge transfer dashboards (illustrated in the following figure) include (i) external partners for research and (ii) research and knowledge transfer income.

RKT uses an interactive map showing countries with which staff have different forms of collaboration, such as research project partnerships and consultancy projects or externally funded projects. An additional dashboard determines the funding received for a wide range of activities, including grants and externally funded research, collaborative research, contract research, training, conferences and various forms of Continuous Professional Development (CPD).

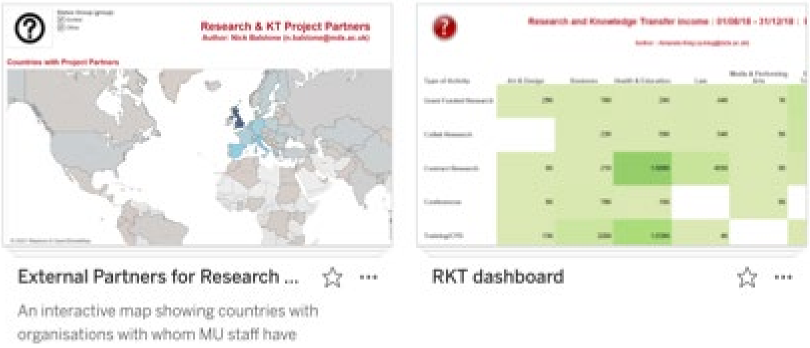

Figure 7-9: Research and knowledge transfer

MDX Management Information Dashboard – Student surveys

The student survey dashboards (illustrated in the following figure) include (i) graduation survey results, (ii) module evaluations, (iii) NSS results, (iv) programme satisfaction survey, (v) postgraduate taught programme experience survey, (vi) postgraduate research programme experience survey, (vii) student experience survey, and (viii) welcome and programme induction survey.

The different dashboards provide overviews of the survey results, enabling stakeholders to assess the views of students. In particular, the NSS survey is supported by a number of dashboards focusing on the sector results, survey demographics, results by programme and department, as well as the summary of the survey results.

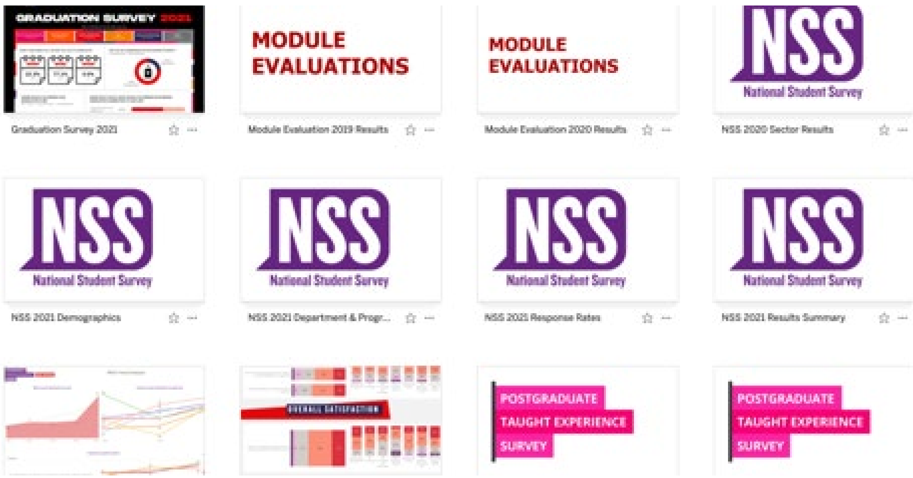

Figure 7-10: Student surveys

MDX Management Information Dashboard – Key Performance Indicators (KPIs)

The Key Performance Indicators (KPIs) dashboards (illustrated in the following figure) include (i) non-continuation, (ii) clarification for each KPI and (iii) recruitment continuation of undergraduate and postgraduate programmes.

The dashboards provide statistics comparing universities and colleges against different criteria such as non-continuation and benchmark trends. Only new students are analysed for cohorts covering the foundation year and the first year of the undergraduate programmes. The analysis is filtered at programme level but also for different departments.

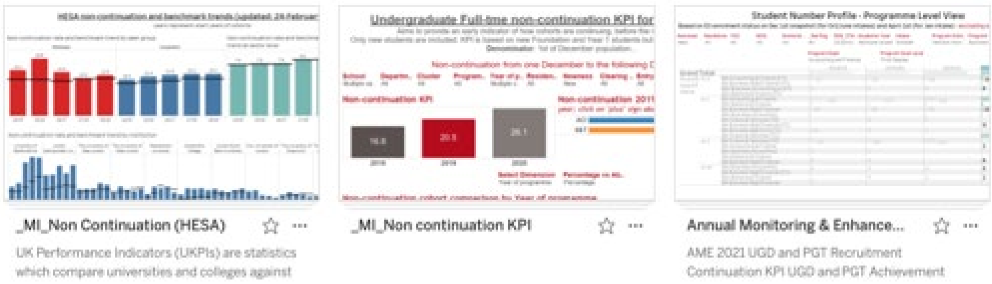

Figure 7-11: Key Performance Indicators

MDX Management Information Dashboard – Marketing

The marketing dashboards (illustrated in the following figure) include (i) applicant engagement, (ii) education liaison and outreach report, (iii) workforce trend data, and (iv) academic partnership report.

The applicant engagement dashboard is currently empty, but the education liaison and the outreach activity report presents the number of activities, engaged students and leads capture over the past few months, as well as an overview of events with details such as date, location, students attended and leads generated. The workforce data includes details about the level of study for staff under different categories, such as managers, directors and senior officials, professional occupations, associate professional and technical, administrative and secretarial, skilled trades occupations, caring, leisure and other services, sales and customer service, process, plant and machine operatives and elementary occupations. The academic partnership report does not provide access to all staff.

Figure 7-12: Marketing

MDX Management Information Dashboard – Visa compliance

The visa compliance dashboard (illustrated in the following figure) includes a progression report. The progression report provides student details, the term of first admission, progression outcomes and progression dates for the student's programme.

Figure 7-13: Visa compliance

MDX Management Information Dashboard – Learning Support Services

The Learning Support Services dashboards (illustrated in the following figure) include (i) LSS management information and (ii) LSS operational reporting.

The LSS dashboards describe the progress of analytical projects in increasing their analytical maturity (for example pastoral care curriculum). There is also a dashboard providing an overview of tickets created, as well as a breakdown of tickets by category (e.g., online UniHelp tickets, tuition fees, extenuating circumstances enquiries and placement or internship enquiries).

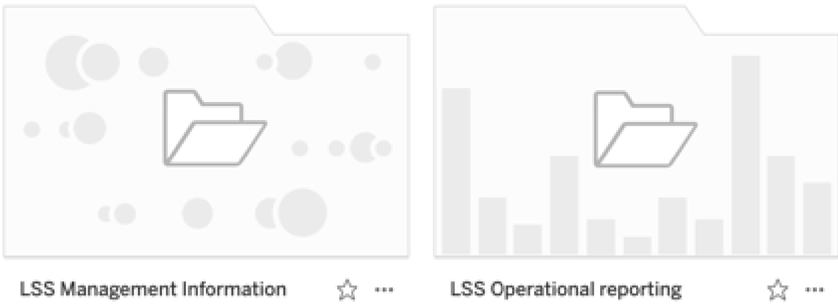

Figure 7-14: Learning Support Services

MDX Management Information Dashboard – Programme administration

The programme administration dashboards (illustrated in the following figure) include (i) CMI claims, (ii) CMI registrations, (iii) programme and curriculum planning, (iv) programme and module leader reports, (v) programme hierarchy, (vi) programme listing and (vii) welcome events. The reports provide information about students registered in programmes and modules.

The CMI registration assistant provides a full record of student information. Programme and curriculum planning dashboards include all necessary information about programme delivery following the impact of the COVID19 pandemic on the delivery of the university programmes.

There is a detailed report on all modules included in academic programmes, the mapping of university modules to those of collaborative partners and a list of the status and level of each programme. There is also a full listing of academic programmes, and detailed schedules for welcome events for each programme.

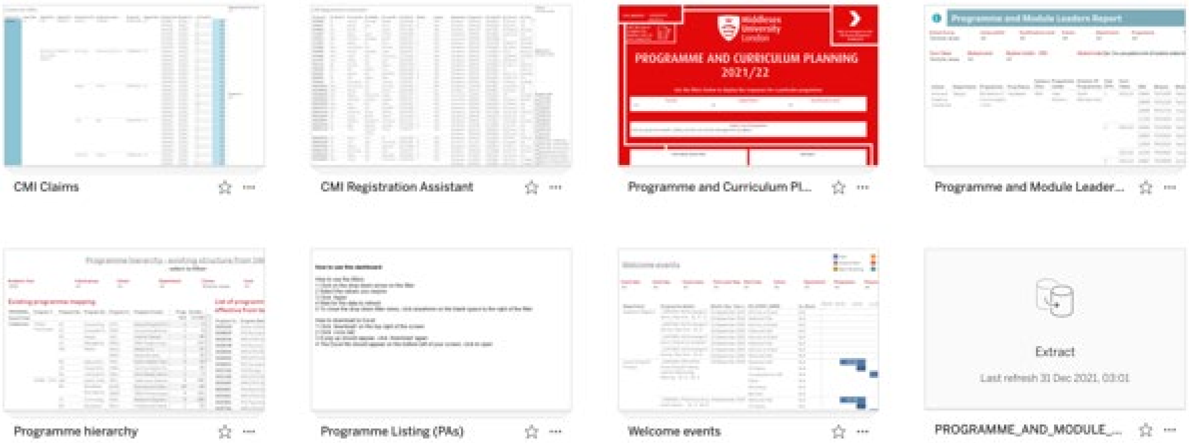

Figure 7-15: Programme administration

<u>MDX Management Information Dashboard – CCSS</u>

The CCSS dashboards (illustrated in the following figure) include the data source log report on CCSS helpdesk tickets. The dashboard is not available to all Tableau users.

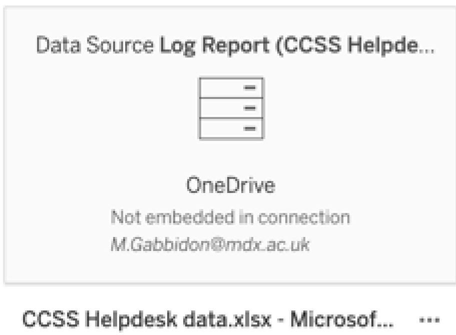

Figure 7-16: CCSS

<u>MDX Management Information Dashboard – Timetable</u>

The timetable dashboards (illustrated in the following figure) include (i) teaching timetable, and (ii) timetable components. The teaching timetable dashboard enables to allocate sessions to each teaching component of the modules of a programme.

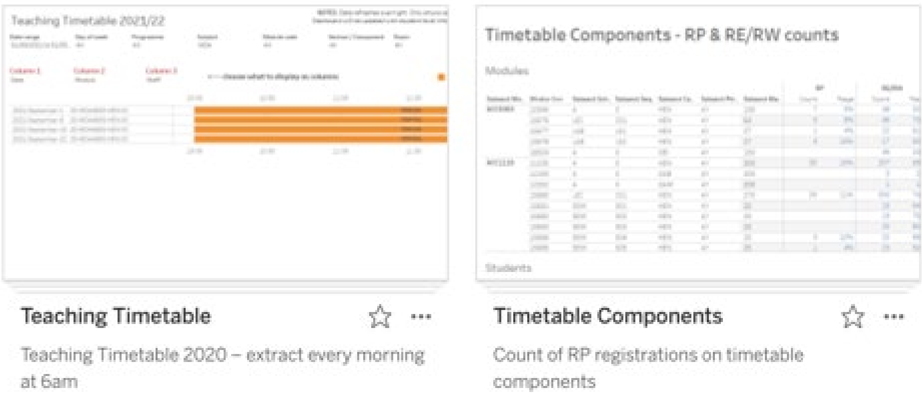

Figure 7-17: Timetable

MDX Management Information Dashboard – Research student monitoring

The research student monitoring dashboards (illustrated in the following figure) include (i) research student milestones, (ii) research student numbers and (iii) research student reports.

All research students are listed with details about their Director of Studies (DoS), and supervisors, as well as expected and actual dates for their registration, transfer and submission. The student overview can be filtered by programme, DoS, supervisor, student status, enrolment, year of study, etc.

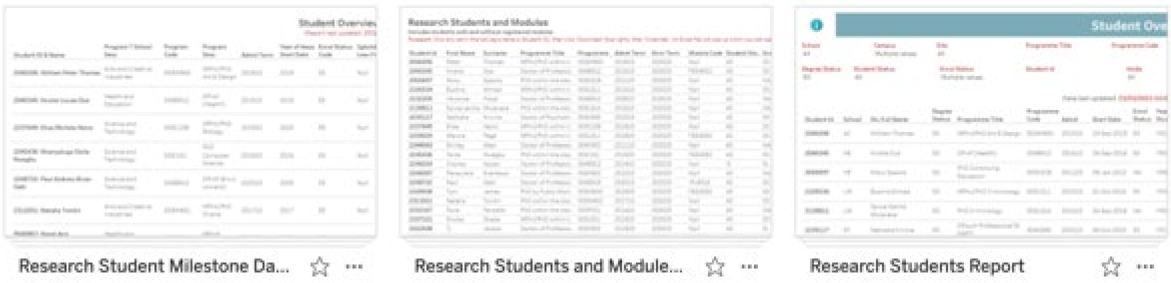

Figure 7-18: Research student monitoring

MDX Management Information Dashboard – Apprenticeships

The apprenticeships dashboards (illustrated in the following figure) include (i) apprenticeships income and (ii) apprenticeships numbers.

The dashboards provide an overview of the apprenticeships funding, classification of funding by faculty and the correlation between apprenticeship funding and learner numbers. There are also dashboards providing the full listings of all apprentices per faculty, including details about employers and their programmes of study.

Figure 7-19: Apprenticeships

<u>MDX Management Information Dashboard – Human Resources</u>

The Human Resources dashboards (illustrated in the following figure) include (i) absence, (ii) equal opportunities, (iii) employee headcount, (iv) recruitment, and (v) turnover.

The absence dashboard provides information about the average number of sick days benchmark for Middlesex University and the public sector at large, as well as average sick days per month, and a breakdown on the sick days per department, organisational unit. The cost of sick days is also provided by month for the current and the previous years. There are also diversity dashboards, illustrating gender demographics and different age bands, as well break downs of ethnicity, religion and sexual orientation. Diversity data are also illustrated in different dashboards for different faculties. There is also a detailed headcount of current core staff by employment type and the full time equivalent for academic, administrative and management staff. HR dashboards also illustrate the number of jobs advertised, as well as the average time to hire in days. The turnover dashboard provides turnover percentages for different employee types, as well as the turnover rates across the university's schools or services.

Figure 7-20: Human Resources

<u>MDX Management Information Dashboard – Student Fees and Finance</u>

The student fees and finance dashboards (illustrated in the following figure) include (i) cash sheets, (ii) deregistration, (iii) monthly pack and (iv) sponsor list.

Student fees and finance analytics illustrate various financial aspects, starting with student payments, classified according to overseas and home students. The analysis also includes those students who have been deregistered from their programmes for different reasons. Additional dashboards provide balances of instalments, and the listing of student sponsors.

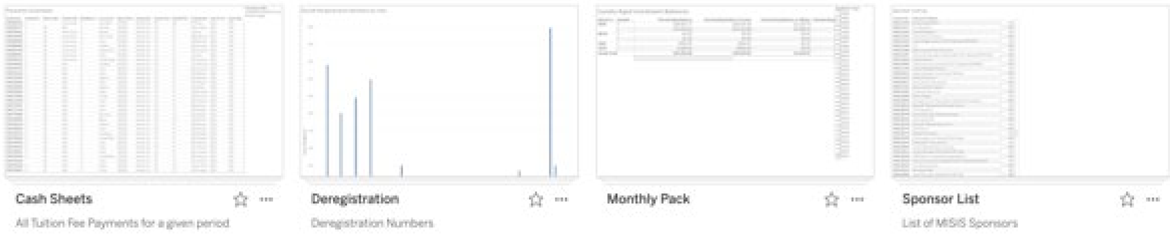

Figure 7-21: Student fees and finance

<u>MDX Management Information Dashboard – Student records</u>

The student records dashboards (illustrated in the following figure) include (i) registrations, (ii) distance education analysis, (iii) e-terms, (iv) site and campus mismatches, and (v) student credit distribution.

The student registrations dashboard provides numbers of students registered per programme and module, as well as summaries for students in each faculty and department. There is a dashboard identifying any location mismatches between in student records. Several dashboards provide details for students registered in distance education mode, with emphasis on their location and programme of study. A separate dashboard provides a detailed checklist for student status timeline.

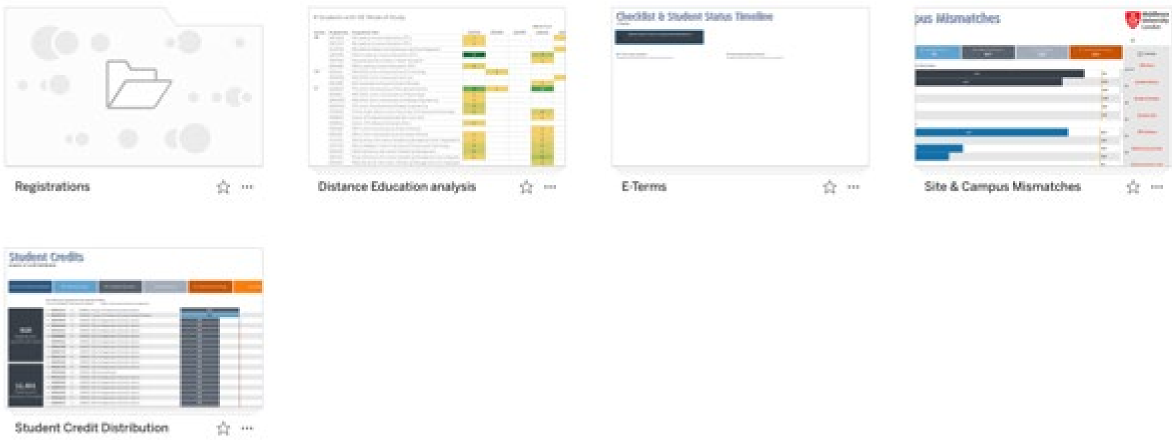

Figure 7-22: Student records

MDX Management Information Dashboard – External returns

The external returns dashboards (illustrated in the following figure) include (i) aggregate offshore students, (ii) student returns and (iii) countries map.

The external returns dashboards show student numbers in overseas campuses, as well as in validated partners. The countries map dashboard illustrates the student numbers from different countries across all programmes.

