DOCTOR OF PHILOSOPHY

Learning Analytics Integrating Student Attendance Data

Gray, Cameron

Award date:
2019

Awarding institution:
Bangor University

Bangor University

DOCTOR OF PHILOSOPHY

Learning Analytics Integrating Student Attendance Data

Gray, Cameron

Award date:
2019

Awarding institution:
Bangor University

PRIFYSGOL BANGOR UNIVERSITY
School of Computer Science and Electronic Engineering
College of Environmental Sciences and Engineering

Learning Analytics Integrating Student Attendance Data

Cameron C. Gray

Submitted in partial satisfaction of the requirements for the
Degree of Doctor of Philosophy
in Computer Science

Supervisor Dr. Dave Perkins

March, 2019

Acknowledgements

> *Dont be satisfied with stories, how things have gone with others. Unfold your own myth.*
>
> — **Jalāl ad-Dīn Muhammad Rūmī**

While it is true that an individual can make quantum leaps in any field of endeavour, most often the transformational cannot be achieved without a plethora of supporting characters. This is true of my work as well. While I cannot hope to name every individual that has contributed support, guidance, and ideas; I would like to single out the following for their efforts.

Firstly; my wife Jill, for putting up with the innumerable rants about students, research, and academia. In addition to her undying support, her inner frustrated English teacher carefully pointed out grammar flaws as if marking my very soul. This work would not be as presentable or flowing without her help. I miss you greatly and I'm painfully sorry you didn't get to see the end of this journey.

Secondly, Dr. Dave Perkins for his guidance through the twisted world that is research and the Ph.D. process. His reassurance was a critical component to finishing this work. His input on teaching matters has also made me a better lecturer, and helped inform and improve this work.

Next, the entire staff of the School of Computer Science. Whether as a preparation during my undergraduate years, or a shoulder to cry on during this endeavour. I would not have been as successful were it not for your guidance, tutelage, and input. Thank you all.

Lastly, I would like to personally thank some of the other postgraduate students within the Schools of Computer Science and Electronic Engineering. James Jackson, for listening to endless rants and appreciating the elder view of things. Shaun Preston, for providing distraction when needed and sensible conversation otherwise. James Monks as an example of how to both be academic (which I would like to follow) and a party animal (which I won't). Paria Yousefi, your ever- optimistic and fun outlook provided much needed lifts on occasions when research did not agree with me. Finally, Will Taplin for being the comedic relief that every good story needs. You all will be remembered fondly. I wish you the best of luck with your future efforts and would be pleased to collaborate with any of you - as long as I am not the one to do the publication part...

Abstract

With UK Higher Education (HE) ever more scrutinised with the Teaching Excellence Framework, changes to finance and media attention student welfare and retention is becoming ever more important. Institutions are turning to technological methods to assist wherever possible. Use of Learning Analytics (LA) systems, as a result, is booming. Institutions are deploying these systems to exploit the wealth of information that they hold on their students. LA systems have historically focused on interactions between a student and their course in the form of results or participation within a Virtual Learning Environment. However, their focus tends to be on analysing outcomes based on performance, with other factors occasionally mixed in. If early identification is to the be goal of an analytics system, new sources of data must be added to the models. These new data sources will need to be timely and robust. Almost all institutions, as they have a visa monitoring requirement, monitor student attendance at timetabled and other events. This attendance data is rarely included in LA models.

This thesis shows how a simple synthetic metric, the Bangor Engagement Metric (BEM), can be used in conjunction with standard Data Science and ML techniques can produce a powerful predictive model as early as Week 4 of Semester 1 (~end of October each year). It proves, through a series of experiments, that a single algorithm is suited to this unbalanced class classification problem. The model is then proven, against both past and future unseen data, to produce prediction accuracies in excess of 93% at Bangor University - reaching as high as 97.33%. These trials go on to show that the only improvement to this level of accuracy is to utilise the attendance data from all of Semester 1. In addition, the results provide a

recommendation that models need to be re-trained every two years to avoid any disproportional effects from a single cohort.

As part of this project a new tool to visualise and communicate student achievement was developed. This tool uses contemporary Information Visualisation techniques to provide both a macro view of the entire cohort, down to the micro view of a single module for one student. One of the views provides an overview of the student's entire academic career by semester. This work has produced a set of sixteen descriptions for the unique (after accounting for time and scale effects) possibilities in this view. The author has termed them 'Degree Pictures'. While the form of the intervention is outside of the scope of this work, these standardised forms can allow educators to develop best practice responses to any pattern that starts towards an undesirable outcome.

During testing of this model, an anomaly was observed where students were highlighted for poor attendance/engagement but were not identified as potentially failing in the model. The attendance data for these students showed a common feature, the disengagement began at week 5. Using statistics tools, this work was able to demonstrate that a significant shift occurred within the student body between week 5 and 8. This timing coincides with the timing of Reading Weeks for that semester. The work goes on to infer a structure to student cohorts which suggest that intervention with certain groups would be more effective than others. This insight shows that educators need to be mindful about introducing any disruption into the student experience, as it can be a trigger for disengagement.

Contents

List of Figures

List of Tables

Nomenclature

The following acronyms and definitions will be used throughout this document. Where these specific terms are used, it should be understood the author means these definitions instead of any common or other definition.

Acronyms

AI	Artificial Intelligence
BEM	Bangor Engagement Metric
CELT	Centre for Enhancement of Learning and Teaching
ER	Engagement Ratio
GDPR	General Data Protection Regulations
HCI	Human-Computer Interaction
HE	Higher Education
HEFCE	Higher Education Funding Council for England
LA	Learning Analytics
LOO-CV	Leave-One-Out Cross Validation
LMS	Learning Management System
ML	Machine Learning
OfS	Office for Students
SIS	Student Information System
SoLAR	Society for Learning Analytics Research
SpLD	Specific Learning Difficulty
TEF	Teaching Excellence Framework
VLE	Virtual Learning Environment

Definitions

Academic Standing Code

A student's Academic Standing Code is a description of the outcome of the current academic year. These are commonplace in the UK but may vary by institution. There are five values used in most situations at Bangor University; PA - Pass All, FC - Fail Conditional, FN - Fail (No-re-registration), RY - Repeat Year, PS - Passed Supplementary. A student that completes a sufficient number of credits to a passing standard, but fails others is given the opportunity to complete Supplementary Assessment. That student is described as conditionally failed until the outcome of the supplementary assessment is known. At that point the student's standing would be changed to either PS if they did achieve a pass or FN otherwise. *Used on pp. 65, 73, 77, 81, 86, 87.*

Big Data

Data-sets and processes adhering to the 'Four Vs' description. It must be too complex or large to be processed by traditional means. See Section 2.1.2 for a discussion and fuller definition. *Used on p. 14.*

Module

A module is one of the separate parts of a programme taught at a college or university. Each module focuses on a separate topic or subject. In UK Higher Education each module runs for either one or two semesters. This is often known as a 'Course' outside of the UK. *Used on pp. 6, 33, 37, 101, 102, 105, 116, 124.*

Programme

A programme is a course of study, with defined learning outcomes and an academic award for completion. A programme is composed of modules, and typically lasts 3 years in the UK for undergraduates. *Used on pp. 33, 37, 65, 101, 124, 137.*

Supplementary Assessment

When a student fails to meet the pass criteria for a given module or their programme overall; they will need to complete additional work before being able to continue their studies or graduate. This extra work is termed supplementary assessment, previously termed resits. This is akin to summer school in the US education system. *Used on pp. 62, 102, 103, 105.*

Symbols

C (Classification Tasks) The number of classes within a given data-set. *Used on p. 18.*

FN (Classification Tasks) The number of instances misclassified as a False Negative. *Used on p. 22.*

FP (Classification Tasks) The number of instances misclassified as a False Positive. *Used on p. 22.*

N (Classification Tasks) The number of features contained in a given data-set. *Used on p. 19.*

TN (Classification Tasks) The number of instances correctly classified as a True Negative. *Used on p. 22.*

TP (Classification Tasks) The number of instances correctly classified as a True Positive. *Used on p. 22.*

k (Bangor Engagement Metric) The number of elements/sessions in the set a student has attended. *Used on pp. 67, 75.*

n (Classification Tasks) The number of instances in a given data-set. *Used on p. 18.*

z (Bangor Engagement Metric) The full set of observations with k elements in the set. This set is encoded with value 1 for attended and 0 for not attended. *Used on p. 67.*

Chapter 1

Introduction

> *It's not worth doing something unless you were doing something that someone, somewhere, would much rather you weren't doing.*
>
> — **Terry Pratchett**

The field of Learning Analytics (LA) has struggled to clearly distance itself from many other forms of analytics tools, strategies, processes, theory, and practice. This struggle remains today; Learning Analytics could be just another form of Big Data endeavour, a specialist form of data mining (often referred to as Educational Data Mining), or something else entirely [210]. To confuse matters, practitioners commonly interchange terminology when referring to the field. "Learning Analytics" and "Learner Analytics" are used synonymously.

Both Ferguson [99] and Clow [66] have attempted to form cohesive definitions of the field, its challenges, drivers, and goals. The Society for Learning Analytics Research (SoLAR) adopted the following definition [281]: "Learning Analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs." This definition has held, although it is not universally accepted. Greller and Drachsler [125] proposed a more holistic definition by defining six 'critical dimensions'. They argue that a combination of Stakeholders, Objectives, Data, Internal Limitations, External Constraints, and Instruments capture the required details to define the qualities, approaches, and requirements to make Learning Analytics a benefit to those involved.

1.1 Motivation

However LA is defined, we must never lose sight of the fact that this endeavour is intrinsically linked to, and for the benefit of the students [113]. Therefore, any motivation should be to better the education, services, support, and other offerings to the student.

There are numerous LA projects and tools employed around the world to significant effect. Each has its own specialist goals and influences. This work is no different. 2017 was a watershed year for the UK Higher Education (HE) sector. Institutions would no longer be solely measured by their research impact and output but teaching as well. The Higher Education Funding Council for England (HEFCE) developed the Teaching Excellence Framework (TEF) to further this aim. As the Office for Students (OfS)[1] has formally re-adopted the TEF [31], institutions must adapt to the (potentially) shifted focal points within their teaching provision.

In the UK HE sector, retention is one of the topics most often discussed when evaluating institutions' teaching. The TEF includes, as part of the metrics, continuation rates. A continuing student is defined as one that enrols on a HE course and progresses to the second year of study. LA is a potential solution, using the wealth of information that institutions hold about their students to provide critical insight. Once adverse patterns are observed, educators can plan and execute appropriate interventions. These interventions should lead to a better outcome for the student. As a by-product, the institution potentially improves TEF scores. As the concept of LA is accepted into an institution's culture, there are significantly more in-depth analytics that could be implemented.

[1] OfS assumed the regulation and supervision of England's HE provisions on April 1st 2018.

Teaching, almost by its very nature, requires the teacher to adapt intuitively to their students. This kind of adaptation is honed over years of experience and through the sharing of experiences by colleagues. The author believes that data, processed by LA, could provide objective guidance to the profession. Explicit assistance could reduce the variability of student experience and feedback. Systems can also reduce the chance that weaker students are not identified until it is too late to help.

1.2 Problem Statement

At Bangor University, as at other institutions that have not adopted LA yet (current as of ~2018), students' academic progression is left to individual tutors to monitor. This introduces a divide between engaged tutors and not, experienced tutors and not, as well as capable tutors and not. Most often, backed by studies [197, 236, 314], early intervention with vulnerable (academically or otherwise) students will produce better outcomes.

The fundamental problem this work aims to solve is the identification of academically vulnerable students promptly, and arming educators with objective insights to best address any issues. In this case, 'early identification' is defined as providing sufficient time to allow intervention with the student before the situation becomes unrecoverable. Along with the time constraint, the work should not fail to identify an at-risk student but may suggest intervention with students that may have succeeded without intercession.

The significant difference between this work and those that have gone before is that our focus is on the earliest possible accurate identification of candidate students. When using specific data-sets, such as achievement (marks), there is an intrinsic delay in obtaining sufficient data points to analyse. There are possibilities of using historical information, but this would not account for cohortal effects, nor changes in regime during the students' prior education.

Any data-set must be rich enough to differentiate student groups and factor in as many aspects of the social, economic, and academic experience of a student. It must also be timely — available for analysis as soon after it is generated as possible. The data must also be indexable by time, so that cause(s) and effect(s) can be linked as well as related to each other.

Once a data-set has been identified, appropriate analysis methods must be sought to convert the raw data into information, to be useful to educators. This necessity exists with almost every other data-driven system and was first highlighted by Ackoff [3]. This information must then be reported in an appropriate form to allow the consumer, in this case educators, to best make decisions on the interventions to address any issues identified.

In order to be of use to educators, the findings will need to be communicated effectively in a manner they are comfortable with. This will rely on a fusion of traditional reporting, interactive tools, and visualisations to best put the point across. Whatever the final solution; it would need to be tested, both for accuracy and for usability/acceptance, and studies conducted to determine what, if any, effect the extra information has on the students concerned.

1.3 Hypothesis

The testable hypothesis for this work is as follows:

> *A system can be developed, using attendance data, to effectively identify students that may suffer poor outcomes early enough for tutors to intervene.*

The completed system may be comprised of multiple components, rather than a single monolithic piece of software. This approach would allow individual aspects to follow appropriate individual methodologies. It will also produce a modular architecture, allowing replacements and upgrades without the need to disrupt the entire system.

This hypothesis will require both quantitative and qualitative evaluation to confirm that the system satisfies these requirements. The component research questions that arise are:

1. What elements of attendance data can be used to predict student outcomes?
2. Can a suitably accurate predictive model be built and maintained from this data?
3. How close to the start of an academic year can the model yield sufficiently high accuracy rates (e.g. > 90%)?
4. What results need to be conveyed to educators to best place them to intervene and in what form do they need to be presented?
5. Can this model be used to provide additional insight into student attendance patterns?
6. How can the findings be used to aid educators in deciding the best course of any subsequent action?