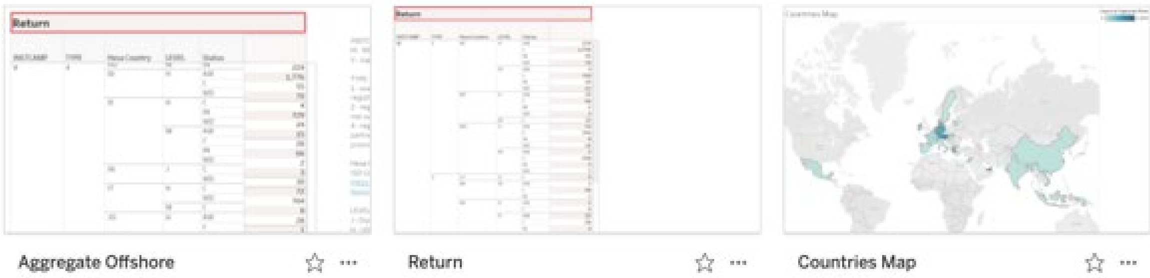

Figure 7-23: External returns

As mentioned earlier, there were no detailed views in this section to ensure the confidentiality of corporate data and the privacy of personal information. Furthermore, there was no access or processing of personal or confidential data as part of this research study.

Following the review of the entire set of dashboards, it became evident that the planning team responds to learning analytics needs from faculties and departments, as well as central services. The Middlesex University case study provides sufficient evidence that there are available resources for generating the necessary dashboards to support specific learning analytics requirements.

There is also available support for the use of these dashboards as part of short training courses offered to staff members. There is clear mentoring, especially between administrators and academics in order to ensure that new members of staff understand the importance of these dashboards for obtaining clarity about the status of certain institutional operations. There are also opportunities for brief Tableau trainings for individuals who are keen to acquire skills for developing simple dashboards.

However, there is no documentation describing the strategic planning of the learning analytics initiatives leading to the creation of dashboards. There also appears to be a gap in the documentation describing or suggesting the way learning analytics are deployed and integrated at operational level. This is the gap that this research study attempts to fill in, and provide the necessary framework for supporting strategy planning, policy making and dashboard creating.

The next section reflects how the research study attempts to fill the identified gap and includes insights from evaluating the C.A.V. framework with stakeholders participating in the focus group and interviews conducted as part of the research study.

7.2. Evaluation of Work

The contribution of this research study is twofold. First, it attempts to contribute to the field of learning analytics by introducing a framework that integrates all aspects of the learning analytics process. Second, it attempts to fill the gap caused by missing guidelines for using learning analytics at both strategic and operational levels for Higher Education Institutions.

The first stage of the evaluation of the C.A.V. framework involved a comparison of the proposed interventions with similar approaches adopted by universities worldwide. The second stage of the evaluation involved key stakeholders participating in the primary data collection of this research study. Following the focus group and interview sessions, the C.A.V. framework was presented to participants in order to receive preliminary views on its usefulness and completeness. Ideally the framework should be evaluated after its application at large scale across the institution or at least a department or cluster of programmes. Unfortunately, this was not deemed feasible during the timeline of this research study, and it would also require the authorisation of senior executives. As discussed earlier, part of the framework was implemented when creating dashboards for a couple of modules taught in the Business Information Systems cluster of programmes in the Computer Science Department.

<u>Reflections from institutional interventions</u>

Jivet et al (2018) "analyse the extent to which theories and models from learning sciences have been integrated into the development of learning dashboards aimed at learners". The authors described different types of data used for the evaluation of dashboards. The different evaluation methods retrieved from their literature review are classified under self-reported, tracked and assessment methods, as shown in the following figure.

Category	Data type
Self-reported	Feedback survey
	Interview
	Focus group
	Evaluation instrument
Tracked	Resource use
	Learning artefacts
	Dashboard use
Assessment	Grades

Figure 7-24: Data used in the evaluation of dashboards (Jivet et al, 2018)

Dashboard evaluation methods include feedback surveys, interviews, focus groups, evaluation instruments, resource use, learning artefacts, dashboard use and grades. As mentioned in the chapter discussing the collection of primary data, a combination of focus groups and interviews was used in this study. During these sessions, the framework components were presented in order to collect views on their usefulness for learning analytics.

Jivet et al (2018) made certain suggestions for the development and evaluation of learning analytics. With regards to the development of the dashboard it is important to ensure that (Jivet et al, 2018):

- Dashboards focus on enhancing reflection and awareness through their use (this is achieved by the C.A.V. framework by enabling awareness and reflection features).
- Dashboard designs are in line with educational concepts (an example is shown in the dashboards illustrating the performance of students according to their profiles).
- Avoid forcing consistent designs across different departments without considering possible adaptation (this is achieved by the C.A.V. framework by enabling different views for different stakeholder groups).
- Seamless integration of dashboards into learning activities (an example is shown with the creation of the G.O.A.L. dashboard used by learners to check their standing in their cohort according to their formative assessment performance).

According to Jivet et al (2018) when evaluating dashboards, it is necessary to prioritise focus on checking whether its goal is achieved, the impact on learners and the dashboard usability. It is also recommended to extend the evaluation beyond usability and usefulness, aiming for an evaluation on how users understand and interpret the visualisations.

The following figure also illustrates the proposed performance model by Jonathan et al (2018) that is divided in two parts. The left pane of the model includes data resources for student data from the different functions of the Learning Management System (LMS) and machine learning techniques used for predicting student performance. The right pane includes Key Performance Indicators associated with strategic goals, staff performance and process effectiveness. These are key aspects that can be used for evaluating whether a dashboard meets the learning analytics needs of the institution and are fully aligned to the key components of the C.A.V. framework.

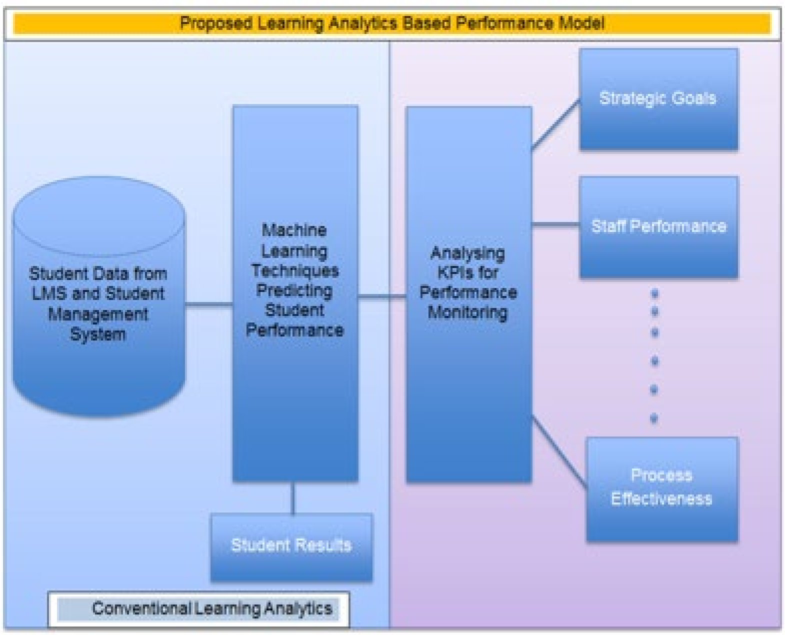

Figure 7-25: Proposed learning analytics-based performance model (Jonathan, 2018)

As mentioned earlier, it is necessary to ensure that the dashboards created are integrated into learning activities. Rienties et al (2017) provide an excellent example of how different "types of interventions have a positive impact on learners' Attitudes, Behaviour and Cognition (ABC)". This is indeed a very useful approach in evaluating dashboards, especially as learning activities may have different impact on individuals. The authors propose a Learning Analytics Intervention and Evaluation Framework (LA-IEF) that is illustrated below. The framework is based on introducing interventions with the use of learning analytics as part of the educational process.

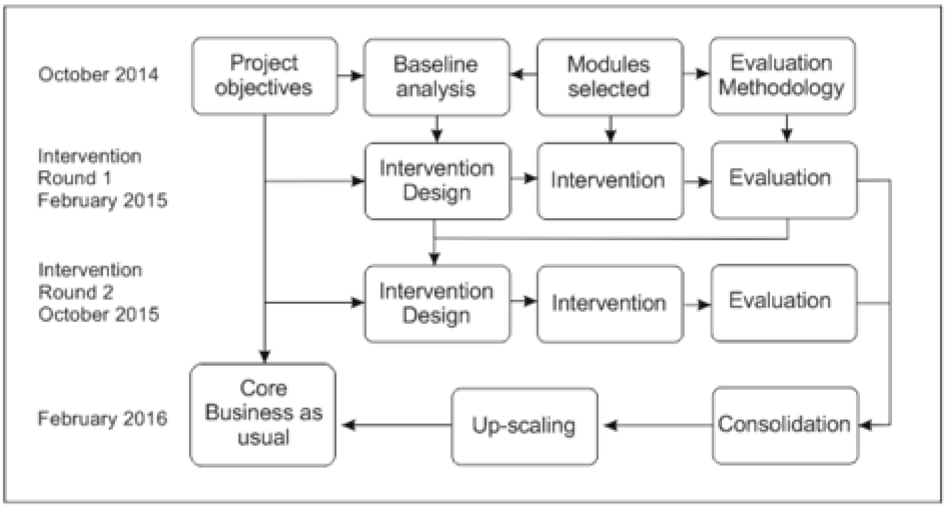

Figure 7-26: LA-IEF framework as implemented at the OU (Rienties et al, 2017))

The framework is based on the concept of iterations of interventions that include the use of learning analytics in selected modules. Similarly, the C.A.V. framework was used in the two modules where dashboards were used to show students their performance through G.O.A.L. assessments and the way student results are affected by their profiles and characteristics.

Ali et al (2012) also used iterations of evaluations for their learning analytics tool. Using the Learning Object Context Ontology (LOCO) framework, they collected qualitative data using open-ended questions to identify areas for improvement. The authors focused on the feedback types provided through learning analytics to the users of such tools, including student performance on learning activities, assessment results, learner activity in discussion forums and chat tools, learner interactions with learning content and student comprehension of learning content. All these features are present in the C.A.V. component functions, especially the analytics and visualisation aspects of the framework.

A different approach is provided by Gasevic et al (2019), which is an approach for systemic adoption of learning analytics in Higher Education Institutions, focusing on data, model and transformation. The authors focus on a number of aspects associated with the adoption of learning analytics to achieve institutional transformation. This can be evaluated against the following areas introduced by Gasevic et al (2019): (i) "building the institutional policy and strategy for learning analytics", (ii) "establishing effective leadership models to drive and oversee the implementation", (iii) "defining principles for privacy protection and ethical use of analytics", (iv) "implementation of learning analytics tools catering the primary stakeholders" and (v) "development of analytics-informed decision-making culture". These aspects of institutional transformation are covered by the guidelines provided at strategic and operational levels as part of the C.A.V. framework.

Another, evidence-based approach towards evaluating learning analytics is proposed by Rienties et al (2016). Their approach is aimed at (i) "accurately and reliably identify learners at-risk", (ii) "identify learning design improvements", (iii) "deliver (personalised) intervention suggestions that work for both student and teacher", (iv) "operate within the existing teaching and learning culture" and (v) "be cost-effective". These aspects are covered by the C.A.V. framework guidelines, apart from the cost-effective dimension, which could not be evaluated within the scope of this research study. This was primarily because there were no means to assess the actual costs associated with the production of learning analytics, such as acquiring dashboard design expertise and calculating staff costs associated with the creation and deployment of learning analytics at institutional level. Rienties et al (2016) suggest the Analytics4Action Evaluation Framework, which involves a six-step intervention process for learning analytics. The C.A.V. framework suggests a similar approach based on interventions and intended transformation that spans across strategic and operational level. The ability to follow the C.A.V. guidelines during policy-making and integrating learning analytics into educational operations serves as an evaluation of the institution's success.

The last influence on evaluating the C.A.V. framework came from the work of Kokoc and Kara (2021) who contextualised the validation of the instrument of the Evaluation Framework for Learning Analytics (EFLA) for the Turkish education system. Their work focused on evaluating learning analytics against data collection, awareness and reflection, as well as impact. These were all criteria used for assessing the C.A.V. framework, with a fundamental difference. Kokoc and Kara (2021) performed their evaluation as part of two studies that were based on quantitative analysis of a survey questionnaire offered to students and staff. This research study considered the qualitative analysis of instructors for the evaluation of the C.A.V. framework, in order to contextualise responses. The use of a survey could mislead the work carried out as emphasis should be given on the conceptual model of how learning analytics should be planned, structured and deployed. Future work could include a survey with students aiming at evaluating the design and content of the dashboard integrated in their learning environment.

<u>Feedback from stakeholders</u>

The C.A.V. framework was presented to the stakeholders participating in the focus group and the series of interviews, collecting different perspectives. These comments helped to determine areas for future improvement, as well as identify strengths of the C.A.V. framework.

The focus group discussed how the G.O.A.L. dashboard was used by the students of two modules as part of their continuous reflection on individual and group performance. All focus group participants (teaching staff at Middlesex University) agreed about the importance of using the visualisation of the student performance as part of the virtual learning environment. Students are provided with a bar chart that determines their ranking within the cohort according to the number of tasks they have successfully completed, after they are observed by lab tutors. Students can access the full set of tasks they have attempted and obtain a full record of their performance using weekly tasks.

The interviews with the Director of Programmes and academics in the Computer Science Department at Middlesex University focused on the need for specific guidelines on how learning analytics can be planned and used. Emphasis was given on the importance of a clear policy that would govern the entire process of suggesting the creation of dashboards, securing access to relevant data and plan visualisation designs. Another interesting point was the need for expert advice on how to integrate dashboards in learning environments and make the use of visualisations part of the learner experience. In particular, the impact of providing learners with the ability to reflect on own performance and view predictions for future performance was considered to be very high.

The interviews with a senior academic manager and teaching staff at Harokopio University emphasised the need for providing visualisations to students as part of their learning environment. Emphasis was given on the fact that current learning analytics are based on ad hoc efforts and that there is the need to have in place a framework regulating the design and use of dashboards. There were particularly positive reactions for the value of introducing guidelines for policy making and operational planning in learning analytics.

The focus group with NUP academics led by a senior manager provided the necessary findings on how the institution uses its business and academic boards as a centralised driving force for its educational data analysis. The views of the focus group were really positive with respect to the role of the framework on improving institutional policy.

Stakeholders involved in the process were as follows:

- Focus Group (MDX)
    - G.D. Module Leader
    - A.T. Associate Lecturer
    - K.M. Graduate Academic Assistant
    - B.L. Graduate Academic Assistant
    - F.A. Graduate Academic Assistant
- Interviews
    - I.V. Associate Professor / Programme Leader
    - T.K. Assistant Professor / Dean / Head of Department
    - S.K. Associate Professor / Programme Leader
    - G.D. Professor / Director of Programmes / Programme Leader
    - S.R. Data Analytics developer / Alumni
- Focus Group (NUP)
    - S.C. Head of Department / Member of QA committee
    - P.C. Module Leader
    - K.Z. Module Leader
    - S.E. Module Leader
    - E.K. Module Leader

The next section presents the revised framework after its evaluation and the comments received. The final version of the framework incorporates all guidelines discussed in earlier chapters.

7.3. Final Framework

The initial reflections on the C.A.V. framework involved comments from learning analytics stakeholders that participated in the interviews conducted during this research study. The C.A.V. framework was also compared to a number of similar conceptual models and frameworks as discussed next.

The original C.A.V. framework identified the need to organise key functions under data collection, analysis and visualisation. This part of the report was further decomposed as shown below to include strategic and operational guidelines for deploying and integrating learning analytics. The figure shows how strategic guidelines include (a) maturity assessment, (ii) identification of the stage of learning analytics development, (c) determining the themes of learning analytics adoption and (d) using a learning analytics checklist. Operational guidelines focus on five steps including (i) determining the added value for the operations of the organisation through learning analytics, (ii) assessing the impact of learning analytics on the institution's stakeholders, (iii) adopting a learning analytics process at department level, (iv) planning the design of dashboards and their contents and (v) identifying factors that affect the implementation of learning analytics at institutional level.

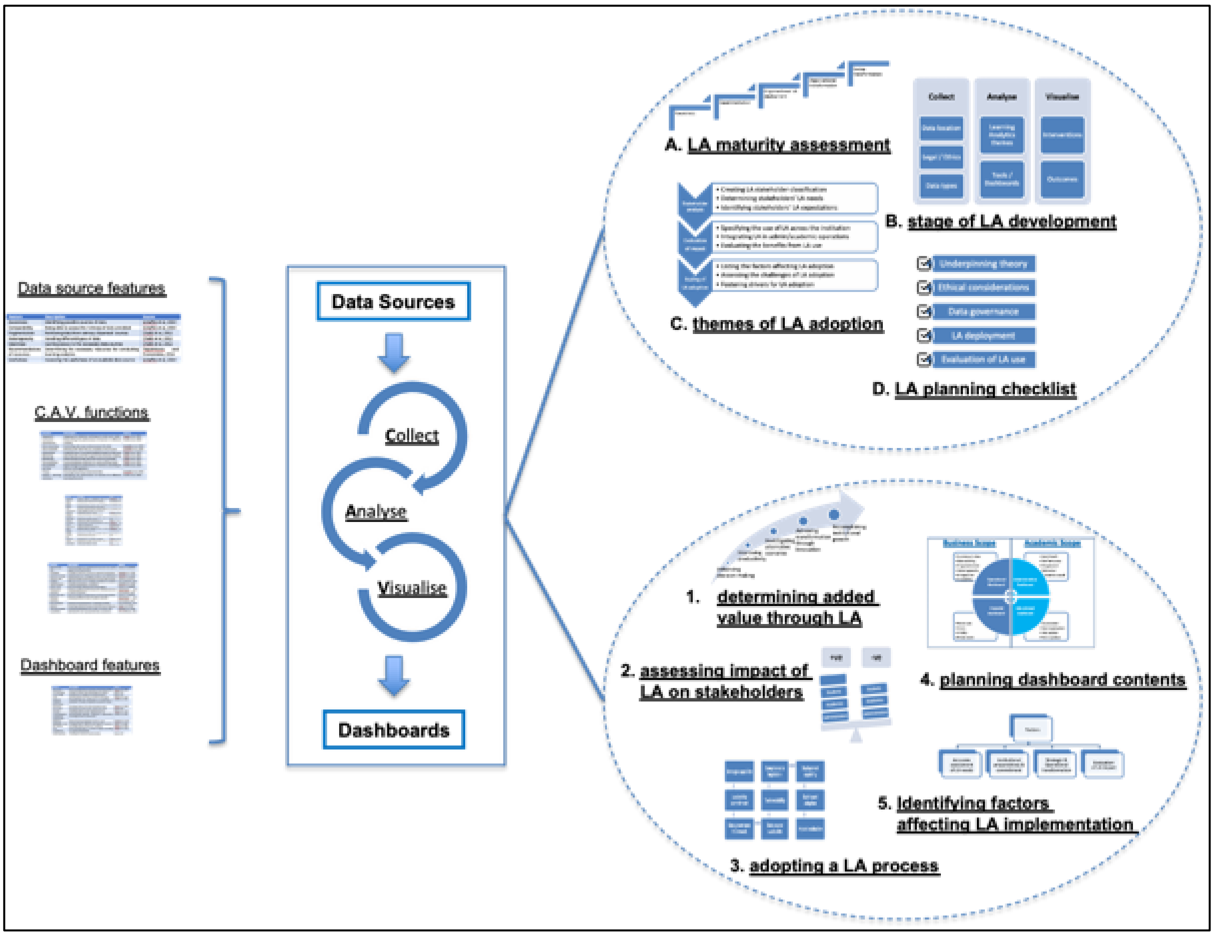

Figure 7-27: The final version of the C.A.V. framework

Apart from the three core components of data collection, analysis and visualisation and the associated guidelines, the framework also includes two components focusing on identification of data sources and dashboard deployment. These are considered to be associated with strategic aspects of learning analytics, and more specifically with the influence institutional capabilities have on learning analytics (i.e., available and well-defined data sources) and the impact of learning analytics on institutions (i.e., effective use of deployed dashboards). The C.A.V. framework supports these two components with the corresponding tables that describe the necessary features of both data sources and dashboards. Similar tables are provided to support the other three components, by describing the necessary functions for data collection, analysis and visualisation.