1.4 Aim & Objectives

The overarching aim for this work is to deploy Learning Analytics (LA) methods using student attendance data to predict which students were at risk of failure. This aim is designed to contribute to the solving of undergraduate retention issues, as well as targeting help to those students that would benefit most in a time of shrinking service budgets.

In order to achieve this aim, the following objectives must be met: -

1. Gain access to, and appropriately clean a data-set which includes student attendance information.
2. Design, execute, and refine a series of Machine Learning (ML) experiments to determine which facets of the data-set hold the most predictive power.
3. Devise, or discover, and validate the most appropriate metric to describe student attendance to be included in a model.

4. Construct and test a model using previous outputs to predict student achievement.
5. In the process of testing the model, discover any non-trivial and previously unknown patterns in the data or meta-data.
6. Utilise both ML and Data Science techniques to derive insight from these new patterns.
7. Devise, refine, and evaluate a method to present these findings to educators in a fashion that allows them to make decisions on further courses of action.

1.5 Scope & Limitations

The focus of this project is to provide further insight into student performance and engagement in HE. The work may apply to other forms of non-compulsory education, but the outputs will not be specifically designed to meet their needs. The secondary focus is for early identification of potentially struggling students. This will change the design and components of the models produced. It is not expected to be transferable to a more generalised form of the problem.

As a result of these foci, a large proportion of traditional LA metrics, such as grades/achievement, will not be available. Typically, the first assessment for a module is set towards the middle of the semester to allow the material to be taught before students start work. We will not be considering any data-set or metric that would not be readily available within the first few weeks of a new academic year.

The limitation of this work is that it is based on data obtained from students at Bangor University only. As the work is designed to discover hidden patterns and infer reasoning for those patterns, other institutions could be hesitant to share their data. While the exact evaluation metric values will change, the methodology should be transferable to any other institution with a similar level of efficacy.

1.6 Contributions

The following contributions are claimed from work presented in this thesis:

Contribution 1 — Bangor Engagement Metric

The Bangor Engagement Metric (BEM) was proven as the strongest predictor variable in the Machine Learning experiments within this work, and the foundation of the earliest versions of the Learning Analytics platform at Bangor University. This underlying, derived metric has been in regular use by tutors across the University since early 2017.

Contribution 2 — Engagement Trace Visualisation

To make the BEM more accessible to non-Computer Science tutors across Bangor, the Engagement Trace visualisation has been integrated into both student and staff facing tutor systems. This visualisation started the Learning Analytics initiative within the University.

Contribution 3 — Predictive Classifier for Student Results

A model utilising the BEM has been able to accurately predict the outcome of students' academic years based on attendance information to the end of the third week of teaching. This model has been shown, across three years, to have predictive accuracy above 90%.

Contribution 4 — Nudge Theory Impacts of Reading Weeks across an Institution

During testing of the Predictive Classifier, anomalies were observed between the predicted set of students and those flagged by previous report-based methods. The prior methods identified a significant number of students that were not identified by the new model. This led to the creation of the Nudge Model showing that schools which do not use Reading Weeks are disproportionately affected by those that do.

Contribution 5 — Student Journey Visualisation

A gap in the current LA offerings is tracking students longitudinally, but it is improving. The Student Journey aims to replace traditional tabular reports during Board of Examiners meetings to show a student's performance throughout their University career. This visualisation is especially useful to identify border-line cases which warrant further attention. The tool was evaluated during a trial in Bangor's Computer Science Board of Examiners (2017/'18) and found to be well received.

Contribution 6 — Degree Pictures

In the same vein as Learning Pictures, part of Precision Teaching [183], Degree Pictures are classifications of a view using the Student Journey Visualisation. These classifications allow faster decisions to be made about a student's overall progress. They can also be used to develop standard pastoral and academic responses/interventions where needed.

1.7 Structure of this Thesis

The remainder of this thesis is structured as follows:

Chapter 2 The Background chapter presents key information which may help readers understand and situate the remainder of this thesis in the field. It covers underlying theory and techniques that are necessary to perform the main body of the work presented. It covers elements of Data Science, Psychology, Moral Philosophy and Education. Also; the drawbacks to educational study design, concerning ethics, are explored.

Chapter 3 This chapter presents a review of the current state of the art of Learning Analytics. The review includes methods, tools, models, metrics and contexts of existing systems as well as research outputs. This survey of the existing literature informs the remainder of the work. In addition, it explores the pedagogic, legal and ethical dimensions to Learning Analytics.

Chapter 4 The initial piece of work created first a probabilistic, then a predictive model using student attendance data over three years at Bangor University. This chapter discusses the creation of the Bangor Engagement Metric, comparisons with other potential metrics, and both seen and unseen data experiments with their methodology and results.

Publication: C. C. Gray and D. Perkins, 'Utilizing Early Engagement and Machine Learning to Predict Student Outcomes', *Computers & Education*, vol. 131, pp. 22–32, 2019. DOI: `10.1016/j.compedu.2018.12.006`.

Chapter 5 In addition to the reporting based on models and disengagement, there is a need to report students' overall progress. This chapter presents a new visualisation to capture the journey through Higher Education, along with a categorisation system to differentiate the potential patterns that emerge.

Publication: C. C. Gray and D. Perkins, 'Visualising the University Degree Journey', presented at the Computer Graphics & Visual Computing (CVCG) Conference, Poster Paper, Swansea, UK, Sep. 2018. DOI: `10.5281/zenodo.1475858`.

Publication: C. C. Gray, D. Perkins and P. D. Ritsos, 'Degree Pictures: Visualizing the University Student Journey', *Assessment & Evaluation in Higher Education*, 2019, Invited paper. DOI: `10.1080/02602938.2019.1676397`.

Chapter 6 Extended testing of the predictive model revealed interesting patterns when determining whether there were students that should have been identified. This work allowed us to perform further numerical analysis to highlight large disturbances in student attendance patterns. This chapter discussed the methodology, results, and links with Behavioural Economics, Psychology, and Learning Design to minimise these effects.

Publication: C. C. Gray, 'Don't Disturb the Student - Investigating Pattern Disturbances with Student Attendance', Presented at AdvanceHE Surveys Conference, May 2018, [Online]. Available: `https://www.heacademy.ac.uk/download/surveys-conference-2018-programme`.

Chapter 7 The final chapter covers the outcomes, conclusions and success of the project. The project and work are evaluated against the original hypothesis and research questions posed. The chapter concludes with a section discussing potential avenues for future work.

Chapter 2

Background

> *The reading of all good books is like a conversation with the finest minds of past centuries.*
>
> — **Rene Descartes**

This chapter provides a review of the relevant background topics from Computer Science, Data Science, Psychology, Moral Philosophy and Education literature. These works all have implications for the contributions claimed in this thesis. This material provides much of the grounding that Learning Analytics is built upon.

2.1 Data Science

Data Science has become one of the leading 'buzzwords' in Computer Science and IT. It has also been pushed into the spotlight, with The Harvard Business Review describing it as the 'sexiest job of the 21st century' [228]. Originally a substitute for the term Computer Science, the term started to develop when Peter Naur proposed 'datalogy' formally in 1966 as 'the science of the nature and use of data'. The contemporary meaning has broadened slightly [239], being described as a set of fundamental principles that guide the extraction of knowledge from data. This definition delegates the processing of data, patterns, and the extraction to a related discipline – Data Mining. Latterly, Data Mining has been augmented, or in some places replaced, by Machine Learning. The field draws upon multiple disciplines;

from mathematics where the use of statistics guides many algorithms, to visualisation which is used to present the patterns and insight unearthed.

2.1.1 The Modern Data Science Pipeline

Hilary and Wiggins [198] coined the acronym OSEMN [1]. The acronym stands for 'Obtain, Scrub, Examine, Model, iNterpret'. Chronologically, these are the steps which must be taken to perform any analytics with a data-set. The OSEMN process has been in wide use since before the acronym's inception in 2010. To many, there is no other way to approach data analysis, especially in the era of Big Data [97]. The acronym has helped recognise that there are four stages [175] to the analytics process.

Before OSEMN can be applied, the first step is to develop a 'data-driven question'. This is effectively identifying a variable to predict or understand. In some areas, such as Higher Education [68], this is a massive change of culture. In other areas, such as retail [267], the need for increasing profit drives organisations in this space to adopt Data Science more easily. Once a goal has been set, data needs to be obtained or collected. In some ways this step is simple to achieve; however, there may be access, copyright, security, privacy and political issues to consider.

Once the data is retrieved, the next process is to scrub or clean it. (In the general description of the pipeline, this is part of the second step along with the collection.) Within Data Science, scrubbing data is where the data is corrected. Such corrections include accounting for format differences between systems, re-coding categorical variables, handling missing and incomplete variables. The precise nature of cleaning the data is specific to the question that is being investigated [49]. Data Scientists must take great care that their cleansing efforts do not alter the meaning of the data, introducing bias into the study/model.

[1] Different groups disagree over the pronunciation of this term, with the two main arguments being 'pronounced awesome', or 'rhymes with possum'.

After the scrubbed data has been assembled into some (problem specific) structure, the third part of the process is to explore the data-set and begin to relate variables to each other. This exploration usually reveals avenues that are viable for answering the data-driven question. These preliminary efforts start to define where and how a data scientist can deploy more sophisticated algorithms or where there is no linkage. While this step can yield a computational benefit, i.e. deploying resources where they make the most sense, it should not dismiss every option. This is due to modern machine learning techniques being able to detect patterns not immediately visible to a human, or not even visible at all.

After the exploration stage has highlighted relations and correlations inherent in the data, the next phase is to build models relating to the data-driven question. This phase entails two essential tasks: selecting the most appropriate facets of the data and selecting the analysis methods that most apply. Before the Big Data boom, Fayyad et al. termed the process 'Knowledge Discovery in Databases' [98]. They describe the process of mining for patterns in the data-set to change aggregated data into information. Lawson [175] describes the facets of a data-set as plot elements in a novel. He observes that not all elements are crucial to the narrative and leaving others out would drastically change that narrative.

Each model will require stringent testing. Firstly, to ensure that the model does answer the data-driven question. This means not only does it provide a plausible answer, but that it meets the intention of the model. This test is needed, as with other endeavours, because there may be unintended consequences [79]. For example; a model may answer a data-driven question 'How much stock will we need to order?'. The model may correctly identify most items, but due to a specific item not being included, it may lead to a drought of that product. Data-driven questions often include some form of prioritisation or direction either of time or resources. Care needs to be taken on the strength of the model that one section is not unduly favoured or disfavoured. The second form of testing is to ensure that there is no *unintentional* bias in the selection, processing, and input of data into the model. As an algorithmic process, the quality of the answers is directly correlated to the input data; or to use a colloquialism ‘garbage in, garbage out’.

Next, data scientists in conjunction with the sponsor of the work will need to interpret the output of the model. For more straightfoward data-driven questions, the interpretation is relatively simple. However more involved models could provide results on different levels. Implicit in this task is the data scientist, or model/tool, communicating the output/response to the sponsors or their colleagues. The form this communication takes will depend on the question and problem that is being solved. Most prevalent are visualisations, tables, predictive and interactive models as software. Care needs to be taken in this step as the correct message delivered ineffectively can lead to sponsors, or other staff/users, potentially drawing incorrect conclusions [15].

2.1.2 Big Data

Big Data is the current fascination for data scientists; however, it is a misunderstood field. The often held, and incorrect, definition is that the sheer size of the data-set qualifies the problem as Big Data. Data scientists have a more formal definition. There are the ‘Four Vs’ of Big Data [132]:

- **Velocity**: The speed the at which the data arrives is high and/or the length of time it remains valid is low.
- **Variety**: The complexity or the vast differences in formats, (inter-)relations, or structure of the data is complex/large.
- **Volume**: The amount of data involved requires different computational techniques or processing of parts at a time. It is important to note that this is not necessarily about how large actual files are on disk.
- **Veracity**: With the volume or velocity of incoming data, there is no longer an option for checking its accuracy. As a result, the usage of the data has to be tolerant of a limited number of inaccuracies.

Crucially, these four factors must combine to make the traditional processing methods unworkable [121]. For example, a 400GB data-set is not automatically considered Big Data, if it can be divided into sections for processing. Similarly, handling 3,000,000 data points per second is not automatically Big Data if the only processing is counting them.

A common, but incorrect, myth is that in order to create an analytics platform/product requires a Big Data data-set [169]. This could not be further from the truth. However; the more facets describing the problem that are available, the more closely matched to reality the resulting model will be [104].

2.1.3 Data Silos

A related, but far less useful, concept is that of data silos. Frequently, in larger organisations, systems and data stores develop independently. These systems had never been intended to be integrated, and most likely are not interoperable. As a result, the databases used by these systems develop and grow with no common linkage [293]. This can occur whether the source is a Big Data store or not. These disparate databases are the data silos. In one regard, this is a benefit — the data is well defined along with its uses. However, the lack of additional context means the data cannot be used to its greatest potential [287]. In order to release that potential, each source needs to be related, potentially inter-related, to other databases and sources. In Enterprise IT, this has become known as creating a 'data lake' [297].

Combining data sources can be a specialist task, requiring the use of data warehouses/data-marts[1]. The structure, of these databases, is fundamentally different from traditional or On- Line Transactional Processing (OLTP) databases. This is due to the workload; analytical queries usually require working with significantly more extensive data sets [149]. This workload is known as On-Line Analytical Processing (OLAP). Facts are discrete data points in whatever problem domain. These facts are described

[1]These terms are sometimes used interchangeably. However, in formal texts [276], a data warehouse is comprised of one or more data-marts.

by dimensions, some of which can be shared — such as dates/times — even with different semantics.

2.1.4 Analytics

The term analytics rose to prominence in 2005. Google's web statistics tool 'Google Analytics' is credited for the popularity [1]. The practice, however, is not new nor are most of the ideas and concepts. However, the definition is usually frustratingly broad. The most common definition is 'using computational methods to discover and report influential patterns in data'.