Finally, the framework provides four compiled lists of elements that should be considered for inclusion in one of the four dashboard types identified during the dashboard design step. These elements are based on an extensive review of literature review papers. It can be described as a literature review of literature reviews! This work resulted in what is perceived to be as complete as possible list of elements that can be used when planning dashboard contents.

In order to reach the final version of the C.A.V. framework this research study included a comparison with several conceptual models and similar frameworks attempting to support learning analytics stakeholders in higher education.

Khousa et al (2015) introduced a social learning analytics approach to cognitive apprenticeship. They “categorise learners into Communities of Practice (CoPs), within which learners thrive collaboratively to build further their career readiness and assert their professional confidence”. The following figure shows how CoP construction and monitoring can help predicting future career developments for individuals. This approach could be utilised as part of the C.A.V. dashboard design guidelines, ensuring that social activities of learners are used to determine their ability to interact with peers after graduation.

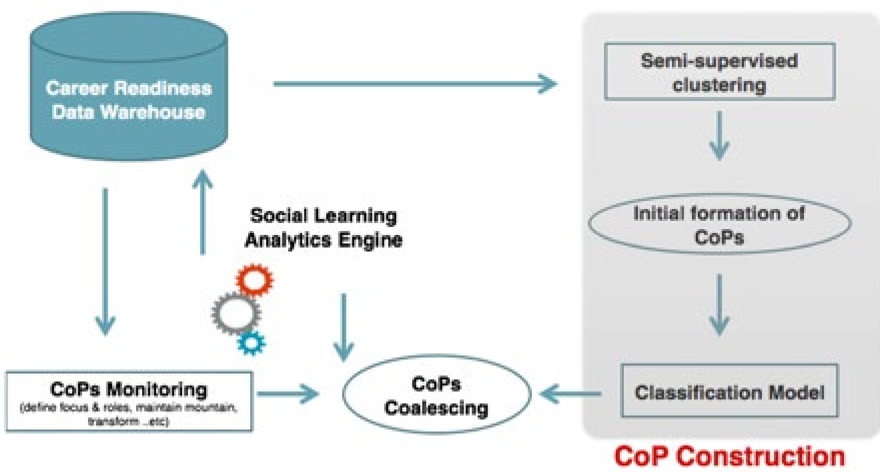

Figure 7-28: Community of Practice (CoP) construction and career prediction (Khousa et al, 2015)

Jo et al (2016) suggest an approach for learner-customised dashboard and treatment model. Their approach is based on three models, namely learning, prediction and action. The Learning Analytics for Prediction and Action (LAPA) model initially focuses on six elements of learning including "the learner's self-regulatory ability, learner psychology, instruction, online learning behaviour, learner characteristics, and types of courses". Log analysis and data measurement are used for creating clusters of learners and supporting a prediction model for each cluster. Finally, the action model identifies the necessary preventative treatment according to the predicted risks for each cluster. This approach lends its cluster approach to the C.A.V. framework in the sense that students in different modules could be grouped in clusters, which can be used to identify common features that help to derive future outcomes for their performance.

Derntl et al (2013) identified common topics between educational data mining and learning analytics including student modelling, data classification, and clustering. These are considered to be core requirements for institutional learning analytics and should be included in all operational guidelines for the deployment of dashboards in taught modules. In line with this approach, the Moodle Analytics API allows Moodle site managers to define prediction models that combine indicators and a target (Moodle, 2021). More specifically, "once new data that matches the criteria defined by the model is available, Moodle starts predicting the probability that the target event will occur".

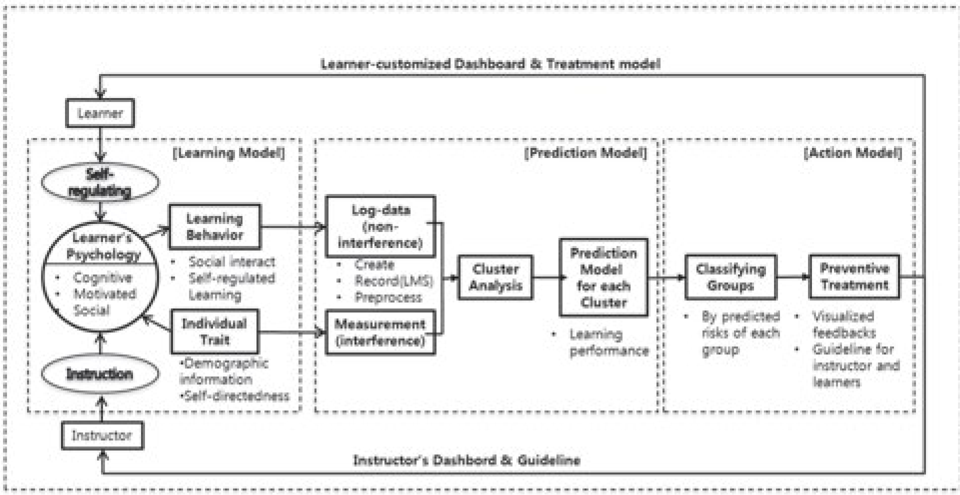

Figure 7-29: Learning Analytics for Prediction and Action Model (Jo, 2012)

The C.A.V. framework was aimed at being quite adaptive to different domains. This is evident when visiting learning analytics models that are created for specific domains, like the proposed model based on the inception of the Internet of Everything (IoE). Ahad et al (2018) discuss how learning analytics can be used in scenarios involving IoE classroom environments, where connected devices may include CO2 detectors, power meters, web cams, lights, humidity monitors, RFID readers, thermostats, fan, smart lighting, air conditioning units, motion detectors, etc. Ahad et al (2018) propose a model that uses inputs from students and wearable devices and sensors, processes the input with the use of machine learning algorithms and produces outputs in the form of corrective measures, improvements and suggestions. This approach is fully aligned with the C.A.V. framework. The analysis component is further discussed by Ahad et al (2018) to include identifying relevant data and data outliers, feeding the data to train the system, identifying key points with respect to analytic requests, and processing the data.

The C.A.V. framework's components have been found in most similar models, such as the process model for learning analytics suggested by Lai and Lahman (2016) that incorporates techniques in a series of processes including data collection, pre-processing, data storage, data selection, data analysis, data formatting and data visualisation. Similarly, according to Siemens (2013), there is a clear impact of the different analysis techniques on the application of learning analytics. Some of the most common application of learning analytics identified by Bienkowski et al (2012) include (i) "modelling user knowledge, behaviour, and experience", (ii) "creating profiles of users", (iii) "modelling knowledge domains", (iv) "trend analysis", and (v) "personalisation and adaptation". These are all applications supported by the C.A.V. framework. The framework is also supportive of all stages of the data loop described by Siemens's (2013) learning analytics model, which includes data collection and acquisition, storage, cleaning, integration, analysis, representation, visualisation and corresponding actions. This is one of the models that are described as quite influential in the field by Moraes et al (2016), along the models proposed by Chatti et al (2012), Freitas et al (2014) and lias and Elias (2011).

Ranjeeth et al (2020) discuss the four dimensions of learning analytics, focusing on what type of data should be considered, who is the affected by the analysis, therefore being the focus of the analytical process, why the analysis is taking place (i.e., the objective) and how the data are analysed (i.e., which techniques to use). The same dimensions are applied by Yuktirat et al (2018) on m-learning analytics, phrased as follows (Chatti et al, 2014):

- What kind of data does the system or theory gather, manage and use for the analysis?
- Why does the system analyse the collected data?
- How does the system perform the analysis of the collected data?
- Who is tagged by the analysis?

Perez-Colado et al (2018) apply the same dimensions on game learning analytics, aiming to answer the 'what', 'how' and 'who' questions. These dimensions are fully covered by the C.A.V. core components that consider data collection, analysis and visualisation.

It is very interesting to observe how learning analytics conceptual models are introduced in different educational contexts. For example, Kazanidis et al (2021) suggest the implementation of a learning analytics framework for interventions supported by Augmented Reality (AR). Their model is illustrated in the following figure, by enabling instructors to follow a series of steps, while engaging with their students through AR applications.

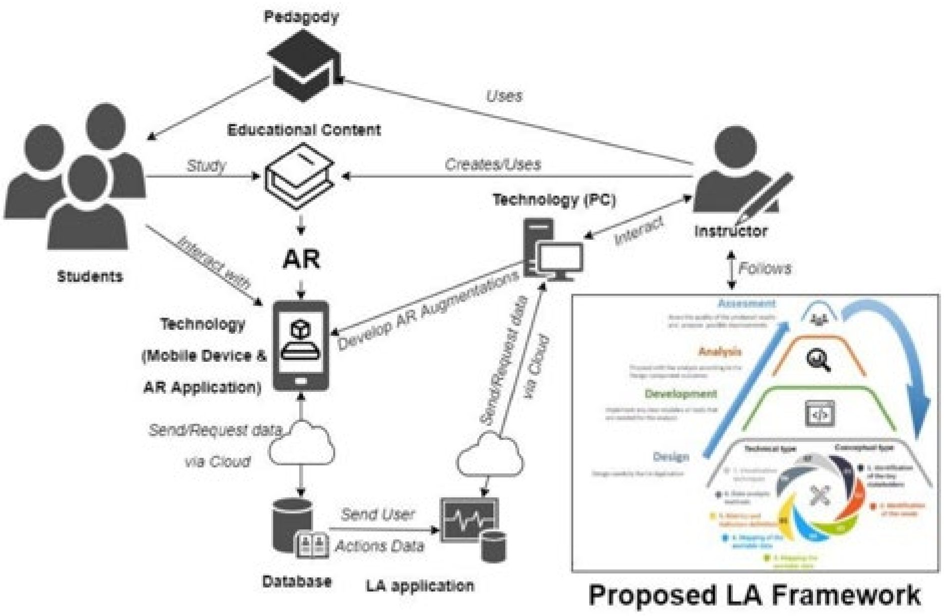

Figure 7-30: Implementation of the LA Framework for augmented reality (AR)-supported interventions (Kazanidis et al, 2021).

The LA framework proposed by Kazanidis et al (2021) involves seven design stages as follows:

- Identification of the key stakeholders – aiming at "identifying all the possible key stake holders".
- Identification of needs – aiming at "providing guidelines on the needs, identification according to the previous stages".
- Mapping the available data – aiming at "studying the available data and the possible ways that they can be analysed"
- Definition of metrics and indicators – aiming at "choosing or creating the metrics and indices that will be used for the LA process".
- Data collection approach – aiming at "adopting an adequate collection approach".
- Data analysis methods – aiming at "deciding on the methods that will be used".
- Visualisation techniques – aiming at "deciding on the visualisation techniques".

The different stages provided by Kazanides et al (2021) correspond to the C.A.V. components. The C.A.V. framework addresses the need to contextualise the support of learning analytics for different learning environments by introducing operational guidelines that enable instructors and dashboard designers to select the most suitable visualisations. The C.A.V. framework also provides operational guidelines that can be moulded to meet the needs of different stakeholders. This helps to shape dashboard designs accordingly, in line with the classification of analytics into learning and academic analytics (Nguyen et al, 2020), which distinguish LA benefits to those intended for learners, and faculty as opposed to those designated for administrators, executives, agencies and funders.

In conclusion, C.A.V. appears to incorporate key aspects covered across different approaches from the relevant literature. The aim of C.A.V. was to integrate the entire set of processes associated with learning analytics. For example, C.A.V. supports all aspects of the cycle of applying data mining in educational systems as suggested by Daud et al (2017). The ability of C.A.V. to adopt different data sources can be used with a wide range of data systems, as identified by Drachsler and Greller (2011), including learning management system logs (e.g., Moodle), student information systems, external services (e.g., Google docs), Intranet (e.g., UniHub functions), course management systems, social networking platforms (e.g., Facebook), e-portfolio systems, mobile platforms (e.g., MDX App) and sensor data (e.g., campus location data). C.A.V. also covers the entire learning analytics process as described by Dyckhoff et al (2012), offering instructions as part of the dashboard design, which can be interpreted by tutors and teachers (e.g., identifying students who need further support, determining which topics affect student performance because of insufficient support or confusing reading materials). Chatti et al (2017) discuss open learning analytics as “an emerging research field that has the potential to deal with the challenges in open and networked environments and present key conceptual and technical ideas toward an open learning analytics ecosystem”. The open learning analytics platform abstract architecture illustrated below is directly comparable to the C.A.V. architecture.

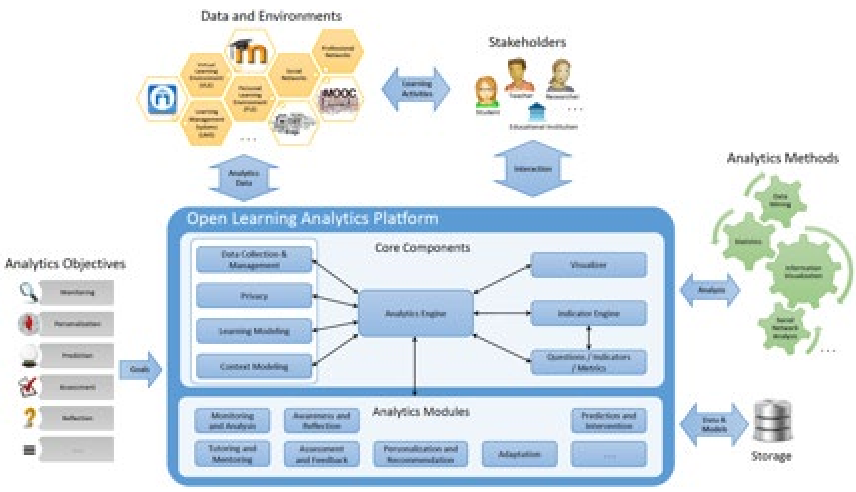

Figure 7-31: Open learning analytics platform abstract architecture (Chatti et al, 2017)

7.4. Summary

In this chapter, the revised C.A.V. framework was discussed and the proposed guidelines for both strategic and operational levels were discussed. This chapter provided the key contributions of this research study in the field of learning analytics in higher education. The next chapter concludes the thesis and focuses on future work to be carried out.

Chapter 8 – Conclusions

In this chapter the conclusions of the research study and future work possibilities are presented. Emphasis is given on how the C.A.V. framework can be applied across the sector.

8.1. Contributions

This research study was triggered by challenges faced by academics who need access to usable and effective learning analytics in their efforts to enhance students' learning experiences. The original idea was to determine the steps required for creating dashboard designs by following widely accepted guidelines. At the early stages of the research the project took another twist by also focusing on how dashboard designs can also incorporate elements required by administrators and senior managers.

The research study contributed in the relevant field in a number of ways, as follows:

- An extensive literature review in the fields of learning analytics and educational data mining.
- A case study analysis in the use of dashboards in Higher Education Institutions.
- A reflection on the state of learning analytics in three universities representing three European countries.
- The C.A.V. framework incorporating:
    - Institutional guidelines for the use of learning analytics (strategic level)
    - User guidelines for the use of learning analytics (operational level)
    - List of features necessary for preparing data resources and dashboards.
    - List of functions necessary for data collection, analysis and visualisation.
    - Dashboard design elements organised across four dashboard types

As a result of this research study, institutions can use the C.A.V. framework even if they do not have any prior work in learning analytics. Institutions can use the institutional guidelines in order to formulate their strategy and prepare policies, as well as introduce procedures at departmental level. Different stakeholders including senior managers, administrators and teaching staff can use the user guidelines to design appropriate dashboards that meet their learning analytics needs.

8.2. Lessons Learnt

The research journey that lasted the best part of two years was challenging but also rewarding. During this research study a number of very useful lessons were learnt. These helped to overcome obstacles, and contributed as valuable experiences for the researcher. The most important lessons that could be shared with researchers in the field are as follows:

- Formulating the research scope – the original idea was to create a set of guidelines for academics who wish to create dashboards for their modules. Eventually it became obvious that this idea could span across different areas of the university and it led to a 'bigger' picture, where guidelines could be provided for more senior staff, as well as non-teaching staff.

Lesson: be prepared to rethink original ideas and re-contextualise the proposed solution.

- Accessing appropriate resources for a grounded theory – there is a plethora of resources in the fields of data analytics, including numerous studies on how academics use dashboards with their data sets. The main challenge was to organise the literature review according to the framework's key components. It was also necessary to proceed with a review of literature reviews in order to assess whether the framework fully addresses the issues identified in the sector.
  Lesson: ensure that the entire field is covered by the papers covered as part of the grounded theory.
- Choosing an appropriate research method – the majority of papers that were retrieved at the early stages of this research study focused on the analysis of student questionnaires and other quantitative techniques. After an initial inclination to adopt a mixed method approach and combine questionnaires with interviews, it was evident that the best approach would combine grounded theory, case study analysis, focus group and interview sessions.
  Lesson: select those techniques that will provide sufficient evidence to support the research tasks and reach concrete findings.
- Collecting primary data – it was evident that this study would require case study analysis, as well as interviews and focus group sessions. This meant that a plan had to be drawn to ensure that data collection was properly organised. Key stakeholders were identified by using contacts from research projects and academic collaborations.
  Lesson: choose the most appropriate participants, aiming for richness of input rather than reaching a specific number of respondents.
- Knowing when to stop – the study continued with further investigations to the point that it was perceived to be extending beyond the scope of an MRes. It was quite a challenge to decide that no more focus groups and papers would be covered in order to determine the elements to be covered by the C.A.V. framework.
  Lesson: determine a suitable number of papers as an initial threshold, be prepared to revise it if needed, and be able to resist the inclination to search more and add more resources.
- Applying the research findings – ideally, the C.A.V. framework would be applied across an entire institution, supporting both its senior managers, but also academics and administrators. Unfortunately, the adoption of the framework at such large scale was unlikely to take place, so the researcher considers applying the framework at specific modules within the Computer Science Department.
  Lesson: be prepared to compromise how extensive the application of the research output can be.
- Dealing with Ethics – it is critical to appreciate how important ethics are in the field of learning analytics. The issue spans beyond compliance with GDPR, as it is necessary to understand how educational data can be misused and also lack of sufficient security and privacy mechanisms can have a dramatic impact on individuals but also introduce risks to the institution.
  Lesson: ensure that appropriate measures are taken for compliance with legislation and the institution's ethical framework.

8.3. Future Work

This research study provided an excellent foundation for extra work. This is likely to lead to a PhD proposal immediately after the outcome of the MRes viva is announced. The research study could be directed towards the following areas:

- Application of the C.A.V. framework – this task will involve the deployment of the framework to a number of modules with emphasis on creating dashboards that can be used by administrators and academics for determining areas that require attention. The aim will be to provide dashboard features for Programme Leaders, Director of Programmes and the Head of Department as well, to achieve a higher-level use of learning analytics.
- Development of specific guidelines – the current version of C.A.V. provides guidelines for institutions and users formed as process steps of what is required for the design and deployment of dashboards. Further works would focus on providing specific guidance on how to perform specific calculations in order to create certain dashboards (e.g., linking attendance to assessment results, associating student engagement in learning activities with progression) .
- Focus on data mining algorithms – a key aspect of future work will focus on using different algorithms to perform data analysis and forecasting.
- Predictive models used in actual case studies – future work would also include the deployment and evaluation of predictive modelling for a range of learning activities such as predicting student performance, cohort progression and student feedback on teaching delivery.

8.4. Summary

The final chapter of the theses provided an overview of the key contributions and paved the way for future research in the field based on the thesis findings.

References

Ahad, M.A., Tripathi, G., Agarwal, P. 2018. Learning Analytics for IoE based Educational Model using Deep Learning Techniques: Architecture, Challenges and Applications. Smart Learning Environments. 5:7 https://doi.org/10.1186/s40561-018-0057-y

Albalowi, A. and A.A. Alhamed. 2017. Big Data and Learning Analytics in Higher Education: Demystifying Variety, Acquisition, Storage, NLP and Analytics. IEEE Conference on Big Data and Analytics. Beijing. China. DOI: 10.1109/ICBDAA.2017.8284118.

Alexandron, G., J.A. Ruiperez-Valiente, S. Lee, and D.E. Pritchard. 2018. Evaluating the Robustness of Learning Analytics Results Against Fake Learners. EC-TEL, Thirteenth European Conference on Technology Enhanced Learning, 3-6 September, 2018, Leeds, United Kingdom, HTTC e.V., 2018.

Ali, L., Hatala, M., Gasevic, D., Jovanovic, J. 2012. A Qualitative Evaluation of Evolution of a Learning Analytics Tool. Computers & Education. 58. 470-489. 10.1016/j.compedu.2011.08.030.

Avella, J.T., Kebritchi, M., Nunn, S.G., and Kanai, T. 2016. Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. Journal of Interactive Online Learning. June 2016.

Bailey, P. 2016. Scoping Learning Analytics for Further Education and Skills. Notes from the 6th UK Learning Analytics Network event in Newport. JISC Data Analytics. (available at: https://analytics.jiscinvolve.org/wp/2016/04/22/scoping-learning-analytics-for-further-education-and-skills-14april2016/, accessed on: 02/01/2022)

Banihashem, S.K., Aliabadi, K., Ardakani, S., Delavar, A., and Ahmadabadi, M. 2018. Learning Analytics: A Systematic Literature Review. Interdisciplinary Journal of Virtual Learning in Medical Sciences. Volume 9. Issue 2. DOI: 10.5812/ijvlms.63024.

Banoor, R.Y., Issack, S.M., and R. Frank. 2019. The Use of Learning Analytics to Improve Online Learning Outcomes: A Systematic Literature Review. Pan-Commonwealth Forum, 9-12 September 2019, Edinburgh, Scotland.

Bienkowski, M., Feng, M., & Means, B. 2012. Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics. Washington, DC: U.S. Department of Education.

Brooks, C. and T.L.B. Thayer. 2016. Institutional Analytics in Higher Education. Research Report. CO: ECAR, March 2016. (available at: https://library.educause.edu/-/media/files/library/2016/2/ers1504ia.pdf , last accessed on: 01/01/2022).

Campbell JP., Oblinger DG. 2007. Academic analytics, EDUCAUSE White Paper. (Retrieved from: https://net.educause.edu/ir/library/pdf/PUB6101.pdf, last accessed on: 26/10/2019)

Carmichael, C. 2019. Big Data on Campus: How IT Leaders Enable Higher Education Analytics. White Paper. The University of Mississippi. pp 1-16. (available at: https://olemiss.edu/tableau/whitepaper_big_data_on_campus_highered_1.pdf, last accessed on: 01/01/2022).

Chatti, M.A., Dyckhoff, A.L., Schroeder, U., and Thus, H. 2012. A Reference Model for Learning Analytics. International Journal of Technology Enhanced Learning (IJTEL) – Special Issue on "State-of-the-Art in TEL", 4(5/6), 318-331. https://doi.org/10.1504/IJTEL.2012.051815.s

Chatti, M., Lukarov, V., Thüs, H., Muslim, A., Yousef, A.M.F., Wahid, U., Greven, C., Chakrabarti, A. and Schroeder, U. 2014. Learning Analytics: Challenges and Future Research Directions. E-learning and Education, 10.

Chatti, M., Muslim, A. and Schroeder, U. 2017. Toward an Open Learning Analytics Ecosystem in Big Data and Learning Analytics in Higher Education (pp.195-219). DOI: 10.1007/978-3-319-06520-5_12.

Chiu, M., & Fujita, N. 2014. Statistical discourse analysis of online Discussions: Informal cognition, social metacognition, and knowledge creation. Proceedings of the fourth international conference on learning analytics and knowledge (pp. 217–225). ACM.

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gasevic, D., Buckingham Shum, S., Fisher, J. 2016. Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement. Sydney: Office for Learning and Teaching. Retrieved from: https://opus.lib.uts.edu.au/handle/10453/117173, last accessed on: 01/01/2022).

Daud, A., Aljohani, N., Abbasi, R., Lytras, M., Abbas, F. and Alowibdi, J. 2017. Predicting Student Performance using Advanced Learning Analytics. 2017 International World Wide Web Conference Committee (IW3C2), WWW 2017 Companion, April 3-7, 2017, Perth, Australia. DOI: 10.1145/3041021.3054164.

Derntl, M., Gunnemann, N., Klamma, R. 2013. A Dynamic Topic Model of Learning Analytics Research. (available at: http://ceur-ws.org/Vol-974/lakdatachallenge2013_01.pdf, last accessed on: 01/01/2022).

Drachsler, H. and Greller, W. 2011. The Pulse of Learning Analytics Understandings and Expectations from the Stakeholders. LAK'12, 29 April – 2 May 2012, Vancouver, BC, Canada. ACM International Conference Proceeding Series. DOI: 10.1145/2330601.2330634.

Dyckhoff, A., Zielke, D., Bültmann, M., Chatti, M. and Schroeder, U. 2012. Design and Implementation of a Learning Analytics Toolkit for Teachers. Educational Technology and Society, 15 (3), 58-76.

Eassom, S. 2014. Using Big Data and Analytics to Create a Smarter University Campus. IBM Government Industry Blog. September 2014. (available at: https://www.ibm.com/blogs/insights-on-business/government/using-big-data-and-analytics-to-create-a-smarter-university-campus/, last accessed on: 01/01/2022).

Elias, T.D. 2017. Data-Supported Learning Design: A Customer Care Training Example. in Proceedings of the 7th International Learning Analytics & Knowledge Conference (LAK17): Practitioner Track.

Ferguson, R., & Clow, D. 2017. Where is the evidence? A call to action for learning analytics. Proceedings of the seventh international learning analytics & knowledge conference (pp. 56–65). ACM.

Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., and Arnab, S. 2014. Foundations of dynamic learning analytics: Using university student data to increase retention. British Journal of Educational Technology.

Gasevic, D., Tsai, Y.S. and Drachser, H. 2021. Editorial Learning Analytics in Higher Education – Stakeholders, Strategy and Scale. The Internet and Higher Education. October 2021. DOI: 10.1016/j.iheduc.2021.100833.

Gasevic, D., Tsai, Y-S., Dawson, S., & Pardo, A. 2019, How do we start? An Approach to Learning Analytics Adoption in Higher Education, International Journal of Information and Learning Technology. (available at: https://doi.org/10.1108/IJILT-02-2019-0024, last accessed on: 01/01/2022).

Guzman-Valenzuela, C., Gomez-Gonzalez, C., Rojas-Murphy Tagle, A., and Lorca-Vyhmeister, A. 2021. Learning Analytics in Higher Education: A Preponderance of Analytics but very Little Learning? International Journal of Educational technology in Higher Education. (2021) 18:23. (available at: https://doi.org/10.1186/s41239-021-00258-x, last accessed on: 01/01/2022).

Hilliger, I., Miranda, C., Schuit, G., Duarte, F. Anselmo, M., Parra, D. 2021. Evaluating a Learning Analytics Dashboard to Visualize Student Self-Reports of Time-on-task: A Case Study in a Latin American University. LAK21: 11th International Learning Analytics and Knowledge Conference. DOI:10.1145/3448139.344820

IBM. 2016. 7 Ways Smart Universities Use Data and Analytics. Center for Digital Education. January 2016. (available at: https://www.govtech.com/education/7-Ways-Smart-Universities-Use-Data-and-Analytics.html , last accessed on: 01/01/2022).

IBM. 2019. Building a Smarter Campus: How Analytics is Changing the Academic Landscape. Campus Technology. pp. 1-8.

Ifenthaler, D. and Yau, J.Y-K. 2020. Utilising Learning Analytics to Support Study Success in Higher Education: A Systematic Review. Educational Technology Research Development 68: 1961-1990. (available at: https://doi.org/10.1007/s11423-020-09788-z, last accessed on: 01/01/2022).

Jivet, I., Scheffel, M., Specht, M., and Drachsler, H. 2018. License to Evaluate: Preparing Learning Analytics Dashboards for Educational Practice. LAK ’18, March 7–9, 2018, Sydney, NSW, Australia. https://doi.org/10.1145/3170358.3170421

Jo, I.H., Park, Y., Yoon, M., Sung, H. 2016. Evaluation of Online Log Variables that Estimate Learners' Time Management in a Korean Online Learning Context. International Review of Research in Open and Distributed Learning. Volume 17, Number 1. DOI: 10.19173/irrodl.v17i1.2176.

Job, M.A. 2018. An Efficient Way of Applying Big Data Analytics in Higher Education Sector for Performance Evaluation. International Journal of Computer Applications. Vol. 180. No. 23. Pp. 25-32.

Joksimovic, S., Kovanovic, V. and Dawson, S. 2019. The Journey of Learning Analytics. HERDSA Review of Higher Education, Vol. 6, 2019. (available at: www.herdsa.org.au/herdsa-review-higher-education-vol-6/37-63, last accessed on: 01/01/2022).

Jonathan, J., Sohail, S., Kotob, F., and Salter, G. 2018. The Role of Learning Analytics in Performance Measurement in a Higher Education Institution. 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)1201-1203. DOI: 10.1109/TALE.2018.8615151.

Kazanidis, I., Pellas, N., and Christopoulos, A. 2021. A Learning Analytics Conceptual Framework for Augmented Reality-Supported Educational Case Studies. Multimodal Technologies and Interaction. 2021, 5, 9. https://doi.org/10.3390/mti5030009

Kellen, V. A Recktenwald, and S. Burr. 2013. Applying Big Data in Higher Education: A Case Study. Cutter Consortium Data Analytics and Digital Technologies Executive Report. December 2013. (available at: https://www.cutter.com/article/applying-big-data-higher-education-case-study-400836, last accessed on: 01/01/2022).

Khalil, M. 2017. Learning Analytics in Massive Open Online Courses. Doctoral Thesis. Graz University of Technology.

Khousa, E.A., Atif, Y. and Masud, M.M. 2015. A social learning analytics approach to cognitive apprenticeship. Smart Learning Environments. 2:14. DOI 10.1186/s40561-015-0021-z.

Kimmons, R. and Caskurlu, S. 2020. The Students' Guide to Learning Design and Research. EdTech Books. (available at: https://edtechbooks.org/studentguide, last accessed on: 01/01/2022).

Kokoc, M. and Kara, M. 2021. A Multiple Study Investigation of the Evaluation Framework for Learning Analytics: Instrument Validation and the Impact of Learner Performance. Educational Technology and Society. 24(1). Pp 16-28.

Krawitz, M. J. Law, and S. Litman. 2018. How Higher-Education Institutions can Transform themselves Using Advanced Analytics. Report on Public and Social Sector. August 2018. pp. 1-6. (available at: https://www.mckinsey.com/industries/social-sector/our-insights/how-higher-education-institutions-can-transform-themselves-using-advanced-analytics#, last accessed on: 01/01/2022).

Lai, F.Q. and Lehman, J.D. 2015. Chapter 9 Learning and Knowledge Analytics in Open Education: Selected Readings from the AECT-LKAOE 2015 Summer International Research Symposium. Springer (available at: https://ebrary.net/45184/education/process_model_learning_analytics, last accessed on: 01/01/2022).

Lauria, E.J.M., Presutti, E., Sokollof, M., Guarino, M. 2017. Crossing the Chasm to Big Data: Early Detection of At-Risk Students in a Cluster Computing Environment in Proceedings of the 7th International Learning Analytics & Knowledge Conference (LAK17): Practitioner Track.

Leitner, P., Khalil, M., and Ebner, M. 2017. Learning Analytics in Higher Education—A Literature Review. In: Learning Analytics: Fundaments, Applications, and Trends. Peña-Ayala, A. (Ed.). Springer International Publishing. DOI: 10.1007/978-3-319-52977-6_1. pp. 1-23.

Lias, T. E., & Elias, T. 2011. Learning Analytics: The Definitions, the Processes, and the Potential (available at https://landing.athabascau.ca/file/download/43713, last accessed on: 01/01/2022).

Manderveld, J., de Wit, M., and van Dompseler, H. 2017. SURF Learning Analytics Experiment: Hands-on Experience for Dutch Higher Education in Proceedings of the 7th International Learning Analytics & Knowledge Conference (LAK17): Practitioner Track.

Mangaroska, K., and Giannakos, M. 2018. Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning. IEEE Transactions on Learning Technologies. PP. 1-1. 10.1109/TLT.2018.2868673.

Mavrikis, M., Charlton, P., Katsifli, D. The Potential of Learning Analytics and Big Data, (available at: http://www.ariadne.ac.uk/issue/71/charlton-et-al/ , last accessed on: 01/01/2022).

Martin, F., Ndoye, A. 2016. Using Learning Analytics to Assess Student Learning in Online Courses, Journal of University Teaching & Learning Practice, 13(3). (available at: http://ro.uow.edu.au/jutlp/vol13/iss3/7 , last accessed on: 01/01/2022).

Matcha, W., Uzir, N.A.A., Gasevic, D. and Pardo, A. 2019. A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. IEEE Transactions on Learning Technologies · May 2019. DOI: 10.1109/TLT.2019.2916802.

Mazriani, G. 2018. "10 ways colleges use analytics to increase student success". eCampus News. August 2018. (available at: https://www.ecampusnews.com/2018/08/13/10-ways-colleges-use-analytics-to-increase-student-success/?all , last accessed on: 01/01/2022).

Moodle. 2021. Analytics API. (available at: https://docs.moodle.org/dev/Analytics_API, last accessed on: 01/01/2022).

Moraes, E.M. de, Silva, M.T. da, and Souza, M.C. 2016. Models to implement Learning Analytics: A literature review. (available at: https://www.pomsmeetings.org/ConfProceedings/065/Full%20Papers/Final%20Full%20Papers/065-0931.pdf, last accessed on: 01/01/2022).

NFETLHE 2013. Report to the European Commission on Improving the Quality of Teaching and Learning in Europe's Higher Education Institutions. Forum Insights – National Forum for the Enhancement of teaching and learning in Higher Education. June 2013.

NFETLHE 2015. Reaching Out: Student Drop-Out. Briefing paper – National Forum for the Enhancement of teaching and learning in Higher Education. 2015.

NFETLHE 2016a. Reaching Out: Why Students Leave. Briefing paper 1 – National Forum for the Enhancement of teaching and learning in Higher Education. November 2016. DOI: 10.6084/m9.figshare.4244960.v1.

NFETLHE 2016b. Transition from Further Education and Training to Higher Education. Forum Insights – National Forum for the Enhancement of teaching and learning in Higher Education. October 2016. DOI: 10.6084/m9.figshare.4056594.v1.

NFETLHE 2017. Learning Analytics: Benefits for Institutions. Forum Insights – National Forum for the Enhancement of teaching and learning in Higher Education. March 2017. DOI: 10.6084/m9.figshare.5092396.

Nguyen, A., Gardner, L., Sheridan, D. 2020. Data Analytics in Higher Education: An Integrated View. Journal of Information Systems Education, 31(1), 61-71.

Oliva-Cordova, L.M., Garcia-Cabot, A., Aado-Salvatierra. 2021. Learning Analytics to Support Teaching Skills: A Systematic Literature Review. IEEE Access.

Ong, V.K. 2015. Business Intelligence and Big Data Analytics for Higher Education: Cases from UK Higher Education Institutions. Information Engineering Express International Institute of Applied Informatics. Vol. 2. No. 1. pp 65-75. (available at: https://pdfs.semanticscholar.org/0dd2/44240c421776d772723b0f13042554df8918.pdf, last accessed on: 01/01/2022).

Ouli, J. and Voutilainen, T. 2019. Learning Analytics and the Processing of Student Data in Universities. Edilex 2019/37, Edilex 30 September 2019, (available at: www.edilex.fi/artikkelit/20064, last accessed on: 01/01/2022).