Analytics is a cross-disciplinary endeavour, sitting at the nexus of statistics, machine learning, information visualisation, data science, and psychology [300]. Often the terms 'analytics' and 'analysis' are used synonymously. However, there is a distinct difference in the Data Science arena. 'Analytics' encompasses more of the OSEMN pipeline, namely the Scrub and Explore phases, whereas the 'analysis' is a sub-component of the Model phase.

Data Science's stated aim is to answer data-driven questions. However, analytics should take this goal further. Its primary driver should be to add insight [214]. What form this insight takes will be tied to the data-driven question. The aim should be for analytics to provide the 'why' for whatever answer is derived. This is known as 'insight analytics', as opposed to 'descriptive analytics' which merely show there is an event/pattern.

The rapid rise of contemporary use of the term and practice of analytics has been driven by commercial self-interest. As companies and organisations collect and generate more data about their subjects (customers, staff, users, visitors, etc.), there has been a natural tendency to attempt to monetise it. This monetisation is not limited to selling customer lists any longer [23]. Analytics users use the wealth of data at their disposal for planning models, forecasting, enticements, detecting inefficiencies, and a range of other business decisions [274].

At its most basic level, analytics can be as simple as bar/column charts in a spreadsheet. However, the modern technical reality is more complicated. Each business sector has introduced its customisations to provide bespoke types of answers to their specific data-driven questions [156]. As with generic Data Science, analytics and the models it generates are only as effective as the communication of results back to the original sponsors [146].

While having appropriate data available is crucial to allow any analysis/analytics, analytics also need to add value themselves. There is little point in building models that state the obvious [173]. The same principle applies to Big Data sources. Care needs to be taken that the additional infrastructure and processing does add value or insight. Unfortunately, often the hype overwhelms necessity, leading to enormous investment without a commensurate return [111].

2.1.5 Predictive Analytics

Predictive Analytics is a sub-discipline of Data Science and Analytics, concerned with mapping causes to effects. It is a form of insight analytics, where the added value is to provide a result based on the input data. Facets of the data are divided into two main groups, the predictor variables and the predicted variables [278]. Not every possible facet will be used in a predictive model, as that facet may be irrelevant to the question being posed. For example, the current price of gold is more than likely unrelated to whether the subject likes to eat fish. Models created with these techniques produce the best-reasoned guess, based on the characteristics displayed by other similar situations/instances. Prior to the interest from the Artificial Intelligence (AI) field, this process was completed manually by statisticians using tools such as regressions and Bayesian inference [21].

Humankind is surrounded by predictive models, from weather forecasts to airline flight planning, credit scores to hunting for a cure to cancer [174, 223, 302]. Since the move to computer-based models, the accuracy and speed at which answers can be obtained have risen significantly. In some areas, notably medical diagnostics, these models can outperform expert humans [235]. One of the earliest examples is the use of credit scoring for mail order companies. It was originally developed by David Durand [301] in 1941. He used discriminant analysis and a data-set of 7,000 historical loans to model which candidates were likely to default. Linear Discriminant Analysis (LDA) is still a foundation of some modern analytical models today.

2.2 Machine Learning

Originally a sub-discipline of AI, Machine Learning (ML) traces its origins to the statistical methods defined as part of classical mathematics [114]. Augmented by Turing's 'Learning Machines' in the 1950s [310], the prospect of Genetic Algorithms was postulated. The 1950s also saw the development of the first recognisable Neural Networks and the Perceptron [263]. By the end of the 1960s, the rate of improvement slowed, and both researcher and public interest in the field waned. As a result, research funding was cut leading to a period named the 'AI Winter' [262]. ML regained popularity in the 1980s when Back-propagation (a method where any error is factored back at each training step) was discovered and added to Neural Networks [261]. By the 1990s, ML had been commercialised, and more capabilities added. The field re-framed the problem as a data-centric question building the rules from known results [313]. Since a host of new traditional classifiers have been added along with new forms of Neural Networks. These advances have spawned a range of applications, capable of defeating humans in strategy games, for example.

2.2.1 Classification Tasks

A standard classification task in ML tries to separate items (called instances represented by n) into two or more groups. These groups are known as classes (C). They can be based on describable aspects of the instances or can be created arbitrarily [159]. This form of ML is known as 'Supervised Learning' as there is an actual, or true, value to compare the classification output against. In this supervised learning process, if the system chooses an incorrect class, the error is fed back to refine the boundary between classes.

The function, mathematical or otherwise, that separates the classes is termed the Classifier. There are many methods of classification, from simple probabilities (the Zero-R classifier [330] which uses the highest prior probability to label all objects) to multi-dimensional processes that attempt to find a hyperplane to differentiate seemingly non-separable classes (Support Vector Machines [294]).

Several discrete steps must be completed to apply classifiers successfully:

1. exploration;
2. feature selection;
3. success criteria selection;
4. classifier selection;
5. training;
6. experimentation and testing;
7. reporting.

2.2.2 Feature Selection

In all but the simplest ML tasks, every instance will be described by multiple facets/data points/readings/metrics/etc. Within the ML field, these are known as Features, represented formally by N. Feature Selection, therefore, is choosing among these features those that allow a classifier to accurately separate the classes represented. Different algorithms exist to complete this process [127].

There are three main methods of operation: Filter, Wrapper and Embedded methods. Filter methods operate irrespective of the data being utilised in the model. This type uses descriptive statistics, e.g. correlation with the predicted class, to rank the features. As this method does not consider relationships between features, it can be prone to selecting redundant ones. However, it is robust to over-fitting and can be completed in relatively short computing times. Over-fitting is an undesirable outcome where the divisions a classifier algorithm has calculated match the original data-set very closely. This then means that the classifier develops errors when previously unseen data is presented, i.e. the classifier has not recognised a sufficiently generalised case to fit other possibilities. Wrapper methods are based around using the output metrics/statistics from a classifier. The user can select which metric to optimise, and the algorithms select the optimal combination of features. The simplest is the pair of Sequential algorithms. These operate in each direction. In the forward direction [326], a set is built up one at a time starting from the empty set. The backward form removes weaker features from the maximal set one by one. They are said to be greedy, at each stage selecting the optimal option in the hope that these will combine to form the globally optimal solution [7]. Embedded algorithms attempt to use the best of both solutions, offering a hybrid set of selected features with a high correlation and optimised classifier metric. This set should be mostly resistant to over-fitting, able to exclude redundancy, as well as perform well in the classification task [264].

2.2.3 Classifier Selection

Selecting the 'correct' classifier algorithm for a given ML task is far less formal than choosing which features to keep/reject. Built up over many projects, researchers develop a feel or instinct as to which classifiers will perform best in what situations. With the growing popularity of ML, tutorials and guides have been created to try to assist in this process [269]. These guides are a best-guess, with researchers still needing to test their selections carefully.

Specific classification algorithms are better suited to a particular type of problem. For example, where one class is drastically over-represented (known as an imbalanced class problem), decision trees perform well [160]. However, when there are complex interactions between collinear features (where more than one feature explains the same aspect of the data), they under-perform [84].

2.2.4 Training & Testing Protocols

Classifier algorithms need to learn specific patterns in each data-set before it can be used on unknown data. This part of the process is called 'training'. During training, the algorithms will adjust the weights, divisions, markers, etc. that are used to determine the output class. This can be done in various ways, such as the iterative method used by Perceptron neural networks. In this algorithm, the weights are altered each time the classifier misclassifies an instance. They are each updated to be more like the misclassified object, mediated by a learning rate [258].

The most basic training/testing protocol is known as Resubstitution. In this method, the same data is used for both training and testing. In theory, this should yield the highest accuracy for the known data-set, as all items are available in the training phase. The drawback is that the parameters can become over-fitted, too tightly matched to the exact data [83]. As a result, the classifier can have significant drops in accuracy with new, previously unseen, data points.

The k-Fold Cross Validation protocol is often used [155] to improve the robustness of a classifier with unseen data points. If k is set to be equal to the number of instances, this protocol is known as Leave-One-Out Cross Validation (LOO-CV)[1]. The principle is that by setting aside one instance for each run, the model cannot become dependent on any single division or split of the data. This does not prevent over-fitting but can reduce the chance significantly. The drawback to using LOO-CV, or k-Fold in general, is that this can take significant computation time. With very complex data-sets or elaborate models, the value of k is often dropped significantly [253].

2.2.5 Classification Metrics

When analysing how successful a classifier has been with a particular problem, ML practitioners use a Confusion Matrix [286]. This grid explains

[1]In more formal notation, LOO-CV is also termed n-Fold Cross Validation.

Table 2.1: An example Confusion Matrix for a hypothetical two class problem. Rows represent the true classes for each instance and columns the predicted or assigned class. At each intersection, there is a count of how many instances with the given true class is assigned the corresponding class by the classifier. In brackets are shown the meaning for each cell: False Negative (FN); False Positive (FP); True Positive (TP); True Negative (TN).

True \ Predicted	Positive	Negative
Positive	34 (TP)	19 (FN)
Negative	8 (FP)	12 (TN)

in one view success and failure on a per class basis. An example is shown in Table 2.1.

The figures listed on the main diagonal of the matrix (where class 1 is predicted as class 1, 2 as 2, etc.) shows the number of correct classifications. A perfect classifier, therefore, has zeros in every cell apart from on this diagonal. The traditional metrics for classifier efficacy are calculated from these grids, shown in Equations (2.1) to (2.5). The terms used in the equations are those used for each cell in Table 2.1.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{2.1}$$

$$\text{Error Rate} = \frac{FP + FN}{TP + FN + FP + TN} \tag{2.2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.3}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2.4}$$

$$\text{F-Measure/F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{2.5}$$

Intuitively, Accuracy and Error Rate (the inverse of each other), are the most obvious metrics. A classifier with a high accuracy/low error rate is presumed to be the best option. However, due to its simplicity, these metrics can be misleading. Precision (also known as the Positive Predictive Value) describes the classifier's accuracy at predicting a correct class from all of the assigned values in that class. This metric has a companion, Negative Predictive Value, looking at the accuracy of not being in the class. The Sensitivity metric also measures accuracy, but this time in relation to the number of instances that are actually in the class rather than just assigned that class.

The F-Measure (also called the F1-Score), or harmonic mean of Precision and Sensitivity, is preferred by many [58, 102, 241]. Specific problem domains, such as those with unbalanced classes, use it almost exclusively [142]. This is due to the fact that a classifier may work for the majority class, and by virtue of there being many more instances in this class achieve high accuracy. For example; in a population of millions, only four individuals have a disease. A classifier that ignores those with the disease, labelling everyone healthy, will achieve over 99.9% accuracy. This classifier would also entirely fail the exercise of attempting to locate disease-affected individuals. Another case that the F-Measure better describes the operation of a classifier is that where one class may be more important than another – for example; detecting cancer cells in a smear. Different classifiers may be designed to recognise different aspects of the sample or have different efficacies in detecting the desired aspect. In these cases, the classifier with the highest F-Measure for that class or aspect should be selected, almost irrespective of the overall accuracy. Unfortunately, no fixed values or rules exist to dictate where these mitigations should be applied. It is left to the designer to determine, usually by trial-and-error, what approaches and metrics are most suitable.

2.3 Visualisation

Visualisation[1] is comprised of three main fields, all at the boundary of Computer Science, Psychology, and Art. The first field is Scientific Visualisation which is concerned with the accuracy of representing 3-D phenomena [179]. While examples are used to illustrate the concepts, Scientific Visualisation is focused on methods and techniques rather than application [39]. The second field is termed Information Visualisation. This field uses computer graphics (including techniques from Scientific

[1] Often the name of the field is Americanised, replacing the second s with a z. The author has decided to use British English spelling throughout this work, unless quoting, or in titles, etc.

Visualisation) to display abstract data, in ways that allow the user to amplify their understanding [51]. Examples can be as simple as representing data in a chart or using colour to grade better and worse results. Examples can also be more advanced where data is overlaid onto a map, image or video in Augmented Reality settings. Interaction and the related Human-Computer Interaction (HCI) concerns form a large part of completed applications. The purpose of this interaction is to allow the user to answer their questions by allowing manipulation of the view and data-set. The final field is Visual Analytics. This exploits the pre-existing abilities of the human visual processing system to convey meaning [147] of information and data.

Much of contemporary visualisation theory originates from the work of Jacques Bertin [30]. ‘Semiology of Graphics’, originally applied to cartography, describes the six retinal variables (also called visual variables): Position, Size, Texture, Shape, Orientation, and Colour. These graphical features can be used to represent selections, associations, ordering, and quantitative data. These ideas hold in modern design, with the variables being combined to present a rich tapestry of complex data in a manner the viewer can reason with.

Kahneman popularised the idea that humans have two systems for making judgements [143]. System 1, the fast mind, is specialised to make snap judgements. This is descended from survival instincts. When using our fast minds, the decisions tend to be based on emotional and biased reasons. System 2, by contrast, is the slow, analytical mind. This is the newer part of the brain, built to evaluate and reason. Within visualisation, this theory is widely acknowledged to be of great use even if different terms are used. Shneiderman’s now-famous mantra [279] “overview first, zoom and filter, then details on demand” appeals to both systems in sequence. The overview allows System 1 to make a quick judgement, termed Pre-attentive Processing by Ware [319], which can then be refined by System 2.

Kirsh and Maglio [152] describe two different forms of interaction with data and systems. Epistemic actions are those where the user is looking for, discovering, or assimilating a data-set presented in a form that cannot be immediately reasoned with. The counterpart is Pragmatic action, something intended to bring the user closer to their current or overall goal. Epistemic actions, almost by definition, appeal to the slow mind. It is attempting to reason with the information presented. They also follow Shneiderman's mantra; however, it is important to recognise that interaction is not always or only about the macro to micro transition [319].