Papamitsiou, Z., and Economides, A.A. 2014. Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. Educational Technology & Society, 17 (4), 49–64.

Patwa, N., Seetharaman, A., Sreekumar, K., and Phani, S. 2018. Learning Analytics: Enhancing the Quality of Higher Education. Research Journal of Economics. 2018, 2:2.

Perez-Colado, I., Alonso-Fernandez, C., Freire, M., Martinez-Ortiz, I. and Fernandez-Manjon, B. 2018. Game learning analytics is not informagic! IEEE Global Engineering Education Conference (EDUCON), 2018, pp. 1729-1737. DOI: 10.1109/EDUCON.2018.8363443.

Queiroga, E.M., Lopes, J.L., Kappel, K., Aguiar, M., Araújo, R.M., Munoz, R., Villarroel, R., Cechinel, C. 2020. A Learning Analytics Approach to Identify Students at Risk of Dropout: A Case Study with a Technical Distance Education Course. Applied Sciences 2020, 10, 3998. (available at: https://doi.org/10.3390/app10113998, last accessed on: 01/01/2022).

Ranjeeth, S., Latchoumi, T.P., and P. Victer Paul. 2019. A Survey on Predictive Models of Learning Analytics. International Conference on Computational Intelligence and Data Science (ICCIDS 2019).

Reichley, C., Phillips, J., Gates, K., Brown, V., Snyder, R., Brooks, L., Kuo, K., Roljevic, J. 2018. Using Analytics to Answer Important Institutional Questions. Educause Review. December 2018. (available at: https://er.educause.edu/articles/2018/12/using-analytics-to-answer-important-institutional-questions, last accessed on: 01/01/2022).

Rienties, B, Boroowa, A., Cross, S., Kubiak, C., Mayles, K., Murphy, S. 2016. Analytics4Action Evaluation Framework: A Review of Evidence-Based Learning Analytics Interventions at the Open University UK. Journal of Interactive Media in Education(1), article no. 2. http://dx.doi.org/doi:10.5334/jime.394

Rienties, B., Cross, S., Zdráhal, Z. 2017. Implementing a Learning Analytics Intervention and Evaluation Framework: What Works? Chapter in Big Data and Learning Analytics in Higher Education (pp.147-166). DOI: 10.1007/978-3-319-06520-5_10.

Romero C., Ventura, S. 2013. Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), pp 12-27.

Romero C., Ventura, S. 2015. Learning Analytics: From Research to Practice. Technology, Knowledge and Learning. DOI: 10.1007/s10758-015-9244-x

Romero, C., Ventura, S. 2020. Educational Data Mining and Learning Analytics: An Updated Survey. WIREs Data Mining and Knowledge Discovery. Volume 10, Issue 2. May/June 2020 e1355. (available at: https://doi.org/10.1002/widm.1355, last accessed on: 01/01/2022).

Scheffel, M., Drachsler, H., Stoyanov, S., Specht, M. 2014. Quality indicators for learning analytics. Journal of Educational Technology & Society, 17(4), 117–132.

Sclater, N. 2014. Learning Analytics: the Current State of Play in UK Higher and Further Education. JISC.

Sclater, N., Peasgood, A., Mullan, J. 2016. Learning Analytics in Higher Education: A Review of UK and International Practice. JISC Full Report. pp. 1-40. (available at: https://www.jisc.ac.uk/sites/default/files/learning-analytics-in-he-v2_0.pdf, last accessed on: 01/01/2022).

Shacklock, X. 2016. From Bricks to Clicks: The Potential of Data and Analytics in Higher Education. Higher Education Commission. pp. 1-76. (available at: https://www.policyconnect.org.uk/hec/sites/site_hec/files/report/419/fieldreportdownload/frombrickstoclicks-hecreportforweb.pdf, last accessed on: 01/01/2022).

Shehata, S., Tan, J.P-L. (Eds.). 2017. Practitioner Track Proceedings of the 7th International Learning Analytics & Knowledge Conference (LAK17). Simon Fraser University, Vancouver, Canada (13–17 March). SoLAR.

Siemens, G. 2012. Learning Analytics: Envisioning a Research Discipline and a Domain of Practice LAK12: 2nd International Conference on Learning Analytics and Knowledge, 29 April – 2 May 2012, Vancouver, BC, Canada.

Siemens, G. 2013. Learning Analytics: The Emergence of a Discipline. American Behavioural Scientist. 57(10) 1380–1400. DOI: 10.1177/0002764213498851

Siemens, G., Baker, R. 2012. Learning analytics and educational data mining: Towards communication and collaboration. ACM International Conference Proceeding Series. 10.1145/2330601.2330661.

Siemens, G., Dawson, S., Lynch, G. 2013. Improving the Quality and Productivity of the Higher Education Sector – Policy and Strategy for Systems-Level Deployment of Learning Analytics. Canberra, Australia: Office of Learning and Teaching, Australian Government. (available at: https://solaresearch.org/wp-content/uploads/2017/06/SoLAR_Report_2014.pdf, last accessed on: 01/01/2022).

Sin, K., Muthu, L. 2015. Application of Big Data in Education Data Mining and Learning Analytics – A Literature Review. ICTACT Journal of Soft Computing: Special Issue on Soft Computing Models for Big Data. July 2015. Volume 05. Issue 04. ISSN 2229-6956.

Slade, S., Prinsloo, P. 2013. Learning analytics ethical issues and dilemmas. American Behavioural Scientist, 57(10), 1510-1529.

StELA. 2021. Final Report – Student Evaluation and Learning Analytics Project. (available at: https://www.ul.ie/quality/sites/quality/files/user_media/documents/StELA%20final%20project%20report%20and%20recommendations.pdf, last accessed on: 01/01/2022).

The Information Lab. 2015. Middlesex University sees "A huge improvement in the delivery and clarity of data" using Tableau. Customer Success Story. https://www.theinformationlab.co.uk/wp-content/uploads/2015/12/Middlesex-University-Case-Study.pdf, last accessed on: 01/01/2022).

Tulasi, B. 2013. Significance of Big Data and analytics in higher education. International Journal of Computer Applications 68: 14.

Tynan, B., Buckingham Shum, S. 2013. Designing Systemic Learning Analytics at the Open University. SoLAR Open Course, Strategy & Policy for Systemic Learning. (Slides available at: http://www.slideshare.net/sbs/designing-systemic-learning-analytics-at-the-open-university, last accessed on: 17/12/2021)

Universities UK. 2016. Analytics in Higher Education. CIVITAS Learning, in partnership with JISC. November 2016. pp 1-11. (available at: https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2016/analytics-in-higher-education.pdf, last accessed on: 26/10/2019).

Universities UK. 2018. Patterns and Trends in UK Higher Education. JISC. September 2018. pp 1-39. (available at: https://www.universitiesuk.ac.uk/facts-and-stats/data-and-analysis/Documents/patterns-and-trends-in-uk-higher-education-2018.pdf, last accessed on: 01/01/2022).

Viberg, O., Hatakka, M., Balter, O., and Mavroudi, A. 2018. The current landscape of learning analytics in higher education. Computers in Human Behavior 89 (2018) 98–110.

Viberg, O., Khalil, M., Baars, M. 2020. Self-Regulated Learning and Learning Analy3cs in Online Learning Environments: A Review of Empirical Research. LAK’20, March 2020, Frankfurt Germany.

Welsh, J. 2020. Learning Analytics. Chapter in The Students' Guide to Learning Design and Research. Kimmons, R. and Caskurlu, S. (Eds.) EdTech Books.

Winne, P.H., and Hadwin, A.F. 1998. Studying as Self-Regulated Learning. Metacognition Educational Theory Practices. Vol. 93, pp. 277–304, 1998.

Wong, B.T.M. 2017. Learning Analytics in Higher Education: An Analysis of Case Studies. Asian Association of Open Universities Journal. Vol. 12. No. 1. pp. 21-40. (available at: https://www.emerald.com/insight/content/doi/10.1108/AAOUJ-01-2017-0009/full/pdf?title=learning-analytics-in-higher-education-an-analysis-of-case-studies, last accessed on: 01/01/2022).

Yuktirat, C., Sindhuphak, A., Kiddee, Krissana. 2018. M-learning for the Art of Drawing: Informal Learning for a Digital Age. International Journal of Interactive Mobile Technologies (iJIM). 12. 152. 10.3991/ijim.v12i5.9207.

Appendix A – University Case Studies on the use of data analytics

1. University of South Florida

Reference	McIntosh, E. 2016. Building Your Organization Capacity for Success: Getting Past the Growing Pains. (available at: https://www.civitaslearningspace.com/building-organization-capacity-success-getting-past-growing-pains/, last accessed on: 02/01/2022)
Objective	improving student graduation and retention rates.
Data collection	using the University's Student Information System and Learning Management System.
Analysis practice	(i) identifying individual students who are predicted at risk of not re-enrolling or completing their degree, (ii) guiding the Academic Care Team in order to get students back on the path to graduation, (iii) figuring out how to deliver the right support, to the right student, at the right time.

2. Wollongong University

Reference	sSclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-J-University-of-Wollongong.pdf, last accessed on: 02/01/2022)
Objective	analysing online student discussion forums, assessing student engagement and identifying students who are isolated from the main discussion.
Data collection	not provided.
Analysis practice	(i) identifying students who are less engaged than others, (ii) identifying groups where just a small number of students are dominating the discussion of a much larger group.

3. Rio Salado Community College

Reference	Scott, A. 2016. Marketplace: Colleges tap big data to help students stay in school. (available at: https://www.marketplace.org/2016/07/20/education/tapping-big-data-help-college-students-succeed/, last accessed on: 02/01/2022)
Objective	identifying at risk students and students to drop their course.
Data collection	(i) students' grades and test scores, (ii) financial aid status, (iii) frequency of interaction with online course, (iv) VLE materials and discussion boards, (v) proportion of students with scholarships.
Analysis practice	(i) predicting high risk students for non-completion, (ii) identifying students who are less engaged than others.

4. University of Edinburgh

Reference	Available at: https://www.ed.ac.uk/information-services/learning-technology/more/learning-analytics, last accessed on 01/01/2022.
Objective	improving course design, attainment, and the student experience.
Data collection	(i) Learning Analytics Report Card (LARC), (ii) Virtual Learning Environment (VLE), (iii) Massive Open Online Courses (MOOC), (iv) Online Video Annotations for Learning (OVAL), (v) multimodal Self-Regulated Learning (SRL), (vi) flipped Classrooms and learning dashboard.
Analysis practice	(i) understanding of the use of computational analysis in education, (ii) becoming aware about student attitudes to data and privacy, (iii) improving the success and experience of MOOC learners, (iv) using digital traces of interactions with OVAL, (v) analysing digital traces recorded by VLEs, (vi) tracing multimodality data measures of students' states during learning.

5. Purdue University

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-A-Purdue-University.pdf, last accessed on: 02/01/2022)
Objective	enhancing student success at a course level, improve overall retention and graduation rates and identify potential problems as early as the second week in the semester.
Data collection	(i) performance metrics, (ii) effort metrics, (iii) prior academic history, (iv) student characteristics, (v) attendance, (vi) students' use of the VLE, (vii) information held in the VLE gradebook.
Analysis practice	developing an 'early warning system' showing at-risk students over a number of years.

6. University of Maryland, Baltimore County

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-B-University-of-Maryland-Baltimore-County.pdf, last accessed on: 02/01/2022)
Objective	focusing on the support of low-level graduates and the high drop-out rates, using predictions to support students, and identifying effective teaching practices with a view to enhancing future provision.
Data collection	(i) student VLE activity like forum usage, to identify usage patterns, (ii) the relationship of final grades to specific types of VLE activity.
Analysis practice	a "Check My Activity" tool enables students to compare their VLE activity on a specific course with the activity summary of other students. When grades are posted by faculty, the students can see how their activity compares with others obtaining the same, higher or lower grades. They can see to what extent other students have used individual VLE tools compared to themselves.

7. New York Institute of Technology

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-C-New-York-Institute-of-Technology.pdf, last accessed on: 02/01/2022)
Objective	intervening early with at-risk students.
Data collection	(i) data on previous students, (ii) key risk factors including grades, (iii) admission application data, (iv) registration/placement test data, (v) survey taken by every student when they did the placement exam, (vi) financial data.
Analysis practice	the dashboard, shows (i) whether students are predicted to return to their studies the following year, (ii) the confidence level in the prediction from the analysis, and (iii) the reasons for the prediction that may include a disparity between fees for the rest of the qualification and the student's funds, student uncertainty about career goal, or students working a large number of hours per week as well as studying.

8. California State University

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-D-California-State-University.pdf, last accessed on: 02/01/2022)
Objective	evaluating student success based on demographic data and VLE use.
Data collection	(i) multiple demographic variables for students, (ii) students' current VLE effort, (iii) student characteristic, (iv) student motivation, (v) learning style.
Analysis practice	points out that earlier studies consistently found that combinations of student characteristics correlated much better with student success than single demographic variables do.

9. Marist College

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-E-Marist-College.pdf, last accessed on: 02/01/2022)
Objective	providing predictive models to help at-risk students.
Data collection	(i) demographic details such gender and age, (ii) aptitude data such as high school scores, (iii) VLE usage.
Analysis practice	the predictive model helps to provide students with earlier feedback on their progress, allowing them to address any issues before it is too late.

10. Edith Cowan University

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-F-Edith-Cowan-University.pdf, last accessed on: 02/01/2022)
Objective	enhancing student retention, success and good rating, as well as offering proactive support for retaining students.
Data collection	(i) grades, (ii) university entrance scores., (iii) language skills, (iv) use of the Enterprise Information Management system functions, (v) demographic information and (vi) student progress information.
Analysis practice	generating a list of students who may need support, using a dashboard that includes case details, contact log, view appointments, add/manage appointments, case attachments and student incidents.

11. University of New England

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-G-University-of-New-England.pdf, last accessed on: 02/01/2022)
Objective	identifying students who were struggling, so they could be offered timely support, as well as developing a "dynamic, systematic and automated process that would capture the learning wellbeing status of all students intra-event".
Data collection	(i) self-reported information about happiness from students via emoticons, (ii) extensive use of social media including Facebook, Twitter and Flickr, (iii) e-Motion student input, (iv) class attendance, (v) previous study history, (vi) prior test results, (vii) assignment submissions, (viii) VLE access patterns, (ix) previous AWE scores.
Analysis practice	qualitative feedback from students showed that Early Alert was successful in increasing the students' sense of belonging to a community, and sharing their experiences of study increasing their motivation.

12. Open University (UK)

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-H-Open-University-UK.pdf, last accessed on: 02/01/2022)
Objective	enhancing student success.
Data collection	(i) students who withdraw, (ii) information about students from the tutors, (iii) VLE, and (iv) e-Library data".
Analysis practice	a student dashboard will enable students to track their progress and make better choices about their study path, while predictive analytics are being developed to address student progression at individual module level and institutional level. Analytics may identify a problem with the content and assessment of modules, and can also identify successful learning designs, enabling the sharing of best practice.

13. Nottingham Trent University

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-I-Nottingham-Trent-University.pdf, last accessed on: 02/01/2022)
Objective	(i) enhancing retention, (ii) improving attainment, (iii) increasing a sense of belonging and (iv) improving student support.
Data collection	actual engagement data comes from four separate systems including (i) the virtual learning environment, (ii) the card access database, (iii) the assessment submission system, (iv) the library system, (v) the student information system and (vi) student background information.
Analysis practice	(i) show a progress line indicating engagement, comparing individual students with the rest of the cohort, and showing the student's engagement rating, (ii) notify tutors when students engagement drops from good to partial, where students don't use the library by a particular cut-off date, or for students with a history of late submission, (iii) help students using the system to challenge themselves and their peers. When students competing with one another to have the highest engagement score."

14. Open Universities Australia

Reference	Sclater, N. 2016. Learning Analytics in Higher Education. (available at: https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-K-Open-Universities-Australia.pdf, last accessed on: 02/01/2022)
Objective	offering a personalised study experience.
Data collection	(i) the student profile, (ii) location, (iii) socio-demographic factors, (iv) prior knowledge, (v) VLE use, (vi) the curriculum profile.
Analysis practice	the PASS framework uses the output from the learning analytics to personalise the output information and to enable interventions like suggestion of additional modules for the student or providing evidence for redesigning a section of the curriculum.

15. University of Oklahoma

Reference	Available at: https://www.mulesoft.com/lp/whitepaper/api/digital-transformation-higher-education, last accessed on: 01/01/2022
Objective	(i) succeeding in the recruitment of talented students, (ii) driving better business intelligence, (iii) achieving digital citizenship for students, faculty, alumni and partners, and (iv) transitioning to real-time student services via mobile.
Data collection	(i) past student data, (ii) financial aid, (iii) MOOC, (iv) student engagement.
Analysis practice	the goals is to have the largest and most academically prepared student body.

16. University of Central Florida (UCF)

Reference	Mariani G. 2018. 10 ways colleges use analytics to increase student success (available at:https://www.ecampusnews.com/2018/08/13/10-ways-colleges-use-analytics-to-increase-student-success/2/, last accessed on: 02/01/2022)
Objective	improving the school's progress toward metrics in student success.
Data collection	(i) Student Information System (SIS), (ii) Learning Management System (LMS), (iii) VLE, (iv) financial information.
Analysis practice	(i) optimising academic program offerings to better appeal to the target student population, (ii) scheduling enough sections of classes, with enough faculty to teach them, to meet demand and help students to graduate on time, (iii) considering enrolment forecasts, budgets, computing resources, parking, and more, (iv). identifying which students are on track, which are near completion, and which are eligible to be auto-graduated.

17. Hillsborough (FL) Community College

Reference	Mariani G. 2018. 10 ways colleges use analytics to increase student success (available at:https://www.ecampusnews.com/2018/08/13/10-ways-colleges-use-analytics-to-increase-student-success/2/, last accessed on: 02/01/2022)
Objective	identifying previously enrolled students who need to complete only 25% or less of their graduation and students who may be closer to another degree than their declared program of study.
Data collection	(i) Student Information System (SIS), (ii) Learning Management System (LMS), (iii) VLE, (iv) students' historical data.
Analysis practice	identifying which students are on track, which are near completion, and which are eligible to be auto-graduated.

18. University of Alabama

Reference	Mariani G. 2018. 10 ways colleges use analytics to increase student success (available at:https://www.ecampusnews.com/2018/08/13/10-ways-colleges-use-analytics-to-increase-student-success/2/, last accessed on: 02/01/2022)
Objective	identifying the correlation between a student asking for an official transcript and then leaving the university.
Data collection	students asking for an official transcript.
Analysis practice	the goal is to find the correlation between a student asking for an official transcript and then leaving the university, as this request is a red flag, and it's actually the best indicator of students leaving the university in their first year. The scope is to identify these students early on and reach out to provide resources to help them.