The goal of any interactive system, especially with visualisation or other data-bound applications, is to allow the user to build their own contexts within the data [74]. Allowing these custom contexts allows users to answer questions beyond those envisaged by the creators of the tool. This is made possible by interwoven feedback between the tool and the user. At the base level; users manipulate the data with hand-eye co-ordination manifested as selections, dragging, and movement. User exploration and navigation, the intermediate level, involves the user creating a logical model of the space. Pre-existing skills and knowledge will heavily influence this model. The highest level is where the user can reason and/or solve problems using the visuals provided. The user's progress between the levels is not linear; they may return to lower levels at any time to refine or rephrase their question.

Visual Analytics is not a direct analogue of traditionally recognised analytics processes. It is a re-imagining of the process of dealing with data to exploit the inherent properties of the human vision system [147]. As such, the HCI and Psychology factors are far more prevalent than in other sub-fields of visualisation. A grand unifying theory is not possible in this area as processes and theory are applied on a per-task basis.

Visual Analytics is often seen as a method to deal with information overload [93, 157]. In this guise interaction is imperative, allowing the user to form a bespoke structure or view. It is this ability that sets visual analytics tools apart from standard visual analysis. Development in this field is focused on combining appropriate visualisations with analytic algorithms to support, rather than direct, decision making. These developments recognise that the reasoning required, from the user, is becoming increasingly complex [124].

In visual analytics systems, there is a pre-processing step added before the overview is presented to the user. In this stage the initial, most often summarised, data-set is analysed. As the user refines their area of interest, the data is reanalysed within that specific context. The results are then shown, ready for the next iteration of refinement in scope and analysis.

2.4 Moral Philosophy & Ethics

Very few, if any, teachers actively decide to disadvantage some students over others. However, as with limited resources in any other field, difficult choices sometimes must be made. There is no 'rule book' to guide these choices specifically. Public policy, though, dictates that all students should have equal and plentiful access to resources that they need to succeed. As a result, teachers have had to form their own codes of ethics [296]. Students are then at the mercy of each particular teacher's whims. The teaching profession draws a particular moral disposition, but as Hansen [130] agrees, those entering the profession should be 'good' and 'moral'.

Normative Ethics is the study of a person's behaviour when faced with a situation that expects an ethical or moral action. Western Normative Ethics is a sub-discipline considering those actions and mindsets common in the western hemisphere [219]. There are three main theories; Deontology, Consequentialism, and Virtue Ethics present in contemporary literature.

Deontological Ethics, or Deontology, measures actions by a rule-based system of morals. This has led to the theory being described as obligation or duty ethics. Emmanuel Kant's work is often included as a manifestation of Deontology for two reasons [222]. First, Kant states that a person acting morally is one who is acting from duty [144]. Kant's second axiom states that the motive or reason behind the action is the defining characteristic of whether the action can be considered good, or right.

A study, conducted by Maxwell and Schwimmer [201], found that there are four codes of professional conduct for teachers that apply Deontology. They are:

- treat pupils fairly, respectfully and avoid discrimination;
- exercise discretion in the exchange of personal information;
- observe respect for authority and workplace hierarchy;
- manage criticisms and complaints respectfully and through proper channels.

These obligations parallel teachers' minimum legal requirements in most jurisdictions [291]. While there are cases where civil disobedience is justified, abiding by prevailing law is commonly accepted as a moral obligation [225].

Virtue Ethics, or Virtuism, revolve around the hypothetical virtuous individual and their response to dilemmas. This person embodies all of the classical virtues while resisting all vices. Carr acknowledges the existence of this theory in his book [54]. The theory is dismissed rapidly as it fails to offer practical guidance in any situation. It is often observed as a thought experiment for students to examine how the virtues interact with situations and each other.

The final theory is Consequentialism. Adherents consider only the consequences of any action, not the motive or virtues of the action itself [116]. There are several specialisations of Consequentialism; namely Utilitarianism, Egoism, Altruism, and Teleologicalism [162]. While all look at results, these take different viewpoints/perspectives when considering the result. The Utilitarian considers the action that maximises benefit for the majority while minimising the harm as the morally correct action to take. Egoists measure consequences only in terms of cost and benefit of the actor. While Altruists, on the other hand, are concerned with the welfare and benefit of everyone but themselves (as the actor). Teleology examines the consequences from an impartial, external party's view.

Education literature has shown, in different contexts, that teachers (as separate from support staff and administrators) hold a strong moral view [120, 224, 305]. Oser goes so far as to say that the entire teaching profession has a ‘moral core’. It is unsurprising to those that have spent time teaching, that this would be the case. Teachers pride themselves on putting the student at the centre of all their efforts. A further study, based on self-identification, showed that most teachers find themselves to be some form of Altruist [333].

2.5 Psychology & Education

Many of the classical sub-disciplines within Psychology (especially Behaviourism, Cognitivism, and Developmental Psychology) apply to teaching in some manner [56]. The most immediate aspects are motivation and perspective. The importance of the former is self-evident; however, perspectives lead to learning styles and methods. These fields provide some of the basic elements for the discipline of Pedagogy [321].

Motivation is critical to teaching and learning, to enable a student to achieve their potential [44]. As each learner’s needs are slightly different, there cannot be a single definition of motivation or the optimal conditions to create it. This fact is especially true when merging learning modalities and styles [65]. Some learners benefit from intrinsic motivation, recognised as ‘learning for the joy of learning’; other learners are extrinsically motivated by rewards and/or recognition [311].

LA can provide extrinsic motivation for learners, either as the reward/recognition or the mirror showing improvement needed. The response can be as varied as the learners themselves [243]. Depending on the focus of the system, or the results shown, differing outcomes are evident. This becomes particularly evident when the analytics platform provides comparison capabilities. Social comparisons are one of the strongest drivers of negative self-image [10]. This could lead to learners being discouraged, more so if the result was (or was assumed to be) public.

The comparison may also harm performance, by sapping motivation [254]. Lower performing students could see their situation as irredeemable, high-achieving students can deem themselves ‘safe’. Both situations result in the student disengaging with the learning process to some degree. On the other hand, the comparison could provide the required extrinsic motivation some students need to improve [72].

Mental Health and the psychological effects of University have become a widely reported issue [153, 189, 317]. The issue is a complex interplay of factors, from peer pressure to fit in through high expectations of family to the isolation that admitting a mental health issue can bring. Universities are being encouraged to recognise and provide care for these conditions, as over half of their student populations report one or more mental health issues. Those that start experiencing problems will also tend to develop more [342] without appropriate support and intervention.

Some LA programmes have been extended to detect the precursors of severe mental health issues. Wang et al. completed a study at Dartmouth University to correlate performance, attitudes, and mental health using a smartphone system linked to an analytics platform [318]. Their study suggests a strong link between pressure on the student, whether in the form of workload, social issues, or health concerns, and their general mental health. Other published works [205, 85] have found similar links and make a case for using analytics systems to aid deployment mental health services to those learners in need. There is a fresh focus (~2018) on the potential benefits analytics can bring due high profile coverage of UK university suicides [196, 320].

2.6 Educational Study Design

When considering studies in education, there are two foundation types; formal quantitative/qualitative research and action research. Formal research methods are further broken down into sub-types, such as longitudinal studies, ethnographic research, meta-analyses, as well as more traditional controlled studies [67]. Each has particular ethical and pragmatic concerns, and idiosyncrasies.

Almost all teachers are actively encouraged to undertake a limited form of action research [177] by varying methods, resources, teaching style to adapt to their class. Allowing teachers to do this makes some basic assumptions [106]:

1. Teachers and other professionals have the authority to make these decisions.
2. Teachers, and supporting staff, want to improve their practices.
3. Those involved are committed to continuous professional development.
4. The teachers, and supporting staff, are capable and willing to make the changes.
5. The underlying motive is to make the educational experience better, or more effective for students.

These assumptions do not automatically prevent, devalue, or dismiss action research as an approach. Action research has significant advantages too. Firstly, it can be completed by almost any teaching professional, at any level, irrespective of subject or type of school, to explore almost any question. Next; action research can more directly influence classroom practice. Ferrance argues [101] that this is more effective as practitioners work best on problems that they have identified for themselves, and help each other collaboratively on shared issues.

However, action research has a fundamentally different focus. The focus (as defined by Kemmis and McTaggart [148]) is on “changing individuals, on the one hand, and, on the other, the culture of groups”. There is an implicit requirement for self-reflection when engaging in action research to achieve the change. This is because in most situations, change on the part of the teacher is more effective than expecting or forcing a change in students. Whereas, formal methods take a passive observational focus. The results are analysed and evaluated from a third-party view. As such, there is a known, and accepted, difference in scientific rigour between formal and action research studies [206]. Changes made to factors directly controlled by a teacher are often not scrutinised in the same way that a formal study would be.

Just as in medical research, conventional ethics holds that all educational studies are conducted with beneficial intent. This goal is in keeping with the Hippocratic Oath, and previous studies into the morality of teachers (see Section 2.4). A dichotomy exists, however, when attempting to follow the rigorous approach formal research requires. Most often, this is in the form of Randomised Control Trials [69], where a group of participants are deliberately not affected. Ethically, this introduces a complication: the researcher actively and consciously denies access to hypothesised benefits. To the strict Altruist, this sort of active- denial is unconscionable. If the cost/benefit ratio is high enough, e.g. in medical trials where a real possibility is death from the technique/drug/approach being studied, a form of conditional altruism can be seen [203]. Many of the rationalisations of this conditional approach can appear to be utilitarian [95]. If the study does not use a control group, the researcher cannot be sure which aspects of the study have led to the changes observed [117]. Action research rarely uses a control group, although if an intervention appears effective, it can be subjected to a randomised trial later [43].

Studies that are designed to be preventative, or those that are predictive if actions/interventions are made, will alter the data collected. These types of studies suffer more when implemented as action research. The practitioner designing the intervention is in the majority of the cases, the same that will implement it. That person should want the intervention to be successful, thus introducing a form of self-serving bias [202].

According to ethical best practice, participants must provide informed consent [41] before the study or trial commences, as well as the right to opt-out. As most action research takes place during regularly scheduled lessons/sessions, the students will not have the ability to opt-out. However, depending on the level of education and availability of sufficient resources alternative provisions could potentially be made. Also depending on level, the participant (student) may not be able to grant informed consent. Their age may mean that parental consent is required. These are acknowledged gaps in the educational research process [220].

Also, in common with clinical trials, various factors can dissuade potential participants from granting consent. This may be due to perceived conflicts of interest [140]; or due to fear that an unproven process may harm the participant more than accepted conventional means [200]. When dealing with research that is investigating developmental and educational factors, emotions can often run unusually high [324].

2.7 Learning Analytics

The use of data to support operational decisions is not new, even in education. The techniques available have evolved, allowing the use of petabyte-scale ‘big data’ but fundamentally the practice of Learning Analytics (LA) is still one of statistics, logic, modelling, and inference [341]. There are finer sub-divisions once the problem domain is considered, requiring one type of analytics over others used in another task.

Learning Analytics (LA) has developed exceedingly fast, drawing on multiple areas of study. As such, its history is filled with missteps, repetition, and even contradiction. There is a fundamental split between the traditional inquiry-led research and the practice being implemented. Institutions are racing to implement some form of analytics to best understand the challenges they face [227]. However, as there is no agreed frame of reference between the raw science and implementations, these deployments do not fulfil their full potential [270].

Due to these factors, there is no one single, overarching, but crucially specific definition of the field. Those that do exist are over-broad; such as those from SoLAR, Ferguson, and Clow (see Page 1). Zilvinskis offers "we define learning analytics as the process of using live data collected to predict student success, promote intervention or support based on those predictions, and monitor the influence of that action" [341]. Perhaps a more appropriate method to define the field is by defining the constituent parts of a typical LA deployment or project. This type of definition comes closer to that offered by Greller and Drachsler cited previously. The two critical components are the level of analysis and the intended audience.

The phrase 'level of analysis' is more easily recognised as the scope of an analytics project from small-scoped projects that focus on an individual module, to the holistic whole-programme scope. The wider scope has been termed 'academic analytics' by Siemens and Long [282]. This repeats the original usage by Goldstein and Katz [118] describing their WebCT business intelligence platform, which has since become Blackboard. They acknowledge that the term has flaws, citing that administrative usages appear to be discounted. These two scopes are interrelated, and co-dependent since improvements on a module level will also influence the programme level. Further complicating matters, learner analytics also looks at a holistic view but focused on the learner rather than the programme [233].

The intended audience of an analytics project is somewhat easier to define academically; however, it is often ill-defined by project teams on the ground [45]. Often LA deployments are built to serve staff or students. There is increasing pressure for transparency, most often citing ethical or governance concerns. (See Section 3.4 for a broader treatise of the ethical dimension.) This pressure leads to a hybrid approach where features are slimmed down to those that are common between staff and students. Neither ethical concerns nor data protection legislation precludes differences in features, despite the common misconception that they do. This gives rise to one of the most common complaints regarding LA: users are not provided with enough support to interpret the information being given to them correctly, nor informed of the 'correct' responses [113].

2.8 Summary

The field of Learning Analytics (LA) builds upon a core of mathematics, data science, statistics, and methods to report the findings such as visualisation. These fundamental components are required for all LA deployments and errors introduced at these levels will result in massive differences in the resulting analysis and models. For most 'customers' of LA, the raw statistics are not much help. They require answers to their data-driven questions in terms and forms that they are comfortable reasoning with.

On the one hand, LA practitioners are not interested in individuals - usually when developing the model. However, they must understand the broader environment for their work. These models and analyses concern real people with errors causing real-world issues. These can become acute when dealing with children and young adults, causing potentially irreparable harm. Issues surrounding ethics, morality, and informed consent can never be fully addressed. The more formal the study design, the more robust the result is likely to be, although this would require actively denying some students the chance to benefit from new advances. This dilemma is common and accepted in medicine where the potential risks are potentially fatal, but less so in education. In the mostly altruistic world of education, local (to an institution, programme, or module) concerns are dealt with at that local level by the educator. This form of research is often more effective but unable to be generalised or robustly repeated.

Chapter 3

Related Work

> *Immature poets imitate, mature poets steal; bad poets deface what they take, and good poets make it into something better, or at least something different.*
>
> — **T. S. Eliot**

This chapter goes on to investigate specific areas of the Learning Analytics (LA) field, highlighting relevant related research where appropriate. It examines the science behind the analytics, concerns when applying the results, means of communicating the results and their basis as well as dissenting opinions regarding the field. The chapter also presents similar research/tools/work to that contained in this thesis, citing the differences directly.

3.1 Data Sources and Metrics for Learning Analytics

In 'Learning Analytics Explained', Sclatter proposes that seven types of data can be collected for analytics [270]. These are:

1. Demographics;
2. Sensitive Demographics (as defined by data protection law);
3. Academic Data (module choice, assignment data);
4. Prior Performance Data;
5. Learner-Generated Data (the work produced by the student);

6. Learning Activity Data (logs of interactions with learning);
7. Educational Context Data (curriculum and learning design).

The first five are typically gathered and stored for general institutional purposes. The sixth is a by-product of using a system and usually is collected without any future use in mind. The seventh is generated as part of educational design, whether LA is used or not.

The vast majority of early LA projects utilised existing Virtual Learning Environments (VLEs) or Learning Management Systems (LMSs) as the data-source. VLEs are appealing to researchers due to the rich, automatic, and consistent set of data stored about each student and interaction with the platform. This quantitative data can be interrogated directly, without the need to collect and interpret responses from the students themselves. In 2012 Abdous, He and Yen suggested that data mining/ML techniques were not sophisticated enough to provide certainty over linkage of factors [2]. Their study suggests combining the model results with traditional statistical analysis to make results more robust. Thai-Nghe et al. promote a radically different method, treating prediction as a recommendation system problem [298]. In their method, the feature space is comprised of student, time, and task only.

Chapter 3 of the ‘Handbook of Learning Analytics’ [29] suggests that the psychological measurement of students is the key task in understanding learners’ efforts and results. Bergner points out that due to the inherently human elements, recording these data points are inherently noisy. Small errors in measurement, recording, or mishandling of an individual case can lead to significant differences in the resulting analysis. This is a significant departure from other methods proposed in the literature.

Table 3.1 provides a summary of the proposed and operational systems covered in the papers found during the literature review. The relative density of the VLE column evidences researchers’ reliance on the automatic gathering of data.

There are two sets of columns, the first shows which data stores or systems the analytics platforms rely on. During the analysis of these sources, the following definitions were used to classify the data source types. A Student Information System (SIS) was deemed to be any data system used on campus to store student information and does not have a day-to-day function in the running of a programme or module. Examples are: Banner [94], PeopleSoft Campus [221], and Workday Student [331]. A VLE, for these purposes, is a data system operated to enhance, organise, or deliver teaching/learning content and/or assessment.

The second set shows which metrics or classes of data is used in the model highlighted by the paper. There is an extensive range of potential types of interactions with both the material and activities in the VLE and the instructor and/or other students through the VLE. As such, this review has categorised all these types as a single 'Interaction' data type.

There are two clear patterns shown by the data, which confirm the earlier assertions about the popularity of certain data-sets and sources. The table shows that all but a handful of references utilise some form of data provided by or collected as a side effect of a VLE. This is unsurprising given the ubiquity and reliance on these systems in the modern university. With the rise of online and blended courses, the importance and reliance on VLEs is only going to increase. The second is that interactions are overwhelmingly favoured as 'proof' of student engagement. Again, this is a logical choice; an interaction is always a positive action on the student's part. Thus, researchers do not have to account for any accidental, inadvertent, or side-effect noise in this data-set.

Attendance metrics, in all their forms, are only represented in four of the 33 studies examined. Where attendance is used, the resulting models factor in a minimum of three other metrics/sets of data. The author has taken this to signify that attendance data potentially has unexplored potential as a predictive metric.

Table 3.1: Cross-Tabulation of the Sources and Metrics/Features used through a selection of the available LA literature. This shows a notable trend of using VLE data, focusing on interactions between students and material or other students.

	Sources		Data							
Reference	SIS	VLE	Registration	Demographics	Grades	Interaction	Employment	Disability	Attendance	Peer Comparison
Agudo-Peregrina *et al.* [6]		X				X				
Alstete and Beutell [9]	X	X		X	X	X				
Araque *et al.* [12]	X		X	X	X					
Baker *et al.* [16]		X				X				
Barber and Sharkey [20]		X	X	X	X	X			X	
Blikstein [33]		‡				X				
Draper and Gittoes [86]			X	X	X					
Duval [89]		X				X				
Freitas *et al.* [107]	X	X	X	X	X	X		X	X	
Gaevi *et al.* [112]	X		X	X		X				
Jayaprakash *et al.* [139]	X		X	X	X	X				
Kovacic [161]			X	X			X	X		
Lin *et al.* [181]		X				X				
Lykourentzou *et al.* [188]	X		X	X	X					
Macfadyen and Dawson [190]		X				X				
Pardo *et al.* [226]		X				X				
Richards [246]		X				X				
Romero-Zaldivar *et al.* [255]		X				X				
Romero *et al.* [256]		X				X				
Romero *et al.* [257]		X				X				
Rovira *et al.* [259]		X			X					
Sclater *et al.* [271] § 1	X	X		X	X	X				
... [271] 2		X								X
... [271] 3	X	X	X	X		X	X			
... [271] 4		X		X		X			X	
... [271] 5		X		X		X				
... [271] 6	X		X	X	X	X				
... [271] 7		X			X	X				
... [271] 8	X	X		X		X				
... [271] 9	X	X				X			X	X
... [271] 10	X	X		X	X	X		X		
Ye and Biswas [334]		X				X				
Zafra *et al.* [338]		X				X				
Total Uses	**12**	**26**	**10**	**16**	**12**	**27**	**2**	**3**	**4**	**2**

‡ While this research does include VLE data, it is taken from related sources, such as web-server logs, rather than those provided by the VLE itself.

§ This source provides several state-of-the-art case studies. The numerals refer to the case study number used in the document.

3.2 Visualisation in Learning Analytics

The entire purpose of Learning Analytics is to enable practitioners to take appropriate action in response to any exceptions in their teaching. An underlying assumption of this is that: a) the practitioner understands the information provided in the first place, and b) could and do take appropriate actions. The response, while not unimportant, is somewhat individual based on the practitioners training, experience, and teaching philosophy as well as institutional policy. This observation indicates that communicating the information or message from analytical and predictive models becomes paramount [280]. The field of Information Visualisation explores how information is communicated and shared graphically [335]. Analytics (not just within the education sector) is about discovering, sometimes complex, patterns within data-sets and communicating them [38]; Information Visualisation provides an obvious possible solution.

Nonetheless, implementing visualisations is not new within the Learning Analytics field. There are dashboards [315], organisational visualisations [178], activity, and path visualisations [105]. Periodic reviews of the state of learning analytics [35, 36, 268] find that the questions being asked of analytics systems do not lend themselves to more advanced visualisation or interaction techniques.

3.2.1 Potential Taxonomies

There have been many taxonomies of visualisation using particular models based on a Data State Model [60]; models of the dataset [307]; specific tasks/usages [279]; the algorithms used [308], and differences between user groups [59]. Some argue [108] that Bertin created the first systematic taxonomy in his work Semiologie Graphique' [30].

Most people, if not all, are inherently aware of the most common visualisation techniques irrespective of knowledge about the wider field. These include the simple line, bar, area, and scatter-plots; pie and donut charts; and tables. As part of the computing revolution, technology has been applied to visualisation as well. The extra power spawned better looking' forms of these charts as well as three- dimensional variants and methods [252]. The research in the Information Visualisation community has defined new chart types, such as the Bubble Chart and Tag Clouds [316], and Stream Graphs [48] among many others.

Visualisations of temporal analyses is similarly not a new concept. The literature shows a wealth of applications from medical to policing. The applications build on the same foundation knowledge, the basic techniques that can be exploited to show relationships in time and with time. The classic example is the Timeline [131]. This method is intuitive and follows human narrative idioms of time being a single straight line coming from the past and into the future. Dassi, Nagay, and Fauvet created a taxonomy to collect and describe various other methods exclusively dealing with temporal data [77].

3.2.2 (More) Advanced Visualisation Techniques

The definition of advanced' in this case would depend on the audience. To the Information Visualisation community, items such as Heat Maps or techniques like Brushing are merely another tool in the bag. However, these same techniques in the educational/learning technology community are considered advanced and not frequently used. For comparison, the educational viewpoint will be used as many practitioners will not be visualisation experts. This section provides an introduction to various techniques used in analytics products. It is assumed the common types are already understood/recognised.

Badges/Glyphs

Badges, alternatively Glyphs [37], are symbolic icons used to define a situation summarily. When designing glyphs and badges care must be taken to ensure that for any set of inputs the same glyph must be used, i.e. the algorithms must be deterministic. Shape, size, colour, and opacity are available to convey different factors, as well as combinations of shapes. Typically, we would strive to include an iconic quality; the ability for the glyph

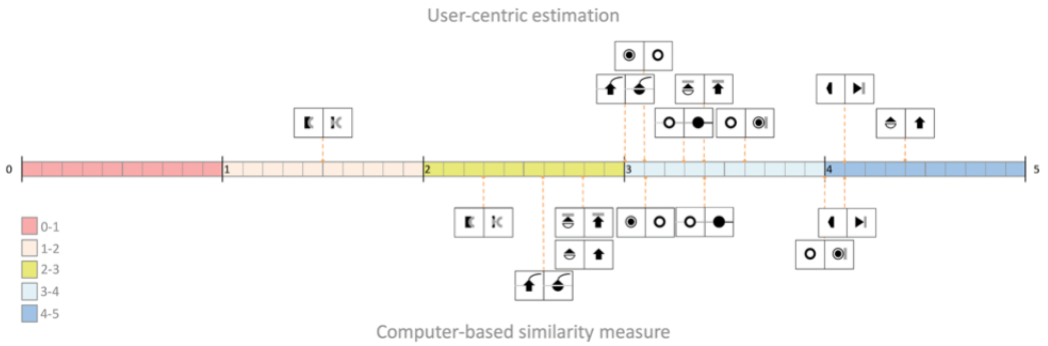


Figure 3.1: Glyph-based Visualisation Example. The Glyphs are shown in the boxes attached to the time-line. Each one represents a different state of the object and so can reoccur along the length of the observation.

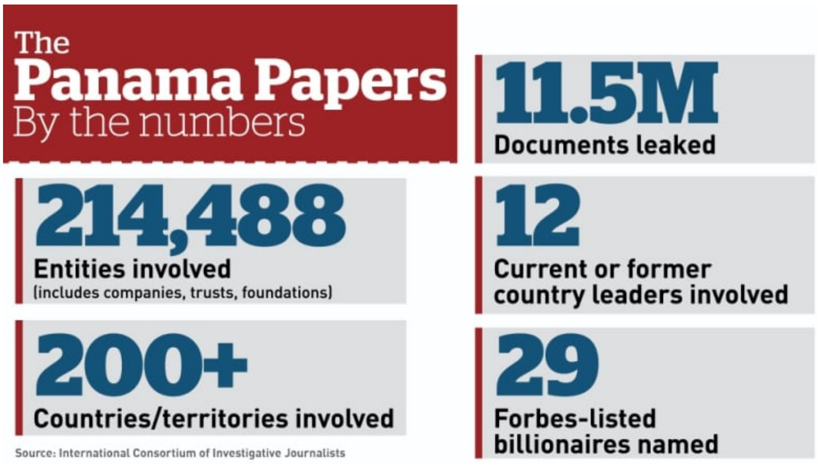


Figure 3.2: Examples of Headline Figures. In this example, the figures do not have any badges or other classification as they just state facts about the data-set. Often Headline Figures show summary or totals rather than individual data points. It is eproduced from https://www.cbc.ca/news/business/royal-bank-panama-papers-clients-canada-revenue-agency-1.3568693.

to be easily recognised at a glance and from a distance. An example is shown in Figure 3.1 taken from Legg et al. [176].

Headline Figures

While not strictly a visualisation technique, the use of a summary statistic can convey a powerful comparative message assuming viewers understand the range and a sense of good/bad for that metric. Typically, these are percentages or integers but may be scaled depending on the range and resolution required. Often the statistic is combined with a badge or traffic light to lend further assistance as the shape and colour give the comparative context. An example is shown in Figure 3.2, taken from CBC Canada's coverage of the Panama Papers.

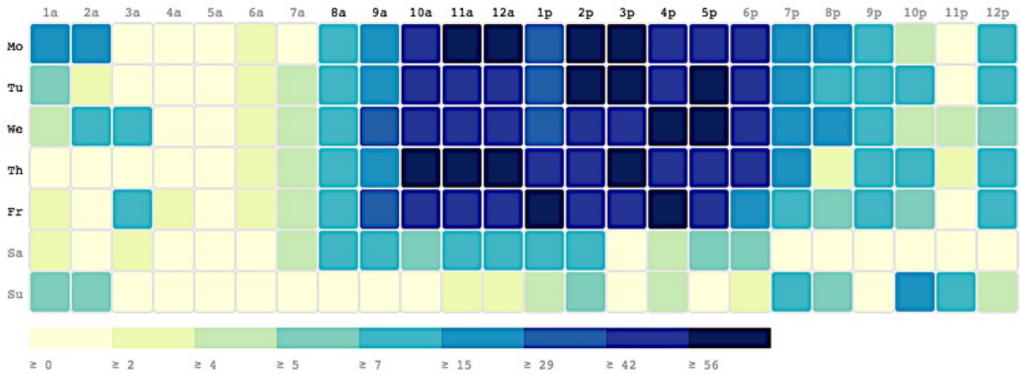


Figure 3.3: An example Heat Map with a Light-Dark Palette. This example also provides the scale for ease of translation to numerical values. It is reproduced from http://bl.ocks.org/tjdecke/5558084, image credit: Tom May.

Heat Maps

Heat Maps are typically a 2-dimensional (but can be 3D) grid comparing two categorical variables in the data-set. Its origins date back to the 19th century but underwent a revival in the mid-2000s [327]. Each cell then represents a count at that intersection, becoming more colourful, more opaque, darker, or similar for a greater or lesser value. An example is shown in Figure 3.3.