19. Des Moines Area Community College (DMACC)

Reference	available at: https://www.sas.com/en_ca/customers/des-moines-area-community-college.html, last accessed on: 02/01/2022)
Objective	attrition among new students.
Data collection	(i) financial data, (ii) students paying full tuition.
Analysis practice	placing students into developmental courses based on entrance exam scores often caused them to drop out. Disheartened students paying full tuition but not gaining college credit would simply quit. Guided by this knowledge, offers a college readiness class and counselling to prepare future students for full admission the next semester, also uses inexpensive online refresher classes for students with low English or math placement test scores, instead of requiring semester-long, no-credit developmental classes.

20. Carnegie Mellon University

Reference	Carmichael C. 2014. Big Data on Campus How IT Leaders Enable Higher Education Analytics (available at: https://olemiss.edu/tableau/whitepaper_big_data_on_campus_highered_1.pdf, last accessed on: 02/01/2022).
Objective	(i) predicting students' future learning behaviour, (ii) discovering or improving domain models, (iii) studying the effects of different kinds of pedagogical support and (iv) advancing scientific knowledge about learning and learners.
Data collection	(i) VLE. Use, (ii) participation in discussion forums, (iii) practice tests scores, (iv) models that categorise student activity from basic behavioural data.
Analysis practice	administrators can look at detailed data across different classes to examine progress for all students at a school, to see what works and what does not in a particular classroom, and to do so with less effort. District administrators can use data from this kind of dashboard as a basis for determining whether a particular learning intervention is effective at promoting student learning, even at the level of individual concepts.

21. Massachusetts Institute of Technology

Reference	Available at: https://www.mulesoft.com/lp/whitepaper/api/digital-transformation-higher-education, last accessed on: 01/01/2022.
Objective	Modernising the systems that widely support MIT's administrative services and educational enterprise, and enabling the MIT community to serve themselves and do things for themselves by giving them better data/better access to data.
Data collection	(i) faculty/department data, (ii) registrar's office data, (iii) financial aid, (iv) participation in Massive Open Online Courses (MOOC), (v) student engagement.
Analysis practice	(i) 77% of students say mobile technology has improved their grades, (ii) 62% of students say that mobile technology leaves them better

prepared for classes., (iii) 85% of students used their smartphones and seems that they want to learn using them.

22. University of San Diego

Reference	Available at: https://www.mulesoft.com/lp/whitepaper/api/digital-transformation-higher-education, last accessed on: 01/01/2022.
Objective	(i) providing better customer experiences, (ii) creating a digital revolution by making a number of new mobile apps available for their students.
Data collection	(i) faculty/department data, (ii) registrar's office data, (iii) financial aid, (iv) Massive Open Online Courses (MOOC), (v) student engagement.
Analysis practice	(i) 77% of students say mobile technology has improved their grades, (ii) 62% of students say that mobile technology leaves them better prepared for classes, and (iii) 85% of students used their smartphones and seems that they want to learn using them.

23. University of Otago

Reference	Daniel B. 2013. Technology Enhanced Analytics (TEA) in Higher Education. (available at: https://files.eric.ed.gov/fulltext/ED557187.pdf, accessed on: 02/01/2022).
Objective	institutional performance and progress in order to predict future performance.
Data collection	(i) curriculum data, (ii) administrative data, (iii) department data, (iv) teaching and learning data, (v) student data, and (vi) research data.
Analysis practice	(i) helping in an effective manner so that operational activities related to academic programming and student strengths and weaknesses can be identified and appropriately rectified, (ii) acquiring the capability to make timely data-driven decisions across all departments and divisions, and (iii) developing rigorous data modelling and analysis to reveal the obstacles to student access and usability and to evaluate any attempts at intervention.

24. University of Bedfordshire

Reference	Sclatter N. 2014. Learning Analytics in Higher Education. (available at: https://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf, accessed on: 02/01/2022)
Objective	(i) improving the student experience, and (ii) identifying at risk students.
Data collection	(i) student information systems, (ii) VLE., (iii) attendance records, (iv) swipe cards, (v) proximity cards, (vi) Management Information System (MIS), and (vii) finance and HR systems.
Analysis practice	monitoring the engagement patterns of individual students over a period of time or making group comparisons and benchmarking reports. The analytics enable staff to be proactive and offer learners

any support they need at the earliest stage of their study, helping them to re-engage, progress throughout the levels and essentially fulfil their academic goals.

25. Bridgwater College

Reference	Sclatter N. 2014. Learning Analytics in Higher Education. (available at: https://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf, accessed on: 02/01/2022)
Objective	ensuring that every learner has the best possible opportunity to be successful and to gain a qualification.
Data collection	(i) previous years' data, (ii) self-assessment judgements, (iii) GCSE results, (iv) socio-economic group and working status, (v) ProMonitor system, and (vi) student logs.
Analysis practice	metrics include progression rates, success rates, destinations (e.g., university or employment), and where the College sits against national Level 3 value added data. There is also information gathered from student surveys some of which e.g., satisfaction rates, can be benchmarked against external data.

26. University of Derby

Reference	Sclatter N. 2014. Learning Analytics in Higher Education. (available at: https://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf, accessed on: 02/01/2022)
Objective	developing an excellent student experience by better understanding learners and their diverse needs.
Data collection	(i) the student information system, (ii) Blackboard Learn (module data), (iii) Turnitin, (iv) Talis / other library systems, (v) attendance data, (vi) swipe cards, (vii) proximity cards.
Analysis practice	a single bespoke dashboard that has been developed presents attendance data from the student information system to academic staff, displaying where the students have recorded themselves as absent. Business intelligence dashboards are also being scoped for areas such as admissions, enrolment, assessment, graduation and drop-out rates. The Career Hub system takes data from the student information system and can accept notes from careers advisors.

27. Lancaster University

Reference	Sclatter N. 2014. Learning Analytics in Higher Education. (available at: https://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf, accessed on: 02/01/2022)
Objective	an interactive transcript that shows students their progress.
Data collection	(i) University Student Information (LUSI), (ii) attendance and submission records, (iii) VLE (Moodle), (iv) Library systems include

	ALMA, Primo, EzProxy, Shibboleth and Aleph Archives, and (v) swipe cards.
Analysis practice	Tableau is used for visualising content in the data warehouse, which is currently focussed on admissions data. The intention is to begin to use Tableau more for visualising operational data. The Library is developing a dashboard for Library staff, formed from different Tableau reports. It will show student book borrowing and downloads of e-books, chapters and e-journals, together with reporting of physical attendance.

Appendix B – Interview & Focus Group questions

The focus of the research study is to provide sufficient evidence supporting the value of the C.A.V. framework as a tool for educational analytics. The C.A.V. framework acts like a bridge between institutional data sources and the design of dashboards to be used for the representation of (i) operational, (ii) financial, (iii) administrative and (iv) educational data. The framework consists of three phases as follows:

- Data collection
- Data analysis
- Data visualisation

It was decided to use a number of interviews as an evaluation of the key aspects associated with each of the C.A.V. phases. Therefore, the questionnaire is organised accordingly in three sections. A fourth section dedicated to the administration tasks of educational analytics is covered by parallel discussion during each section aiming to find out how the entire process is taking place in the participants' institution.

<u>Collection</u>

1. How do we determine the necessary data sources for the intended analysis and visualisation?
   (clarification: what criteria should be used to identify the most suitable data sets for analysis and visualisation)
2. How do we assess whether sufficient volume of data is available for the intended analysis?
   (clarification: how many records are needed for different analysis tasks)
3. How do we prepare the collected data for the analysis that must be performed?
   (clarification: what criteria should be used for ETL to ensure that the data set is suitable for analysis and visualisation)

<u>Analysis</u>

4. How do we ensure that data analysis techniques are adapted to suit the specific domain?
   (clarification: what changes are required in the way data are analysed when we are working with different businesses, parts of the organisation, or data sets)
5. How do we differentiate between the analysis of strategic and operational data?
   (clarification: in the case of educational institutions., what is different when analysing business data such as HR performance indicators and finance when comparing to academic data such as student progress and programme evaluation)
6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
   (clarification: how do we decide which calculations are necessary for the available data set in order to produce effective visualisations)

<u>Visualisation</u>

7. How do we select appropriate visualisations for the analysis conducted and the data sets available?
   (clarification: what criteria should be used to identify the most appropriate visualisation techniques for the data sets available and the analysis required)
8. How do we align the various dashboard components to different decisions that must be made?
   (clarification: what plans are required so certain dashboard areas are mapped to specific decisions that must be made based on the provided visualisations)
9. How do we evaluate whether the dashboard contents meet the analysis needs?
   (clarification: what criteria can be used to assess whether the dashboard completely meets the needs of its intended users)
10. How do we design the dashboard in order to maximise its usability by the intended users?
   (clarification: what heuristics, techniques or guidelines should we follow when creating the visualisation interface of a dashboard)

Appendix C – Interview & Focus Group transcript

This appendix includes the transcripts of the sessions used to collect primary data from a wide range of participants. The sessions included the following:

- Focus Group (MDX)
    - G.D. Module Leader
    - A.T. Associate Lecturer
    - K.M. Graduate Academic Assistant
    - B.L. Graduate Academic Assistant
    - F.A. Graduate Academic Assistant
- Interviews
    - I.V. Associate Professor / Programme Leader
    - T.K. Assistant Professor / Dean / Head of Department
    - S.K. Associate Professor / Programme Leader
    - G.D. Professor / Director of Programmes / Programme Leader
    - S.R. Data Analytics developer / Alumni
- Focus Group (NUP)
    - S.C. Head of Department / Member of QA committee
    - P.C. Module Leader
    - K.Z. Module Leader
    - S.E. Module Leader
    - E.K. Module Leader

Focus Group (MDX)

1. How do we determine the necessary data sources for the intended analysis and visualisation?
    - (KM) – depending on number of students involved, but also their prior experience with analytics (e.g., use of software – maybe we need to include trainings to help end users to become more familiar)
    - (AT) – engagement and attendance, logs from virtual learning environment, assessment submissions
    - (BL) – divide GOALs and assessment in terms of their difficulty level, monitor performance according to level of task
    - (KM) – the role of gamification in learning analytics
    - (BL) – we need comparison between courses as students may progress differently and have different grades in different modules
    - (KM) – different progress in modules may be a sign of different motivation levels

2. How do we assess whether sufficient volume of data is available for the intended analysis?
    - (KM) – depending on how many credits/weeks the module is to decide how many weeks we need to collect weeks for, perhaps some students are not engaged in the first couple of weeks, also some students are not in groups for quite a few weeks
    - (FA) – depends on the number of students of the module and then we can aim for certain thresholds
    - (BL) – as modules have different numbers of students, we need to determine how comparable certain modules are
    - (AT) – lab tutors can tell the quality of students from the first couple of weeks, but some information cannot enter analytics (e.g., students getting help from others)
    - (BL) – obtaining historical data from previous modules to anticipate performance
    - (KM) – we may have issues when assessing a student's performance using data from other modules

3. How do we prepare the collected data for the analysis that must be performed?
    - (FA) – missing data such as last names or erroneous student numbers
    - (KM) – group numbers must be double checked, same with the file names of student submissions
    - (AT) – double entries, duplicated data
    - (FA) – cleaning the data
    - (BL) – wrong names and student numbers, group work does not show who has actually done the work and the proportion of effort
    - (KM) – specific requirements for the data, for example whether gender affects the data collected
    - (FA) – the type of data and the variables used for analysis need to be considered.

4. How do we ensure that data analysis techniques are adapted to suit the specific domain?

- (KM) – depends on the features or functions being used, for example we shifted from LinkedIn, Twitter, and Facebook and now we use wikis and blogs
- (AT) – making sure that students are familiar with the technology and techniques used, and the platforms being used, also the time needed to gain familiarisation with tools and techniques, younger users use different apps compared to older ones, the experience of individuals in using digital technologies
- (KM) – accessibility in getting the data, for example different programs using different forms to access the data, in contrast some platforms provide reports that generate analysis of data (e.g., timestamps), the API of different software such as LinkedIn and Facebook may cause issues to collect data.

5. How do we differentiate between the analysis of strategic and operational data?
    - (GD/BL) – we should have a list of needs and then use the analysis needs in order to select which data are more relevant to give us the answers.
    - (AT) – apply different filters

6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
    - (KM) – comparing results of the analysis with previous years and investigate whether there is consistency
    - (AT) – using thresholds
    - (BL) – impossible to have full proof analysis without the use of human intervention, perhaps using data analysis results with contextualisation and interpretation from individuals who can comprehend the data sets and the analysis conducted
    - (FA) – considering use keywords or qualitative techniques to analyse the data

7. How do we select appropriate visualisations for the analysis conducted and the data sets available?
    - (KM) – relevance to the data collected (for example collecting the gender of students can help us derive more meaning to the students results)
    - (KM) – clarity of the data (pie charts may be difficult to use with, say with groups having 10 members

8. How do we align the various dashboard components to different decisions that must be made?
    - (GD/KM) – align data to the decisions that must be made

9. How do we evaluate whether the dashboard contents meet the analysis needs?
    - (BL) – set aims for the dashboard, and assess the extent the aims were met, also compare with similar techniques, and test them out.
    - (KM) – check whether it meets the data collection requirements, using certain criteria, aspects

10. How do we design the dashboard in order to maximise its usability by the intended users?
    - (FA) – radar chart is a good technique
    - (KM) – depends on the volume of data that are represented

- (FA) – depending on whether it is qualitative or quantitative analysis
- (KM) – use keyword word cloud
- (BL) – keep it simple, avoid excessive use of data, perhaps maximum of 3-4 areas
- (AT) – aim for a clear view with the first look at the board of what is included and what is the key message
- (FA) – use criteria to decide what is more important in the dashboard
- (AT) – apply filters and sorting for the visualisation

Interview: I.V. Associate Professor / Programme Leader

1. How do we determine the necessary data sources for the intended analysis and visualisation?
    - Module evaluation from students – questionnaires at the end of each term filled online. Focus on delivery, content, and learning experience.
    - Annual report at institutional and departmental levels are required based on specific KPIs. Focus is on internal evaluation and includes grades for different modules (A, B, C, D classifications), research output for individual FT staff, research output in collaboration with MSc students and researchers, external funding (projects), collaborations with external institutions including Erasmus and various mobilities.

2. How do we assess whether sufficient volume of data is available for the intended analysis?
    - Typically, less than 10 evaluations mean that the module is not considered for annual teacher awards.
    - There is no concrete plan on thresholds for establishing learning analytics.

3. How do we prepare the collected data for the analysis that must be performed?
    - Data collected with the use of online forms and then manual ETL is performed.
4. How do we ensure that data analysis techniques are adapted to suit the specific domain?
    - A significant part of the analysis is done in an ad hoc manner.
    - There is a dedicated institutional group that is discussions with the national group in order to have certain directions when establishing institutional KPIs.

5. How do we differentiate between the analysis of strategic and operational data?
    - Strategic targets are in the form of specific aims that are communicated as KPIs at institutional level.
    - There is an ad hoc approach in collecting such data.
    - Operational data are collected for each module and focus on evaluation of the modules.
    -
6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
    - The KPIs drive the analysis of the data that are collected. The KPIs are used as a framework to ensure the institution collects the necessary data. For example, the institution now focuses on collecting data regarding its alumni.

7. How do we select appropriate visualisations for the analysis conducted and the data sets available?
    - This depends on the KPIs that are used, the types of aims whether they are strategic or operational, the aims the institution has, etc.

8. How do we align the various dashboard components to different decisions that must be made?

- It would be a departmental decision to decide the structure and storyline of the dashboard.
- An educational independent body (ETHAAE) could provide guidelines on dashboard components.

9. How do we evaluate whether the dashboard contents meet the analysis needs?
    - At institutional level, a committee would approve the dashboard before its deployment.
    - Departments can suggest dashboard components and designs.

10. How do we design the dashboard in order to maximise its usability by the intended users?
    - Uncertain whether specific testing would be in place before the dashboard launch.

Interview: T.K. Assistant Professor / Dean / Head of Department

1. How do we determine the necessary data sources for the intended analysis and visualisation?
    - There is a student record database (e-studies) that can provide statistics to a certain extent. The student record system is controlled by a sub-contractor offering access to the database so the department can execute its own queries. There is also a module evaluation survey (based on questionnaires) obtaining feedback from students for each individual module. Live survey (open-source form-based system that is hosted in house) is used for the module survey with participants including registered students for each module. There is a central unit responsible for the quality of HUA (MoDiP – Quality Assurance Unit), which authorises any changes to the evaluation questionnaires. There is one such unit for each HEI.
2. How do we assess whether sufficient volume of data is available for the intended analysis?
    - MoDiP will be responsible to determine the thresholds in collaboration with the Ministry of Education. There is also an independent body (ETHAE) that provides independent QA audits for all institutions. This agency would likely to offer input for the data volume deemed sufficient. There is also an issue with the number of registered students in each module as there may be students from previous cohorts still registered in the system as enrolled to specific modules. There is no attendance system in place recording student presence, attendance, interaction. Attendance is recorded only for those lab sessions that are compulsory.
3. How do we prepare the collected data for the analysis that must be performed?
    - Again, MoDiP would be in charge to assess the validity of the data collected. Currently the questionnaire is offered only to those students that have not exceeded a certain number of years (to avoid skewed data due to students who do not attend). OMEA is a team responsible for the QA of each department. OMEA teams from each department liaise with MoDiP that is responsible across the entire institution. OMEA will identify any issues stemming from the evaluation results. A report is prepared for each module leader following the student evaluation (survey). There is no dashboard generation mechanism or any survey to run mid-term. There is also no historical data kept. There is no archive kept or historical records to show the evolution of certain metrics. This is implemented manually.
4. How do we ensure that data analysis techniques are adapted to suit the specific domain?
    - The university governing body (Senate) would be in charge of a decision for the way any LA techniques would be adopted and possibly adapted for different departments. MoDiP would have to propose an initiative and request the necessary permission and supporting resources including funding. MoDiP could recommend a working group or alternatively the university committee would recommend the membership of the working group. This would be unlikely a decision to be made by the Vice Chancellor alone.
5. How do we differentiate between the analysis of strategic and operational data?

- This year for the first time a number of performance indicators (e.g., research, internationalisation, infrastructure) will be used to assess whether the departments of the institution will get the entire budgeted funds or receive only 80%. This is a directive provided by the Ministry. The HoD is responsible for the teaching delivery of all modules. OMEA will be responsible to assess certain performance indicators for teaching. Data such as catchment areas and recruitment patterns may be collected centrally. The HoD can propose certain initiatives to the assembly of the department. Certain strategic data are likely to be directed to the programme leader, as well as the HoD. For example, programme viability and need to increase marketing budget. MoDiP also needs to have access to all data and the vice-rector of academic affairs who is responsible for MoDiP.