Sunburst Charts

Originally descended from the TreeMap [141], the Sunburst Chart combines a radial design (similar to a donut or pie chart) with layers to show relative contributions to the whole. In this chart, the segment size is used to show proportion and layers to show the hierarchical relationship to the root population. An early example is visualising the amount of data stored in a file system, where each ring is a sub-directory [285]. An example, depicting the number of source files in each directory of the Flare JavaScript library, is shown in Figure 3.4.

Florence Nightingale Rose/Aster Plots

The Florence Nightingale Rose is named after its creator, the selfless British Nurse. She first used this visualisation to support her arguments regarding sanitary conditions and their link to disease in the Crimea [218]. Originally

Figure 3.4: An example Sunburst Visualisation showing numbers of source files in different directories within the Flare library. It is reproduced from https://bl.ocks.org/mbostock/4063423, image credit: Mike Bostok, design credit: John Stasko.

used in 1859 (see Figure 3.5), this style of visualisation has enjoyed a resurgence in contemporary Information Visualisation [180, 167].

Bubble Charts

Similar to scatter plots, Bubble Charts show a relationship between two variables. However, as the size, colour, and other properties of each bubble can be varied, they can be used to encode other information as well. The origins of this chart appear to be lost. However, visualisation websites[1] point to French civil engineer Charles Joseph Minard (1859) as one possible answer. When using Bubble Charts care must be taken to ensure that information is not double encoded; for example, in both position and size. An example of this technique is shown in Figure 3.6 using both position and size to encode variation and averages in the same graph.

Radar / Spider Plots

Originally termed 'star plots', in use from 1877 [109]; this visualisation type has seen a modern revival. This plot provides a multivariate comparison for multiple series of data. It offers a similar view as Parallel Coordinate

[1] https://cartographia.wordpress.com/2008/06/16/minards-map-of-port-and-river-tonnage/ and https://visage.co/data-visualization-101-bubble-charts/

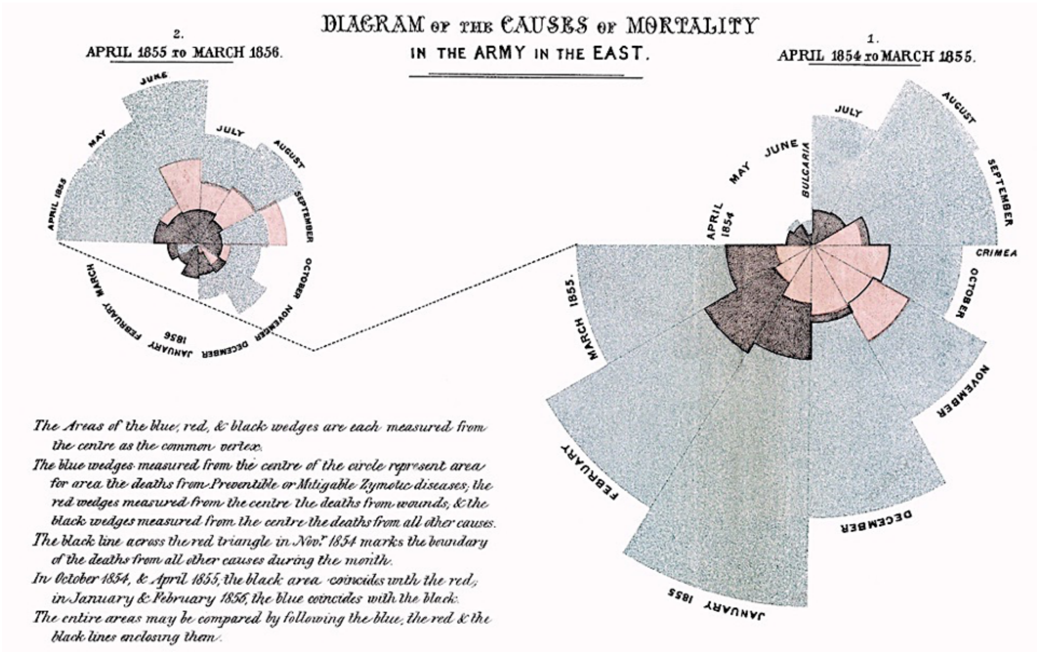


Figure 3.5: Reproduction of the original Nightingale Rose. The image is taken from http://www.historyofinformation.com/expanded.php?id=3815.

Figure 3.6: An example Bubble Chart showing the difference in Price Variation and Sales Volume Variation for two Store Types. Sizes of the bubbles show the extent of the variation of the value. The centre of the bubble is plotted at the average for that store on both axes. The roundness illustrates the relationship between the differences of the two factors. The image is taken from http://dimplejs.org/examples_viewer.html?id=bubbles_standard.

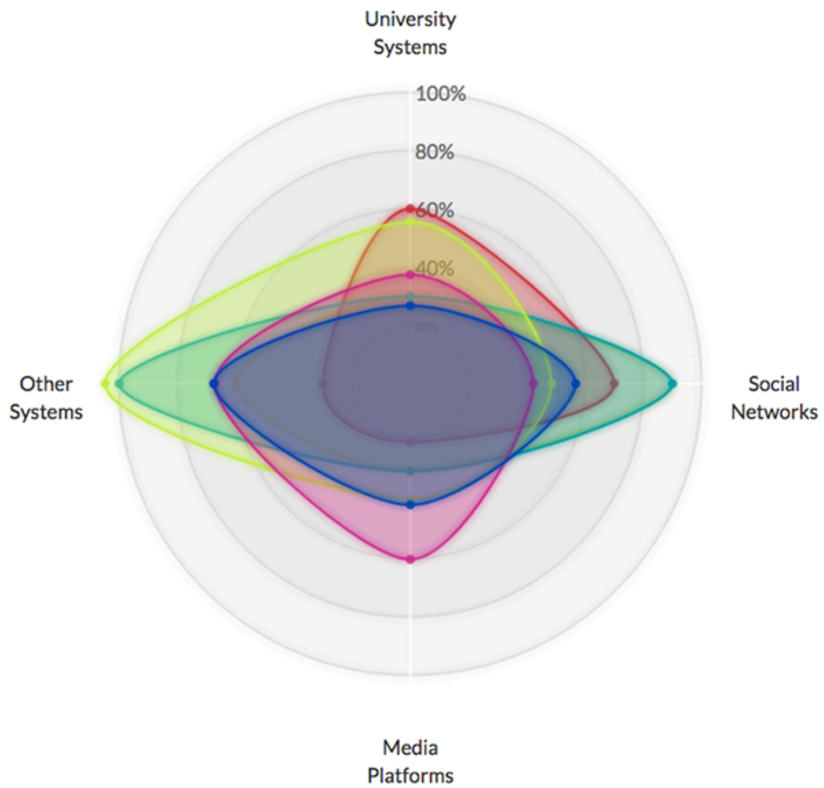


Figure 3.7: An example Radar Plot comparing the use of different media platforms by a HE professional. In this particular plot, the goal is not to maximise neither minimise the overall area but to compare relative positions on each of the axes independently.

Plots but around a common centre using polar coordinates. Differentiation between objects comes by either trying to maximise or minimise the area of the polygon created for each set of observations. Quite often a radar plot will be used to show how well rounded' a solution is; creating a large near circular polygon rather than one with a distorted spike toward one factor. An example comparing different forms and modes of the use of media platforms in HE practice is shown in Figure 3.7.

3.2.3 Visualisation Techniques Used in Learning Analytics Tools

As part of the literature review, a list of the most popular or most cited tools/environments/systems was created. Each was then investigated separately using either published academic works, public data sheets/product literature, or demonstrations. The author acknowledges that these sources will never produce a feature-complete evaluation in all situations. This has been deemed a low-risk eventuality as the investigations

Table 3.2: Comparative Tabulation of the Tools/Systems available in the LA space and the visualisations that are used within them. The visualisation types are split as either standard or advanced. The advanced visualisations are described in Section 3.2.2. There is a strong tendency to rely on the standard visualisations, with the notable exception of badges and headline figures. These were often used in student facing versions of the visualisation. Data gathered from public sources, correct as at September 2017.

	Standard Visualisations							Advanced Visualisations							
Tool	Table	Line Chart	Bar/Column Chart	Area Chart	Pie Chart	Donut Chart	Traffic Light	Badge	Headline Figure	Scatter-plot	Heat Map	Sunburst Chart	Nightingale Rose	Bubble Chart	Radar/Spider Chart
Commercial															
Acrobatiq [4]	X		X		X					X	X				
BB Predict [32]	X	X	X		X	X	X	X	X						
Captivate[5]	X		X			X		X							
Cornerstone [71]	X	X	X	X		X								X	
D2L [76]	X		X									X	X		
Eurekos [209]	X		X						X						X
KlassData [154]	X	X	X												
MGH Insight [204]	X	X	X												
Pearson [229]	X							X							
Renaissance [244]	X	X	X		X					X					
SAP [265]	X		X												
SEAtS [273]	X		X			X		X	X						
Tribal [309]	X		X	X		X			X						
Xyleme [332]	X	X	X		X										
Academic															
LOOP [73]	X	X	X	X	X					X	X				
OU Analyse [166]	X	X		X			X	X	X						
NTU [271]	X	X	X				X	X	X						
Signals [13]	X						X								
UMBC [110]	X	X													
Total Uses:	**19**	**10**	**15**	**4**	**5**	**5**	**4**	**6**	**6**	**3**	**2**	**1**	**1**	**1**	**1**

are only concerned with the types of visualisations offered. This is irrespective of whether any institution has/wishes to deploy them nor whether the use to which they are put is appropriate in the wider Information Visualisation sense.

Table 3.2 shows the results of this investigation. There are 19 tools investigated, 14 are commercial products, and 5 are the results of an academic project. Wherever possible, demonstrations of these tools, or videos thereof, were used to gain an appreciation of the features available. However; in situations where these were not available, pictures, data sheets, or other media related to the product were substituted.

Most Learning Analytics tools are produced commercially. Our comparison highlights a fundamental difference between commercially and academically produced tools. In order to target the broadest possible market, commercial producers most often offer a standalone and VLE integrations. When developing internal solutions, the institution would have access to documentation, support, and even development resources for their chosen VLE. Targeting that VLE (or other systems) would make both business and technical sense instead of generalising the work. Commercial systems must remain generic in order to appeal to a mass market. These questions are most likely going to be answered with a standard set of visualisations. Internal tools, on the other hand, could be far more tailored to the pedagogic needs of the institution. With this kind of tailoring, the insight may be evident from the data and not require any visualisation support.

From the comparison, the use of advanced (or less traditional) visualisations is lower, something that Sclatter confirms in his research [270]. The most often included advanced visualisation is badges; this is due to the current trend of gamification [82]. While there are many possible reasons for a lack of more advanced visualisation, there is an often-overlooked external factor the users. From a Human-Computer Interaction standpoint, developers should offer tools appropriate to the users of the system.

Contemporary Information Visualisation theory does not only cover charts and other graphical representations; it, by definition, includes elements of interaction. This means that charts should no longer remain a static picture, but allow the viewer an element of control [336]. The most basic element of interaction is selection (or deselection), being able to pick an item in some form [145], enabling almost all other interactions. Once items can be selected, the user gains the ability to decide what items should be shown, known as filtering. Filtering can be used to reduce information overload, but only once a user has identified a focal point for themselves. It is vital to stress Interaction does not necessarily need to reduce the items or information shown, but the view may be refocused. Brushing is analogous to highlighting but would also show related items if they were not shown before. When introducing interaction patterns, the tool moves from reporting data graphically to allowing users to explore their data-set. Exploration may, therefore, lead to more profound insights separate from the original analytic question. Ritsos and Roberts agree with this view. They argue that there should be more reliance on visual reasoning [250], a task at which humans excel. If we assume this premise is correct, standard visualisations have limited ability to reveal any extra patterns or insight.

Fundamentally, practitioners and learning analytic designers need to focus on what outcome they wish to achieve. Dyckhoff et al. suggest that the ultimate response to analytics should be self-reflection on the part of the teacher [90]. If this is the case; there is no reason to trawl for deeper patterns, making the visualisations a mere reporting exercise. Dyckhoff's view makes a fundamental assumption that the teacher is responsible for differing performances within the cohort. While 'learning styles' have been discredited [248], there are other solutions teachers can try. Rather than trial and error, better analytics could directly inform these choices. If the overall goal is to place the learner in control of their learning, then the communication will need to be on their terms. Often with commercial solutions, the intended audience is the institution, and as such make assumptions of the users' pedagogical knowledge.

So far, only the system-teacher and system-institution views have been considered. Learning Analytics will always include the student/learner as well. In non-compulsory (including adult) education, transparency concerns would demand that teachers, students, and institutions are provided with the same view. It is not possible to make the same pedagogical expectations of students as of teachers. As a result, any solution must either present a different view to various stakeholders, possibly breaking transparency; or find alternative solutions. Each group will have differing motives, interests, and use cases. If the variety is small enough, fixed solutions may be able to work. If, however, there are considerable differences a more free-form and exploratory environment could be more appropriate. Only after detailed requirements analysis, taking the multiple user groups into account, could this question be answered.

The types and methods of visualisation used in existing tools are important as the users of these tools will be familiar with reading and interpreting these graphics. The review can also be used to evaluate the existing depictions and determine where improvements can be made. A thorough grounding in Information Visualisation may yield results that communicate the intended message more effectively or by showing more facets in a single view.

3.3 Goals of Learning Analytics Systems

Various Learning Analytic predictive models exist in both literature and commercial use. One class attempts to predict students as either passing or failing based on their educational resource usage, such as Virtual Learning Environments (VLEs), Libraries, and other support services [13, 78, 217]. Usage patterns can vary drastically meaning these systems require time to build up a profile of each new class. There is usually a correlation between reduced engagement with support and teaching resources, but the analytics can only flag a potential issue with any given student. A weak correlation means that the 'usage' is only one factor in the model, and usage alone cannot be used to make a definitive prediction due to the variances observed.

The second class of products/tools and accompanying research uses marks/grades to predict the likelihood of a poor outcome or retention issues [135, 260]. These models have a high success rate in flagging poor outcomes as they are directly descended from the constituent components of that outcome. There is one drawback to using grades as the predictor variables. At the point grades are awarded, the student cannot do anything to influence them. This deficiency may be acceptable where there are multiple or formative assessments but fails in courses with single and/or significant summative assessments.

During the late 2000s and early 2010s, a third meta-class became popular where the previous two styles of analytics were combined with in-person interviews or oral/written depositions [134, 295]. The findings in this work show that both teachers and students are looking for more insight into why any particular flagged event was significant, rather than mere reporting of statistics and lists. Within Learning Analytics this is known as moving from descriptive analytics to insight analytics [232].

3.4 Ethics of Learning Analytics Systems

Despite educators self-identifying as some form of Altruist, the age of analytics and data-driven decisions can still be troubling for many students. The literature examining the ethics of LA systems is broadly split into two discrete areas: the ethics and privacy protections surrounding the gathering and storage of student data, and the ethics of any subsequent interventions. These issues require a critical view of the potential impact on students balanced with the benefits that they may expect to see.

3.4.1 Data Collection, Storage, and Privacy

Universities are in the privileged situation of having a massive amount of descriptive data about their customers (the students) but are hesitant or disinterested in using it. This stems from a historical quirk where different departments and/or functions maintained their own set of data [125]. These same institutions are now struggling to adapt to the new data-driven reality, often unsure of the legal, ethical and technical environment in which they find themselves.

Governments the world over are racing to update their data protection and privacy laws in the wake of technology's constant march. This is no more so than the European Union [158], introducing the General Data Protection Regulations (GDPR)[1]. This change in the law requires data controllers (such as universities) to seek explicit permission to process their data, except where doing so is required to complete a contract or requested service [50]. In many ways, universities will proceed in the same ways despite the GDPR. Handling submissions, grades, correspondence, and biographic information is needed to provide the education students have signed up for.

[1]This commentary is assuming that the GDPR remains in force after the outcome of Brexit.

Institutions have already had to grapple with underlying issues [245]. All of the data that is forming part of analytics models is already stored and managed. These include biographical, achievement, attendance, financial, some elements of medical, disability, some aspects of social, and extra-curricular information. Each institution has had to put in place access and security policies to prevent unauthorised use or disclosure of student data.

The issue arises when institutions (including Bangor) want to begin using student information for other purposes, such as analytics. The integrity of any analytics model relies on having access to the broadest range of data possible. In this new legal climate, students would need to provide specific permission for this use [136] as a result, institutions are needing to consult with their students and the students' unions in order to form specific policies on the use of their data [88]. The first widely accepted policy for student data use and the legal/ethical dimensions was created by the Open University [234]. This work included a set of student-facing materials to assist students in understanding the agency they have with their personal data. JISC has since adapted, extended and published guidelines for its members to use in formulating their versions [272]. The JISC version similarly advocates that institutions create accessible student-facing explanations of detailing the processing that they will complete on students' data. Others have suggested that each institution create a 'Data Governance Panel' [123] to handle local conditions concerning the collection and use of student data.

In the LA field's zeal to provide bigger, better, and more effective use of student data we often lose sight of the constraints that student data was collected under. Given a strict interpretation of many of the world's data protection laws (including the UK Data Protection Act, and EU GDPR), students should provide consent for their data to be processed in each new endeavour [270]. Realistically this would place an undue burden on institutions to collect this consent before each new development can be deployed [339]. Care needs to be taken when deciding consent policies, including using vague blanket terms, as LA may face the same counter-intuitive drop in participation seen in medical practice and research [55].

The standard response is to permit students agency over their data [237]. In the ethical sense, this is an appropriate response. If students can choose if, when and what data is made available to various databases and efforts then they are in control and any choice becomes a personal one. However, when considering the implications on LA, it harms all users of the system/model. This is due to less data being made available in the design and/or training phase which may cause entirely different distinctions to be included. Each institution will need to decide if this possible outcome is too significant a risk or if an individual should reasonably be empowered to decrease the quality for the majority.

Another suggestion is to de-personalise all data [24]. In the development of a LA model, this is not as problematic as it would first sound. Statistics, patterns, and ML do not place any emphasis on the identity of an instance in their calculations. However, later use of the output will need to be intrinsically personal in order to be of use.

3.4.2 Interventions and Use of Analytic Models with Students

Most LA systems seek to improve student success. This will inevitably mean examining precursors to target individuals and situations that, on the balance of probabilities, resulting in weak or less than expected outcomes. In other situations, this practice could be labelled predetermination or preemptive action. Humanity has long struggled with the ethical implications surrounding taking action to correct situations we don't categorically *know* to be undesirable This has been observed in many fields, for example the case for the second Iraq war [137], a preemptive organ transplant [231], and conducting research in an educational setting [303]. This trope is the driving force behind Philip K. Dick's 'Minority Report', and the subsequent Hollywood blockbuster of the same name. In these cases, society appears to accept the potential harm(s) when the severity of the alternatives is sufficiently large or widespread.

The specific complaint, in an LA setting, is that educators would be intervening or discussing remedies for potential failure at a point where the student has yet to fail [283]. There are obvious and simplistic, possible courses of action to redress these issues. First, students could be permitted to ‘opt-out’ of data-driven interventions. This is significantly less harmful to the overall LA endeavour, but some students will remain sceptical that data may have prompted any intervention.

There are pedagogic concerns to account for as well; these fit into two categories. The first surrounds data literacy of the educators receiving information and insights from an analytics platform [288]. In this situation, there is a shared responsibility between the developers/implementation team and the academic expected to intervene. If the results are not clear, or the results are poorly interpreted the ensuing intervention may be harmed [129, 150]. While this concern could be levelled at any educational decision support system, there is more faith placed in computerised systems, and even more in analytical systems [208].

The second category involves the design of the actual interventions and their respective usage of the underlying data. This area of HE is relatively under-developed, with interventions being left to the individual to design and implement. There are evaluation methods that have been suggested, such as the Learning Analytics Intervention and Evaluation Framework (LA-IEF), developed by the Open University [249]. Even if such a support mechanism is in place, it is difficult to prove direct causal relations between the intervention and any effect. Without a body of evidence, such as with traditional teaching methods and their evaluation, all interventions become a best-efforts attempt on the part of one educator for the specific student [207].

A current (at the time of writing) debate among both LA practitioners, researchers, and educators is whether comparisons or rankings of students help or harm. The institution has a duty of care to students at both ends of the achievement spectrum. Demonstrating that there is an implicit ranking, assessment, or comparison among students whether acted upon or not will achieve different types of motivation. For high flying students, it has the potential to encourage them to coast rather than strive for their own potential. For low achievers, it can have a similar effect causing them to simply not bother as they can never reach that comparison target. If a band (rather than single comparison point) is used, then the scope for coasting may extend further, from the top of the cohort to the lower bound of the band chosen. It is entirely possible that the comparison will have the opposite effect on any group, causing them to strive for better results.

While comparisons are not a form of assessment, in the strictest sense of the word, as soon as comparisons are drawn an implicit assessment takes place. This implicit assessment may drive some students to improve their standing by 'gaming the system' [17]. This same research suggests that the only sure-fire method to prevent this sort of avoidance is to remove the desire to or sense of reward that may be attributed to 'success' when measured by an attendance metric.

Ethically, as presenting a comparative view of performance may motivate as well as demotivate different elements of the student body, any solution must minimise potential harms. This harm [64] may not be just to low achieving students deciding, on the basis of the presented bands, that there is no more point in trying. It may also affect high achievers, convincing them that they no longer need to strive so hard when 'enough will do'. This is fundamentally decided by the psychology of the individual student; will they react in the way we hope they would? It is almost impossible for any member of staff to reliably answer that question correctly for their tutees, let alone for the entire student body. The Psychology research suggests [10] that social comparison is one of the strongest factors influencing negative self-image when the focus is on winning or 'passing a normative standard'. This effect is not seen when the focus is on 'improvement' or 'trying hard'.

A potential solution is to allow students to set their own targets or measurement. This would bypass the ethics and motivations issues but would not further the educational aim of improvement. Within the Learning Analytics field this is known as 'Self-Regulated Learning' [254]. Ordinarily this approach is applied to curriculum or tasks rather than a comparison metric/target/assessment level. While the approach has the merit of being personalised and may aid individual achievement, the exact level may create a false sense of security. If the student sets a low bar that they meet easily, it could be just as harmful as displaying an average that appears multiples of times more than their performance.

3.5 Linking Attendance and Achievement

Attendance at timetabled sessions is a consistent predictor of likely student retention [103]. The same study also correlated attendance and outcome of 'developmental' courses, which usually occur in the first year of a degree program. While the debate over a cohesive model including all plausible causes of student departure rages; there has been broad agreement that engagement with a course usually leads to higher achievement [275, 304, 277].

West et al. conducted a study linking learning analytics specifically to retention indicators and efforts [322]. Their findings show that students self-reporting of issues is the most common data source when provided with fixed categories. However, when given a free-form answer field, the majority of comments singled out 'class attendance' as the most offered answer. This result was also reported by Anderson, Whittington, and Li [11] confirming that attendance can be used as a strong indicator of a student's final grade. Attendance as a predictor of success has been shown across disciplines, from Economics [292], through Medicine [171], to Psychology [126]. There appears to be a gap in the literature when testing the theory in non-science-based subjects and faculties.

With the inclusion of more technology into education, something that has lagged in the university sector, some have attributed lower attendance to the online availability of materials [119]. This can be particularly acute when applied to lecture capture, either audio [172] or video [122]. A controlled study has since corroborated anecdotal evidence from lecturers [328] that the substitution of lectures does not aid the student. This study shows that student performance only increases when a student both attends the live lecture and uses the recorded form as a revision aid.

3.6 Existing Work on Attendance Analytics

The prior work that utilises attendance data uses it as a proxy measure for engagement. The prevailing theory being that engagement with a learning activity would likely lead to better performance and outcomes [52]. Several studies [61, 164], as well as those highlighted in Section 3.5, have already made the case that using this surrogate measure is not unreasonable. However, raw attendance metrics were only used in four of the systems found in the available literature (see Table 3.1 for references and listing). In these models, attendance is used as one component of the model, without determining the potential power of attendance metric alone.

Some effort has been spent on extending the quantitative view provided by attendance/engagement metrics. Strang [290] designed a smaller scale study looking to augment the raw data with qualitative measures to widen the insight.

Some of the existing work also uses ML techniques, such as Neural Networks [199] and Self Organising Maps [107]. These works are exploratory and examine the possibility of using these techniques rather than maximising success. When reviewing literature combining the phrases ‘learning analytics’ and ‘machine learning’, most results comment on the potential of using ML techniques rather than applying them.

There are only a few examples, found during this literature search, that were specifically designed to highlight issues at the earliest possible juncture. (This is one of the defining characteristics of the work presented in this thesis.) Barber and Sharkey [20] attempted to use basic statistical models to predict failure early. Their work used a similar timescale (the first three weeks of the course), however, included more factors than simple attendance counts or ratios.

3.7 Criticism of Learning Analytics

Not all educators nor researchers consider LA a force for good. There are three main categories of objections; that the measurements and analysis are to be used solely for ranking, that the measures used do not accurately reflect the learning experience, and that in all the efforts to analyse data practitioners lose sight of the human reality of education. The remaining category of criticism is surrounding the ethics of using data to predict a possible future for a student. This concern, among others, has already been covered in Section 3.4.

There is a growing trend towards the use of analytics, metrics and scores as a surrogate for measuring actual learning. In effect, making the practice of measurement more important than the measurements themselves [193]. This has been born out of the concept that students must be shown that their fees are being used to demonstrate that they are in fact learning. Macfarlane argues that when attendance is used as the key metric, the mere act of attending matters more than the learning that should occur in that time. When this concept is taken to its logical extreme, we end up with an academic surveillance programme [192].

This sentiment is echoed by Lodge and Lewis [185], questioning whether the practice of analytics is trying to assist in measuring quantities (such as attendance) or the quality of the experience, learning, etc. Beattie, Woodley, and Souter – as part of their critique of analytics in HE – coined the term 'creepy analytics' [24]. Their central themes align with the ethical objections that have already been examined; however, their work does pose a related question 'when is there too much analytics?'. By probing the data, all manner of conclusions and patterns could be found. However, little consideration is put into whether that pattern should be used.

Macfayden and Dawson [191] argue along the same lines. Their position is that arming educators and management with data does not necessarily nor automatically improve the experience for students. They postulate that institutional and social change must also occur to make effective use of any analytical platform. Ferguson [99] also argues that institutions can rush into a deployment without considering the educational element, treating LA more like an enterprise IT project than an educational one. Gašević agrees [113], arguing that while data may appear to be rife, little of the collected data accurately reflects the conditions of the learning environment. Macfarlane and Tomlinson also decry the use of student engagement and analysis of this data [194]. They argue that the use of these emerging technologies – without appropriate social, policy, or governance support – can lead to negative behavioural effects in students.

Following in the footsteps of Dijkstra, Dringus contributed another '... considered harmful' essay [87] to decry various elements of LA. His first criticism is that institutions deploy systems but do not consider the source and accuracy of underlying data until too late. The catchphrase Dringus employs "getting the right data, and getting the data right"is entirely correct. It highlights an issue that HE has not grappled with on a large scale. A similar position is put forward by Ferguson and Clow [100], suggesting that the limitation of LA projects occurring within institutions presents unintended bias of the data.

3.8 Summary

Learning Analytics is a relatively new field of study and has largely been approached from a Technology Enhanced Teaching/Learning (TET/TEL) viewpoint. This has resulted in little cross-disciplinary efforts, missing out on the wealth of knowledge and techniques from other fields. For example, Table 3.2 shows that LA could benefit from the input of the Information Visualisation community. The available literature also shows a preoccupation with certain types of metrics and data, such as a fixation on 'engagement' or 'interaction' with a VLE. This is not to say these are the only metrics used but, less effort has been expended on finding new metrics or methods to incorporate other data into existing models.