6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
    - OMEA and MoDiP will decide on the dashboards required.
7. How do we select appropriate visualisations for the analysis conducted and the data sets available?
    - MoDiP recommended the structure and process followed for the module evaluation. Any changes in the reporting mechanisms would go through MoDiP for approval. Ideally this would be in alignment with HoDs.
8. How do we align the various dashboard components to different decisions that must be made?
    - Each department (up to now) has identified its own KPIs. For example, the department decides the thresholds used for assessing performance in different criteria. Therefore, the members of a department would expect to see dashboard components in line with the criteria used for assessing individual performance. This would mean the HoD, MoDiP and OMEA would need to reach agreement.
9. How do we evaluate whether the dashboard contents meet the analysis needs?
    - This will be decided between OMEA and the HoD after checking whether the dashboard provides sufficient visualisations for the required KPIs. OMEA is meant to be the representing body for the academic members of the department. It would be too chaotic to open the discussion to all members of each department.
10. How do we design the dashboard in order to maximise its usability by the intended users?
    - HoD, OMEA and MoDiP would agree. It could be an agenda item for a departmental meeting to obtain more feedback.

Interview: S.K. Associate Professor / Programme Leader

1. How do we determine the necessary data sources for the intended analysis and visualisation?
    - At micro level I would need to investigate how student data is stored for different modules and programmes. I would also check whether any historical data are available. Assessment data would be at the core of the data sets used, together with, the use of forums and any other student interactions. I would consider whether learning analytics could be used to make students aware about their progress, how they would be assessed, and the criteria used for assessing them. At macro-level we could evaluate the performance of different modules of the programme; perhaps focusing on students' skillsets or how the modules are being taught.

2. How do we assess whether sufficient volume of data is available for the intended analysis?
    - Data are not focusing on providing insights on student behaviour and interactions. For example, student emails are sent to the wrong recipients. Sometimes students do not provide the correct information about their needs. Ideally a 10-15% of students should respond to small surveys of a few questions that is conducted even weekly, providing a basic strengths/weaknesses review of the module.
3. How do we prepare the collected data for the analysis that must be performed?
    - I would like to have fully measurable data sets. For example, ranking questions are not useful as they can be affected by subjective views. There should be a way to ensure that data sets are prepared so they can be comparable and measurable to help specific decisions to be made. Fully quantifiable data sets are required.
4. How do we ensure that data analysis techniques are adapted to suit the specific domain?
    - It is necessary to consider the module learning outcomes but also focus on the different departmental focus. Considering the topics taught and the concepts covered in modules, we could adapt the data collection to the different domains. The focus should be on determining the skillsets associated with each domain.
5. How do we differentiate between the analysis of strategic and operational data?
    - If we can use external sources to validate the results of the analysis it would certainly help. For example, using external identifiers for hot topics that are offered by the competition. It is necessary to use external benchmarks for strategic or marketing analytics. Operational data can focus on internal benchmarks and comparison between different modules.
6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
    - I would try to identify which data sets represent the needs of different stakeholders. Different data sets are intended to be used by teachers or students. Do we aim for more students, better students, or avoid losing students? Analytical tasks should be mapped to LA requirements.
7. How do we select appropriate visualisations for the analysis conducted and the data sets available?

- It depends on what the visualisation needs are. For example, are we looking for the outliers of the data sets or are we looking for mean or median values? Are we planning to use any benchmarks or baselines to evaluate different modules? Are we aiming for certain issues to be addressed or fix specific problems? It is necessary to avoid repeating the design of analytics but instead to have a series of questions that would cover all strategic and operational aspects and then select from the bank of questions which ones to be used.

8. How do we align the various dashboard components to different decisions that must be made?
    - Depending on who will be using the dashboard we can determine the required components. For example, we may need the executive view of certain areas, followed by role-based visualisations. For example, HoD, DoP, PL and ML views requires using different KPIs.
9. How do we evaluate whether the dashboard contents meet the analysis needs?
    - I would match the dashboard design to the requirements for the LA. Can I directly find the answers to my questions? Are there any visualisation techniques that provide answers?
10. How do we design the dashboard in order to maximise its usability by the intended users?
    - Role-based design makes the dashboard more compact and geared towards the intended user. UX designers could quickly spot areas for improvement and identify interface problems. Technical experts are not necessarily enough so they should have UX designers involved.

Interview: G.D. Professor / Director of Programmes / Programme Leader

1. How do we determine the necessary data sources for the intended analysis and visualisation?
    - Across the institution there are several data collection projects aiming to provide sufficient data sets for key operations. For example, we have detailed recruitment data, as well as progression and assessment information. Data sets include collaborative partners, locations of institutions we work with and also information about research funding. At department level there are several ongoing initiatives focusing on learning analytics on student performance,
2. How do we assess whether sufficient volume of data is available for the intended analysis?
    - There are not specific thresholds imposed on learning analytics, however for student surveys the institution aims for at least 50% return, preferably much higher in line with national average. In ad hoc learning analytics it is necessary for the analysts to have a clear target for a percentage that is considered representative as a sample of the entire cohort.
3. How do we prepare the collected data for the analysis that must be performed?
    - All data should undergo a thorough ETL process, removing duplicates and any noise. This should be taken for granted and the institution should ensure that anyone involved in learning analytics is trained in data preparation techniques.
4. How do we ensure that data analysis techniques are adapted to suit the specific domain?
    - Before the analysis is conducted analysts need to ensure that specific techniques are identified and certain calculations are selected to produce the required dashboard elements.
5. How do we differentiate between the analysis of strategic and operational data?
    - It is critical to increase awareness about strategic requirements for learning analytics. Teaching staff should be supported in understanding the various KPIs and how they drive certain analytical processes. There should be an alignment between strategic and operational needs for learning analytics.
6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
    - The institution's decision-making process on selecting dashboard designs needs further dissemination at departmental level. It is also necessary for departments to initiate bottom-up processes for the inception of dashboard designs and identification of learning analytics requirements
7. How do we select appropriate visualisations for the analysis conducted and the data sets available?
    - A unit of experts is responsible at institutional level to discuss with stakeholders the dashboard designs. Departments can also introduce ad hoc visualisations to meet their needs. There is room for introducing a process with clear steps on how to decide visualisation designs.
8. How do we align the various dashboard components to different decisions that must be made?
    - Dashboard users and stakeholders of learning analytics communicate their requirements to dashboard designers. There is a clear need for guidance of what visualisation options are available for the institution's learning analytics users.

9. How do we evaluate whether the dashboard contents meet the analysis needs?
    - There should be an evaluation checklist that could be used by dashboard users. Furthermore, the visualisation designers and the end users should reach consensus before signing off the dashboard and finalising the visualisation project.
10. How do we design the dashboard in order to maximise its usability by the intended users?
    - Users must make their requirements clear to designers. Furthermore, there should be a prioritisation of the dashboard elements that are required. Ideally, a list of options should be provided during consultation sessions in order to jointly select wit designers the most appropriate calculations and associated visualisations.

Interview: S.R. Data Analytics developer / Alumni

1. How do we determine the necessary data sources for the intended analysis and visualisation?
    - In the first instance sensors could be used to monitor student movement, entry or exit from rooms and various other data. Questionnaires and evaluation surveys can be used to collect student views. Surveys should be based on 10-15 questions and have specific metrics associated with them (mainly to avoid student boredom and ensure that students take the surveys). To incentivise the student, the questionnaire should be designed in a way that it becomes attractive. There were no clear pointers on how to access Moodle analytics for own interaction with the modules.
2. How do we assess whether sufficient volume of data is available for the intended analysis?
    - If there is a small proportion of engagement to the survey (Say 30%) there should be no assumptions based on the collected sample. There should be a combination of factors, on one hand the incentivisation of students as discussed above and the students' own commitment and motivation to participate. There are also technical aspects to be considered as sensors or web forms may be used to collect the data.
3. How do we prepare the collected data for the analysis that must be performed?
    - Perhaps the most important element of data analytics should be ensuring that data are consistent, accurate and need no further engineering. For the data set to be at the highest level of preparedness (which is critical in any analysis), we need to assess which portion of the data is collected in a way that requires no more analysis.
4. How do we ensure that data analysis techniques are adapted to suit the specific domain?
    - It is important to adapt the analysis according to the domain – this is directly associated with the dashboard designs as well. For example, certain disciplines are more exposed to the use of dashboards.
5. How do we differentiate between the analysis of strategic and operational data?
    - There is a clear differentiation between analysis that is intended for strategic tasks when compared to internal analysis for own purposes. It is important to consider the needs of the intended users of the dashboards.
6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
    - Academic staff are also involved with administrative tasks. Therefore, it would be ideal to have an academic member of staff responsible for the analysis and visualisation of educational data.
7. How do we select appropriate visualisations for the analysis conducted and the data sets available?
    - This relates to the previous question. It is necessary to focus on the end users. During my recent employment, we had some clients who had no idea about dashboards. We had to design the dashboards in a way that would enable the novice, inexperienced users would not have much difficulty in using the dashboards. Our dashboard designs were therefore adaptive in terms of their usability and automations according to the experience level of their users.

8. How do we align the various dashboard components to different decisions that must be made?
    - In a previous dashboard I associated students with the most negative comments on the use of Google Glasses, with their answers on open ended questions. Therefore, it is necessary to associate data analysis with additional data sets. For example, connecting survey results to comments.
9. How do we evaluate whether the dashboard contents meet the analysis needs?
    - There is no specific technique that I am aware of. I would ask potential end users to provide their views. This may include CEO, CTO, or random member of the technical team. I would focus on their views on the intended use of the dashboard. My personal inclination is for dashboards that are not over loaded – I prefer to design a dashboard that is more basic that may provide more insights. From my own experience overloaded dashboards seem to be used less.
10. How do we design the dashboard in order to maximise its usability by the intended users?
    - What is common across all dashboard users is that they wish to use the dashboard momentarily so they can get the needed information without spending too much time. It is a combination of good data sets, appropriate KPIs, good balance avoiding data overload.

Focus Group (NUP)

1. How do we determine the necessary data sources for the intended analysis and visualisation?

    For research

    - (SC) – First we check whether IRP needs to be involved to decide which data sources will be used. There are two committees an institutional one and a national one. The coordination is the responsibility of the project coordinator. If there is an involvement of the national committee the data collection will happen by the research project buy the national committee will ensure compliance.
    - (SE) – In order to collect data for children, their parents need to provide approval.
    - (SC) – The necessary data sources are identified according to the needs of the different projects.
    - (KZ) – Data are collected by certain databases according to the project needs.

    For learning analytics

    - (SC) – KPIs take under consideration student assessment in various types of examination. NUP observes the student progress in every 6-month term. This helps to identify whether something has happened during a specific term. The KPIs determine what data sources are used. Attendance is taken to ensure if there is a penalty for students based on the 85% threshold.
    - (SC) – A Moodle-based LMS is used, and data can be obtained from this platform. Data provided includes frequency of downloads and interaction with content.
    - (PC) – The LMS provides attendance percentag4e for each student per module.
    - (SC) – KPIs also include 4 days of attendance per week and that the hours of attendance need to be either AM or PM to allow attendance for students who work in parallel. Personal tutor policy requires compulsory evaluation of the personal tutor from the student sop the student is given an incentive to contribute to three evaluations in order to access grades. The evaluations include personal tutor, module content and module teaching staff.
2. How do we assess whether sufficient volume of data is available for the intended analysis?
    - (SC) – Evaluation results are considered only when 60% or above students participate in the evaluate. Q&A focuses on evaluations for the learning materials and the teaching staff. Now the threshold is almost 100% for student participation so staff can be given the results.
    - (SC) – The QA committee includes one academic from each department. A QA manager and two members from the executive board of the University.
3. How do we prepare the collected data for the analysis that must be performed?
    - (SC) – The results come directly from the software. The process transparency is regulated by the GDPR officer. The data results are connected directly from the QA officer. A Moodle plug in is used to collect the data after they are anonymised. The results are prepared after the anonymisation as a statistics summary for each module.
    - (KZ) – We need to get 360-degree feedback including evaluations not only from the students but also from peers.

- (EK) – Early evaluations were negative for my modules. The results were affected by the students' view of the module or the topics rather than the delivery. Data should also include open ended questions and rich information on how improvements could be implemented. Focus should not be only on the quantitative data – students should provide more detailed feedback. We should get more detailed information nit just averages.

4. How do we ensure that data analysis techniques are adapted to suit the specific domain?
    - (SC) – Student evaluations are only one of five evaluations we include. The five dimensions are supervisor's (the departmental representative to the QA committee who is a different person every 6 months) report for academics, course observation where the QA committee evaluates the content, institutional effectiveness (comparing the curriculum with other institutions) and the self-assessment. The data analysis is done across different dimensions. For example, modules of the same term are compared to identify any patterns of whether a specific module or module leader deviates from the pattern. The same module leader is evaluated across different terms as well. Q&A has adopted heuristics for evaluation after the validation from MUHEC of its programmes. The evaluation focuses whether there is a pattern or there is noise in the data.
5. How do we differentiate between the analysis of strategic and operational data?
    - (SC) – There is a big gap as we do not have access to the finances and the way recruitment takes place. The strategic aspects are managed centrally and are not connected to the academic part. Due to GDPR we cannot correlate certain data sets. There is the need to tune data sets to be compatible with the sector's needs. The scientific and the business boards decide together with the Senate.
6. How do we confirm the analysis tasks required to produce concrete findings that can be visualised?
    - (SC) – The scientific and the business boards decide together with the Senate.
7. How do we select appropriate visualisations for the analysis conducted and the data sets available?
    - (SC) – The scientific and the business boards decide together with the Senate and the QA. There is not a specific mechanism to be used. The responsibility would like with QA and Senate. Academics can produce visualisations of the assessment results that are discussed at departmental level. As NUP is a private institution, there is an important aspect associated with the student finances and the recruitment of students. When access to resources is disallowed following a financial hold then the academic department is not involved in any financial agreements.
8. How do we align the various dashboard components to different decisions that must be made?
    - Same as above
9. How do we evaluate whether the dashboard contents meet the analysis needs?
    - Same as above
10. How do we design the dashboard in order to maximise its usability by the intended users?
    - Same as above

Appendix D – Ethics

As part of the ethics approval process the following documentation was submitted leading to the approval from the ethics committee:

- Middlesex University Data Protection Checklist and Declaration for Researchers.
- Participant Information Sheet (PIS)
- Consent Form

Middlesex University Data Protection Checklist and Declaration for Researchers

REC no:______

Project title:	**A framework for strategic planning of data analytics in the educational sector**	
PI/Supervisor:	Xiaohong Gao	**Date:**
04/01/2022		

There are now **7 Data Protection Principles**, which states that information must be:

1. Fairly and lawfully processed;
2. Processed for specified and lawful purposes;
3. Adequate, relevant and not excessive;
4. Accurate and kept up date where necessary;
5. Not kept for longer than is necessary;
6. Kept secure;
7. Necessary to actively demonstrate compliance with all of the above principles

processing in accordance with individuals' rights and not transferring to countries without adequate protection are no long principles but have specific

Article 89 of the GDPR and Schedule 2 Part 6 of the Data Protection Act 2018 (DPA) provides exemption to some of the data protection principles and individual rights for processing personal data for 'research purposes' including statistical or historical purposes. These are noted in the checklist below.

For guidance on the Data Protection Act for Social Research please see the MRSGuidelines for Social Research, April 2013 which can be accessed using the following link: https://www.mrs.org.uk/standards/legislation/tab/data_protection

Guidance on large data sets can be found at the Information Commissioner's Office website – BBig data, artificial intelligence, machine learning and data protection September 2017.J https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

You may also find JISC Legal Information on Data Protection and Research Data Questions and Answers, (last updated July 2018)helpful. http://www.jisclegal.ac.uk/guides/data-protection

Note: Personal data which is anonymised[i], permanently, is exempt from compliance with the DPA and registration process. See endnotes for further details.

Conditions which must be met for a research exemption to apply under Schedule 2 Part 6 of the DPA 2018	**Please indicate**	
1. The information is being used **exclusively for research** purposes	Agree ✓	Disagree
2. The information is not being used to **support measures or decisions relating to any identifiable living individuals**	Agree ✓	Disagree
3. The data[ii] is not being used in a way that will **cause** or **is likely to cause, substantial damage or substantial distress** to any individuals or very small groups *If you 'Disagree' please provide details why an adverse effect is justified:*	Agree ✓	Disagree
4. The results of the research, or any resulting statistics, will **not be made available in a form that identify individuals** *If you 'Disagree' please provide details why identification is intended:*	Agree ✓	Disagree

If you 'Agree' to all of the above conditions then the use of personal data is exempt from the Second Principle and the Fifth Principle, but you must comply with First, Third, Fourth, and Principles of the DPA alongside protecting certain individual rights and not transferring to countries without adequate protection. If a research exemption does not apply then you must ALSO comply with the Second and Fifth Principles of the DPA			
First Principle: Fairly and lawfully processed			
5. Will you have appropriate **informed consent**[iii] secured from participants for the **personal data**[iv] that you will be analysing? i.e., **inform participants** of a) What you will **do** with the data? b) Who will **hold** the data? (Usually MU, unless a third party is involved) c) Who will have **access** to the data or receive copies of it? (e.g., for **secondary data sets**, are you sure that appropriate consent was secured from participants when the data was collected?) *If 'no' please provide details and any further actions to be taken:*	Yes✓	No	N/A
6. If you plan to analyse **sensitive (known as special categories of personal data under the new legislation) personal data**[v], have you obtained **data subjects'**[vi] **explicit informed consent**[vii] (as opposed to **implied consent**[viii])? *If 'no' please provide details:*	Yes	No	N/A ✓
7. If you do not have the data subjects' explicit consent to process their data, are you **satisfied that it is in the best interests of the data subject** to collect and retain the sensitive data? *Please provide details:*	Yes	No	N/A ✓
8. If you are processing[ix] personal data about **younger individuals or those with reduced capacity**, have you put a process in place to obtain consent from parents, guardians or legal representatives, if appropriate? *Please provide details:*	Yes	No	N/A ✓
9. Will you have a process for managing **withdrawal of consent**? *If 'no' please provide details:*	Yes	No	N/A ✓
10. Will it be necessary or desirable to work with external organisations e.g., charities, research organisations etc. acting as **a third party** i.e., directly providing a service for us or on our behalf that involves them accessing, collecting or otherwise processing personal data the third party will become a data processor under the DPA? *If 'yes' then you will be using a third party as a data processor you must take advice from the Middlesex University Data Protection Officer about the planned contractual arrangements and security measures.*	Yes	No✓	N/A
11. Have you written an appropriate privacy notice to provide to individuals at the point you collect their personal data? *(Please see '**Guide to Research Privacy Notices'**)*	Yes✓	No	N/A
Second Principle: Processed for limited purposes			
Will personal data be obtained only for **one or more specified and lawful purposes**, and not further processed in any manner incompaible with the purpose(s)? (Research data subjects should be informed of any new data processing purposes, the identity of the **Data Controller**[x] and any disclosures that may be made.) **Research Exemption Note** (GDPR Article 89): Personal data can be processed for research purposes other than for which they were originaly obtained if that processing does not take fmeasures or decisions with respect to the particular data subjects (unless necessary for approved medical research); and no likelihood of substantial damage or substantial distress to any data subjects That data may also be held indefinitely.	Yes	No	N/A
Third Principle: Adequate, relevant and not excessive			
12. Will you only collect data that is **necessary** for the research? *If 'no' please provide details and any further actions to be taken:*	Yes✓	No	N/A
Fourth Principle: Accurate and where necessary, kept up to date			

13. Will you take reasonable measures to ensure that the information is **accurate**, **kept up-to-date** and **corrected** if required? *If 'no' please provide details:*	Yes✓	No	N/A
Fifth Principle: Not kept for longer than is necessary			
14. Will you check **how long data legally must be kept** and **routinely destroy** data that is past its retention date and **archive data** that needs to be kept? **Research Exemption Note** (section 33(3)): Personal data processed for research purposes can be kept indefinitely.	Yes	No	N/A
Chapter 3 GDPR: Processed in accordance with individuals' rights under the DPA[xi]			
15. If you are intending to publish information, which could **identify individuals**, have you **made them aware** of this when gaining their informed consent? *If 'no' please provide details:*	Yes✓	No	N/A
16. Will you allow **access to all personal data** held about a data subject if an individual makes this request? **Research Exemption Note** (Schedule 2 Part 6 DPA): Where the results of processing personal data for research purposes do not identify a data subject, that data subject does not have a right of access to that data.	Yes	No✓	N/A
17. Will you ensure that all researchers who have access to **personal data** understand that it **must not be provided to any unauthorised person or third party** (e.g., family members etc.) unless consent has been given?	Yes✓	No	N/A
Sixth Principle: Kept secure			
18. Will you ensure that personal data will be **stored in locked cabinets, cupboards, drawers** etc. (regardless of whether data is on paper, audio-visual recordings, CDs, USBs, etc.)?	Yes✓	No	N/A
19. Will you ensure that if personal data is to be **stored electronically** it will only be kept on **encrypted devices**?	Yes✓	No	N/A
20. Will you ensure that individuals who have access to the personal data are aware that **email is not a secure method** of communication and should **not be used for transferring the data**?	Yes✓	No	N/A
21. Will you ensure that **disposal of personal data** will be via **confidential waste services** or in the case of **electronic media and hardware** should be **destroyed** in line with Middlesex University guidelines and procedures?	Yes✓	No	N/A
Chapter 5 GDPR: Not transferred to other countries without adequate protection			
22. Will you **ensure that personal data is not transferred outside the EEA** unless one of the following applies? i. The country you are transferring the data to has been approved as providing adequate protection ii. You have obtained explicit informed consent from the individual(s) iii. You have a contract in place with the recipient of the data, which states the appropriate data protection requirements. iv. You have completely anonymised the data.	Yes✓	No	N/A

Declaration

I confirm that I have noted the main points related to Data Protection GDPR for researchers and I understand my responsibilities for data protection as a researcher.

The following video summaries the main points: https://www.youtube.com/watch?v=hBzqELMe2nY

Please refer to the following documents that can be located with within the MORE Help/Templates for further information:

1. Specific Issues to Consider regarding GDPR
2. GDPR Guidance for Researchers: Summary of main points of the GDPR related to research
3. Personal and Sensitive Personal Data – ICO definitions relating to GDPR on personal and sensitive data

4. Criminal Offence Data Protection Requirements
5. Data Protection Policy
6. MU Data Management Plan Policy
7. Data Management Plan example
8. Personal data flowchart – check is you are collecting personal data
9. Anonymous and Pseudonymous Data – definitions, guide and further references on how to anonymise data
10. ICO website address for full details of GDPR: https://ico.org.uk

Print Name: Ariadni Tsiakara____

Signature: _____ Ariadni Tsiakara ____

Date: _04/01/2022____

Any concerns in relation to compliance with the DPA should be discussed with the Middlesex University Data Protection Officer.

MIDDLESEX UNIVERSITY

Participant Information Sheet (PIS)

More Than Minimal Risk or High-Risk Projects

(Must be used with a Consent Form that is signed by the participant and retained by the researcher)

Participant ID Code:……………………………………………

SECTION 1

1. Project/Study title

A framework for strategic planning of data analytics in the educational sector.

2. Invitation paragraph

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

3. What is the purpose of the study?

The scope of this research is to investigate the different factors that affect the successful deployment of data analytics in educational contexts focusing both on strategic and operational aspects of academia. The research study attempts to identify those elements necessary for introducing data analytics practices across an institution. The main contribution of the research is a framework that models the data collection, analysis and visualisation in higher education. The specific contribution to the field comes in the form of generic guidelines for strategic planning of HEI data analytics projects, combined with specific guidelines for staff involved in the deployment of data analytics to support certain institutional operations.

The research is based on a mixed method approach that combines grounded theory in the form of extensive literature review, state-of-the-art investigation and case study analysis, as well as a combination of qualitative and quantitative data collection.

4. Why have I been chosen?

It is important that we assess as many participants as possible, and you may have indicated that you are interested in taking part in this study. The reason why you have been selected as an interviewee is because as a teacher in HEI you have the knowledge and experience to identify those elements necessary for introducing data analytics practices across an institution.
Finally, there will be 8 interviewees who are over the age of 18 that will be studied in this research.

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. If you do decide to withdraw from the study then please inform the researcher as soon as possible, and they will facilitate your withdrawal. If, for any reason, you wish to withdraw your data please contact the researcher within a month of your participation. After this date it may not be possible to withdraw your individual data as the results may have already been published. However, as all data are anonymised, your individual data will not be identifiable in any way.

A decision to withdraw at any time, or a decision not to take part, will not affect you.

6. What will I have to do?

- Your involvement will only be required for the interview.
- The interview will take place through an online meeting via Zoom.
- There will be a Zoom meeting code sent to you and date/time arranged for the interview.
- There will be interview questions that will be emailed to you so you can understand what questions would be asked.
- There will be questions relating to the research that would be asked.
- The duration of the interview is expected to be 30 to 60 minutes.
- The interview will be recorded in Zoom.

Please note that in order to ensure quality assurance and equity this project may be selected for audit by a designated member of the committee. This means that the designated member can request to see signed consent forms. However, if this is the case your signed consent form will only be accessed by the designated auditor or member of the audit team.

7. Will I have to provide any bodily samples (i.e., blood/saliva/urine)?

No

8. What are the possible disadvantages and risks of taking part?

There is no known risk in participating in this project.

Appropriate risk assessments for all procedures have been conducted, and will be followed throughout the duration of the study.

9. What are the possible benefits of taking part?

We hope that participating in the study will help you. However, this cannot be guaranteed. The information we get from this study may help us to investigate the different factors that affect the successful deployment of data analytics in educational contexts focusing both on strategic and operational aspects of academia.

There is no intended benefit to the participant from taking part in the study.

10. Will my taking part in this study be kept confidential?

The research team has put a number of procedures in place to protect the confidentiality of participants. You will be allocated a participant code that will always be used to identify any data you provide. Your name or other personal details will not be associated with your data, for example, the consent form that you sign will be kept separate from your data. All paper records will be stored in a locked filing cabinet, accessible only to the research team, and all electronic data will be stored on a password protected computer. All information you provide will be treated in accordance with the UK Data Protection Act.

11. What will happen to the results of the research study?

The results of the research study will be used as part of a Postgraduate dissertation. The results may also be presented at conferences or in journal articles. However, the data will only be used by members of the research team and at no point will your personal information or data be revealed.

12. Who has reviewed the study?

The study has received full ethical clearance from the Research ethics committee who reviewed the study. The committee is the Computer Science REC.

13. Contact for further information

If you require further information, have any questions or would like to withdraw your data then please contact:

Ariadni Tsiakara, A.Tsiakara@mdx.ac.uk
Xiaohong Gao, X.Gao@mdx.ac.uk
Franco Raimondi, F.Raimondi@mdx.ac.uk

Thank you for agreeing to take part in this study.
You (the participant) should keep this “Participant Information with Consent” sheet since it contains important information and the research teams contact details.

SECTION 2

Middlesex University Guide to Research Privacy Notices

Privacy notices need to be presented whenever data is collected and should be understandable and accessible. Privacy notices must explain the type and source of data that will be processed. They will also set out the processing purpose, data retention schedules and data sharing. Privacy notices must include details of the subject's rights and who the subject can complain to.
The following example may be used and completed for your research purposes.

Middlesex University Privacy Notice for Research Participants

The General Data Protection Regulation (GDPR) protects the rights of individuals by setting out certain rules as to what organisation can and cannot do with information about people. A key element to this is the principle to process individuals' data lawfully and fairly. This means we need to provide information on how we process personal data.
The University takes its obligation under the GDPR very seriously and will always ensure personal data is collected, handled, stored and shared in a secure manner. The University's Data Protection Policy can be accessed here: https://www.mdx.ac.uk/__data/assets/pdf_file/0023/471326/Data-Protection-Policy-GPS4-v2.4.pdf.
The following statements will outline what personal data we collect, how we use it and who we share it with. It will also provide guidance on your individual rights and how to make a complaint to the Information Commissioner's Officer (ICO), the regulator for data protection in the UK.

Why are we collecting your personal data?

As a university we undertake research as part of our function and in our capacity as a teaching and research institution to advance education and learning. The specific purpose for data collection on this occasion is to investigate the different factors that affect the successful deployment of data analytics in educational contexts focusing both on strategic and operational aspects of academia. The research study attempts to identify those elements necessary for introducing data analytics practices across an institution. The main contribution of the research is a framework that models the data collection, analysis and visualisation in higher education. The specific contribution to the field comes in the form of generic guidelines for strategic planning of HEI data analytics projects, combined with specific guidelines for staff involved in the deployment of data analytics to support certain institutional operations.
The legal basis for processing your personal data under GDPR on this occasion is Article 6(1a) consent of the data subject.

Transferring data outside Europe

In the majority of instances your data will be processed by Middlesex University researchers only or in collaboration with researchers at other UK or European institutions so will stay inside the EU and be protected by the requirements of the GDPR.
In any instances in which your data might be used as part of a collaboration with researchers based outside the EU all the necessary safeguards that are required under the GDPR for transferring data outside of the EU will be put in place. You will be informed if this is relevant for the specific study you are a participant of.

Your rights under data protection

Under the GDPR and the DPA you have the following rights:

- to obtain access to, and copies of, the personal data that we hold about you;
- to require that we cease processing your personal data if the processing is causing you damage or distress;

- to require us to correct the personal data we hold about you if it is incorrect;
- to require us to erase your personal data;
- to require us to restrict our data processing activities;
- to receive from us the personal data we hold about you which you have provided to us, in a reasonable format specified by you, including for the purpose of you transmitting that personal data to another data controller;
- to object, on grounds relating to your particular situation, to any of our particular processing activities where you feel this has a disproportionate impact on your rights.

Where Personal Information is processed as part of a research project, the extent to which these rights apply varies under the GDPR and the DPA. In particular, your rights to access, change, or move your information may be limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we may not be able to remove the information that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible. The Participant Information Sheet will detail up to what point in the study data can be withdrawn.

If you submit a data protection rights request to the University, you will be informed of the decision within one month. If it is considered necessary to refuse to comply with any of your data protection rights, you also have the right to complain about our decision to the UK supervisory authority for data protection, the Information Commissioner's Office.

None of the above precludes your right to withdraw consent from participating in the research study at any time.

Collecting and using personal data

The interviews will be used as an evaluation of the key aspects associated with each of the C.A.V. phases. Therefore, the questionnaire is organised accordingly in three sections. A fourth section dedicated to the administration tasks of educational analytics is covered by parallel discussion during each section aiming to find out how the entire process is taking place in the participants' institution.

Data sharing

Your information will usually be shared within the research team conducting the project you are participating in, mainly so that they can identify you as a participant and contact you about the research project.

Responsible members of the University may also be given access to personal data used in a research project for monitoring purposes and/or to carry out an audit of the study to ensure that the research is complying with applicable regulations.

Individuals from regulatory authorities (people who check that we are carrying out the study correctly) may require access to your records. All of these people have a duty to keep your information, as a research participant, strictly confidential.

If we are working with other organisations and information is shared about you, we will inform you in the Participant Information Sheet. Information shared will be on a 'need to know' basis relative to achieving the research project's objectives, and with all appropriate safeguards in place to ensure the security of your information.

Storage and security

The University takes a robust approach to protecting the information it holds with dedicated storage areas for research data with controlled access.

Alongside these technical measures there are comprehensive and effective policies and processes in place to ensure that users and administrators of university information are aware of their obligations and responsibilities for the data they have access to. By default, people are only granted access to the information they require to perform their duties. Training is provided to new staff joining the University and existing staff have training and expert advice available if needed.

Retention

Under the GDPR and DPA personal data collected for research purposes can be kept indefinitely, providing there is no impact to you outside the parameters of the study you have consented to take part in.

Having stated the above, the length of time for which we keep your data will depend on a number of factors including the importance of the data, the funding requirements, the nature of the study, and the requirements of the publisher. Details will be given in the information sheet for each project.

Contact us

The Principal Investigator leading this research is Ariadni Tsiakara
a.tsiakara@mdx.ac.uk

The University's official contact details are:
Data Protection Officer
Middlesex University
The Burroughs
London
NW4 4BT
Tel: +44 (0)20 8411 5555
Email: dpaofficer@mdx.ac.uk

Thank you for agreeing to take part in this study. You (the participant) should keep this "Participant Information with Consent" sheet since it contains important information and the research teams contact details.

Middlesex University London

Version Number…

Participant Identification Number:

CONSENT FORM

Title of Project:

Name of Researcher:

Please initial box

1. I confirm that I have read and understand the information sheet dated ....……………..….for the above study and have had the opportunity to ask questions. [1]

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without penalty . [2]

3. I agree that this form that bears my name and signature may be seen by a designated auditor. [3]

4. I agree that my non-identifiable research data may be stored in National Archives and be used anonymously by others for future research. I am assured that the confidentiality of my data will be upheld through the removal of any personal identifiers. [4]

Delete 5 and or 6 if not applicable:

5. I understand that sections of any of my medical notes may be looked at by responsible individuals from [company name] or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. [5]

6. I understand that my interview may be taped and subsequently transcribed. [6]

7. I agree to take part in the above study. [7]

Name of participant	Date	Signature
Name of person taking consent (if different from researcher)	Date	Signature

Researcher Date Signature

1 copy for participant; 1 copy for researcher;

Remember that a signed consent form is not required for an anonymous questionnaire, instead the following statement is recommended to be included on the survey questionnaire:

'Completion of this questionnaire is deemed to be your consent to take part in this research.'

[i] **Anonymous data** is prepared from personal information but from which, an individual cannot be identified by the person holding the data. **Anonymisation** is a **permanent** process. Personal data must be treated so that it cannot be processed in such a way as to link the data to a specific individual (e.g., using an identifier). Coded data is not anonymised and therefore not exempt from compliance or registration.

[ii] **Data** covers information that is held on computer, or to be held on computer to be processed. Data is also information recorded on paper if you intend to put it on computer.

[iii] **Informed consent** means providing participants with a clear explanation of the research project in order for them to give informed consent regarding the use of their data. Individuals should be informed that their involvement is voluntary and that they have the right to refuse or withdraw at any time without any negative consequences.

Informed refers to the following information being provided to the data subject/participant:

- i) Who you are, the organisation you work for and who else is involved in the research project or using the data.
- ii) What data will be collected and how.
- iii) Who will hold the data, control access to the data and how it will be stored and kept safe and whether it will be transferred to a third party.
- iv) How the data will be used.
- v) How long it will be kept and what will happen to it at the end of the project.
- vi) Risks related to any aspects of the research project and data, benefits of the research project and any alternatives.

[iv] **Personal data** (sometimes referred to as personal information) means data which relate to a living individual who can be identified from those data whether in personal or family life, business or profession, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller. The data is of biographical significance to the individual and impacts an individual in a personal, family, business or professional capacity. It includes any expression of opinion about the individual and/or statements of fact.

[v] **Sensitive/special categories of personal data** means personal data consisting of information about the **data subjects'**,

1. Racial or ethnic origin,
2. Political opinions,
3. Religious beliefs or other beliefs of a similar nature,
4. Trade union membership
5. Physical or mental health or condition,
6. Sexual life,
7. Genetic or biometric information

Criminal matters are technically now not part of the list of special categories of data and have their own section in the legislation but for practical purposes it should be treated the same as the above.

Also personal financial details are vulnerable to identity fraud and should be handled confidentially and securely although not defined as sensitive under the Act.

[vi] **Data subject** is a living individual to whom the personal data relates. If an individual has died or their details have been anonymised then their data does not fall within the Act. Personal data relating to deceased individuals may still be owed a duty of confidentiality.

[vii] **Explicit informed consent** is where an individual actively opts to participate.

[viii] **Implied consent** is where an individual must inform the researcher that they wish to opt out.

[ix] **Processing** of personal information includes collecting, using, storing, destroying and disclosing information.

[x] **Data controller** is the person who either alone or jointly on in common with other persons determines the purposes for which, and the manner in which, any personal data are or are to be, processed. The fact that an individual or institution holds or processes personal data does not make them a Data Controller if they do not determine the purpose and manner of that holding or processing. (This is probably one of the most widely misunderstood definitions of the Act.) In most cases the Data Controller will be Middlesex University, however further guidance and clarification can be sought from the Middlesex University Data Protection Officer.

[xi] **Data subject rights** include:
The GDPR provides the following rights for individuals:

The right to be informed
The right of access
The right to rectification
The right to erasure
The right to restrict processing
The right to data portability
The right to object
Rights in relation to automated decision making and profiling.

. Access means an individual can make a subject access request for all copies of all personal data held about them and ask to whom it has been disclosed. An individual potentially has access to personal comments written about them. It is an offence to deliberately edit or destroy data once a subject access request has been received. Third parties do not generally have access to subject data unless an exemption applies or there is overriding public interest. There may be limited third party access to ordinary personal data relating to a business or professional capacity in the public interest through the Freedom of Information Act.