There is a significant body of work proving, or at least supporting, the link between attendance and performance. This tallies with anecdotal evidence from practitioners. This work also includes investigations linking the use of technology to a lowering of attendance. This work found that when students substituted attendance for use the of this technology, their performance remained the same. Work has been conducted looking into the use of attendance in analytics, usually as only a part of the overall risk or classification. Those works that have used attendance have not focused on the earliest possible point where attendance patterns become predictive of the student's performance. Similarly, investigations into utilising Machine Learning techniques to predict student achievement have been completed before this work. In these cases, the studies look into whether it is possible to use ML rather than maximising its efficacy.

Not all work regarding LA is positive. Some see this work as another invasion of privacy, prejudging students before they have a chance to perform. As a result, there is a large body of work dedicated to the ethics of using student data. This work debates the merits and deficits; it recommends that institutions must consult students and create transparent policies and processes for dealing with data. The remaining criticisms strike at the methods and motivations of LA implementations. These new metrics could be utilised as just another way of monitoring and ranking students, cohorts, and institutions while not improving the experience or results of those being monitored. Other arguments suggest that the data being used in these models is not suitable for the conclusions being drawn from it. This use of proxy metrics can also include unintended biases.

Practitioners and their institutions, are very much left on their own to find solutions to rectify patterns discovered through their analytics. There has not, yet, been sufficient research into the courses of action that educators should take. This area remains based on the individual experience of the tutor concerned making the best efforts intervention with a single student. There is also the genuine possibility that intervention with students that are ultimately going to perform well may demoralise them and cause the very thing that the intervention was intended to solve.

3.8.1 Reflections on LA

Learning Analytics has a significant amount of promise, potentially being able to add insight to explain and improve all teaching. However, there is a split forming within the published work. On one side is the purely technical investigations, looking at what is possible with application of both traditional and non-traditional analytical methods. On the other side are educationalists and ethicists examining whether this endeavour is moral or compatible with the long-held ideals in education. The author takes the view that analysis and data can not be inherently ethical or not – in the same vein that science itself is not dangerous. However, the author acknowledges that both science and analytics can be applied in ways that are immoral and damaging. In the wider context, this is what has led to regulation in the form of the EU General Data Protection Regulations and the equivalent national laws.

Much of the technical literature focuses on 'low-hanging fruit', using data-sets that are easily obtained and combined into a model which can provide an answer. There is little exploration of the other wealth of information held about students and the learning environment. Neither is there an appetite to explore how soon technology, in particular LA and ML, can provide assistance to educators to prevent a poor outcome.

The work surrounding the ethics of LA can be further sub-divided into two sets; those proposing potential codes of practice or implementations to preserve student rights, and those decrying LA as an intrusion into student lives or a technological overreach. Simply demonising a new set of developments ultimately hampers valuable developments, therefore the author recognises the importance of the second set of work but respectfully disagrees. The application of data (and therefore data science) to HE is inevitable, just as compulsory education is now measured with metrics and league tables. Institutions (and researchers) can therefore either adapt to the changing times or attempt to turn back the tide. This work was aimed at filling the gap left between these camps, applying established ML techniques to a new class of data that had been largely ignored.

Furthermore, once models or systems identify a potential issue these systems leave educators to decide on an appropriate intervention with little to no guidance. This may be an appropriate strategy if the implementing institution is filled with expert and caring tutors, but does nothing to support inexperienced staff. The author views this as a problem in desperate need of a solution. However, it is also one that cannot be completely solved by the application of technology. As a result, analytics tools must communicate a complex narrative involving many differing factors. So far, the discipline has yet to capitalise fully on the use of modern visualisation techniques to convey this narrative.

Learning Analytics has a bright future, with the possibility of bringing many advances in education and many benefits to our students. The research and implementations in this discipline must become more open and inter-disciplinary to usher in this change. Combining the best of breed technologies from data science, analytics and visualisation with the latest advances in pedagogy and psychology, scientists and learning technology experts could revolutionise how we view our students and their behaviours.

Chapter 4

Predictive Model of Student Outcomes Using Machine Learning

> *Numbers have an important story to tell. They rely on you to give them a clear and convincing voice.*
>
> — **Stephen Few**

This chapter describes the development of a probabilistic, then followed by a predictive model for student outcomes — whether each student will pass or fail their current academic year, or will be required to complete supplementary assessment. It examines potential sources of data, the development process, and results obtained. The model, developed throughout this chapter, will become a key enabling technology for the overall system this work aims to construct. The work in this chapter helps to provide answers to research questions 1-3. As such, it is not intended to have direct user interfaces, but results can be extracted for comparison and validation at this stage.

This work has been published as; C. C. Gray and D. Perkins, 'Utilizing Early Engagement and Machine Learning to Predict Student Outcomes', *Computers & Education*, vol. 131, pp. 22–32, 2019. DOI: `10.1016/j.compedu.2018.12.006`.

4.1 Model Creation Methodology

This work started with a definition of metrics and data points that may be useful in determining students' performance at an early juncture. This list was cross-checked with a survey of relevant literature (within Education, Machine Learning, Learning Analytics, and Operations Research) to determine which have been investigated before. The list of metrics was further restricted to those that can be obtained in real-time or near real-time and would be available from the start of a student's studies.

Once candidate metrics were identified, work began to obtain the data from the institution. It is expected that, due to Privacy and Data Protection controls, not all metrics or information will be available for use in research. Controls on the raw data must be observed. Identifiers must be retained to relate data to each other, but they may be changed to decouple the results from individuals. From this point, the work mirrored the standard Data Science Pipeline (as described in Section 2.1.1).

After collecting an appropriate data-set, initial explorations into it were conducted. This served two purposes; first to infer structure and/or relationships, and second, to understand where data was suspect. The data could be suspect for a variety of reasons, most often coding errors and missing data. As a result, an appropriate strategy to mitigate these flaws, and others, was applied.

After cleaning and exploration, a series of trials were conducted. These aimed to find the best combination of metrics and machine learning tools to produce the most informative results for tutors and students. Aspects of the problem, most notably the unbalanced nature of the classes of student outcomes (pass, fail, etc.), will influence these selections - was aided by guidance from relevant literature. These trials used both seen and previously unseen historical data. The parameters for the trials were further revised based on the results until an appropriate model was reached.

The resulting model will require testing and evaluation beyond the trials that provided the specification. It is proposed to also test the model on the succeeding academic year. The predictions made by the model will be made available to tutors to determine if interventions were made, and if so did they have a positive effect.

4.2 Possible Metrics Selections

The overriding objective of this model is to predict which students will not be retained, either through academic failure or other factors. This makes the availability of various metrics and data crucial. Each candidate must be both available, i.e. collected rather than implied or stated anecdotally, and timely, i.e. available for analysis in near real-time rather than an end-of-semester summary or only calculated periodically.

Based on local institutional knowledge and experience, an initial list of candidates was compiled. This list divided potential candidates into different categories, to avoid over-sampling from each category. (See Appendix A.1 for the full list.) In theory, all of these metrics should be stored in one manner or another by University data controllers. However, Bangor — just as with many other large organisations — suffers from Data Siloing. This information is fragmented across, up to, seven differing systems. These systems are disparate and require data to be transformed before it can be matched with other systems. Figure B.1 shows a conceptual layout of data systems within the University at the start of this work (~2016).

The metrics envisioned/described so far are positive measures of a student's engagement with their studies. However, there is no description of the student as a person. This demographic data is just as important as any other feature of a student but can be somewhat more contentious. Including biographical and demographic data can be seen by some as a step toward racial, ethnic, gender, orientation, religious, or other profiling [24]. The demographics that were initially considered can be found in Appendix A.2. This list contains both personal and sensitive categories of information.

Despite a Data Protection Registration [138][1] that permits the use of all categories of information for research purposes, this work was not permitted to use them.

The final set of metrics and information provided by the University used in this work is: -

- Student ID
- Level Code (Undergraduate/Postgraduate)
- Year of Programme
- Home Department
- Programme
- Academic Standing Code
- Degree Standing Code (In Progress/Completed/Failed)
- Attendance - Scheduled Sessions
- Attendance - Tutor/Supervisor Meetings
- Attendance - Other Events

This data was provided with express consent of Bangor University's Student Engagement Unit, Compliance Team and Information Services Department. This work was completed with approval under the University Research Ethics Policy and Process.

The Academic Standing Code is the metric that maps directly to the student's outcome. (See the Definitions section, on Page xii, for a full definition of this code.) This work uses the Academic Standing as the *predicted* variable or in ML terminology, the class label. The student ID is included to be a unique reference (to ensure instances are not included twice) but will not be used as a metric in the resulting models.

This work has a keen focus on student attendance for two main reasons. Prior studies [11, 25, 91] have found that while not a perfect analogy

[1] Since the introduction of the General Data Protection Regulations (GDPR), the Information Commissioner no longer shows the categories of data permitted and the use on their public register.

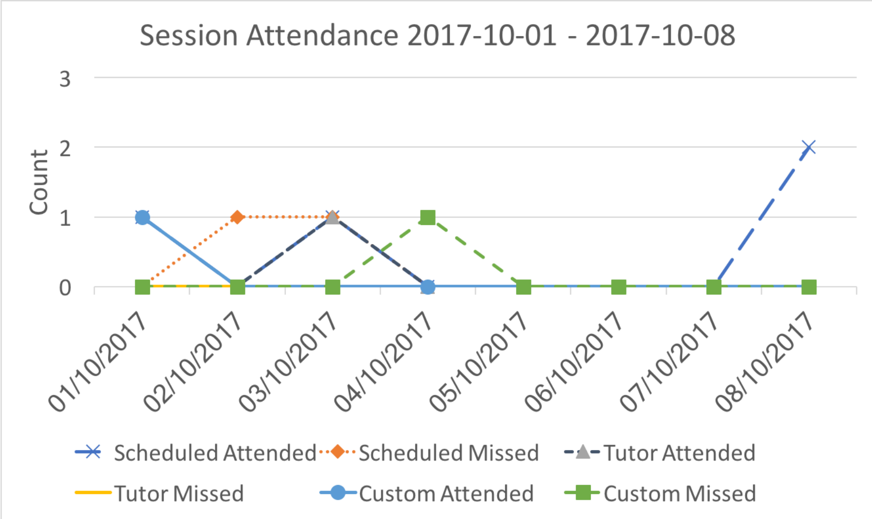


Figure 4.1: Recreation of the original display in Bangor's analytics for student attendance. This representation was not created as part of this work, and was almost universally condemned as not useful. The x-axis represents time, and the y-axis represents the number of occurrences in each series.

for achievement, attendance and achievement are nonetheless strongly correlated. Engagement/attendance has also been previously linked to the likely retention of the student [103]. Secondly, that attendance data can be analysed as soon as it is created, i.e. the session has occurred, or the class register is input.

4.3 The Bangor Engagement Metric

In Bangor University's first iteration of adding LA to their systems, there was a single line graph showing the various attendance metrics for the student. The attendance was shown as the count of sessions attended and missed from day to day. A recreation of this display is shown in Figure 4.1.

Due to the amount of over-plotting (where multiple parts of the visualisation directly overlap), users reacted negatively to this display. From an Information Visualisation point of view, there are several incorrect elements. Line Graphs automatically suggest that there is a relationship between the plotted points, as this chart is based on fixed readings (in the same way a histogram is) there is no intermediate value between those points. This view also makes it hard for a user to reason about any overall performance as there is no cumulative element, neither in the plot itself or accompanying text.

As a result, the Centre for Enhancement of Learning and Teaching (CELT) requested a new method for displaying this data. As part of this work, the engagement metric was created, with the aim of combining the various different counts as well as maintaining a graphical analogy – an upward trend is good, and vice-versa. This became formalised as the Bangor Engagement Metric (BEM). This metric is created from an intuitive rule; a student gains a point for each scheduled event attended and loses one for each event missed. A set of five alternatives were devised; a) showing all data as cumulative totals, b) cumulative totals with missed events being negative, c) the ratio of sessions attended, d) the engagement metric (defined later in this section), and e) engagement metric with sessions attended/missed overlay. These alternatives are shown in Figure 4.2.

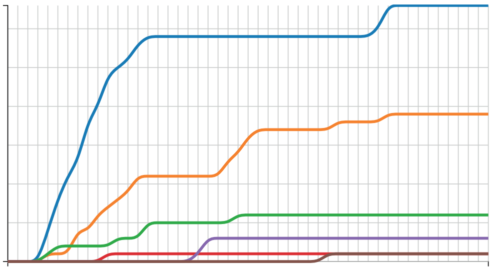
In the raw data-set, there would be k observations for each student. Each observation is coded as either one for attendance and zero otherwise. These observations are recorded in the set $z = \{z_1, z_2, ..., z_k\}$. Therefore, the ratio of attended/missed for each student (termed Engagement Ratio (ER)) is defined mathematically as Equation (4.1).

$$ER_k = {}^1\!/_k \sum_{j=1}^{k} z_j \tag{4.1}$$

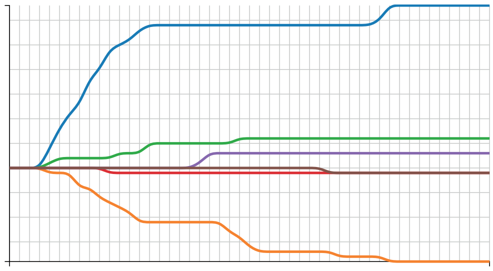
Using the same base notations for the set and observations; the Bangor Engagement Metric (BEM) is defined as Equation (4.2).

$$BEM_k = \sum_{j=1}^{k} (-1)^{(1-z_j)} \tag{4.2}$$

While these two measures may at first appear directly related, it can be proven that there is no straightforward relationship using a series of Monte

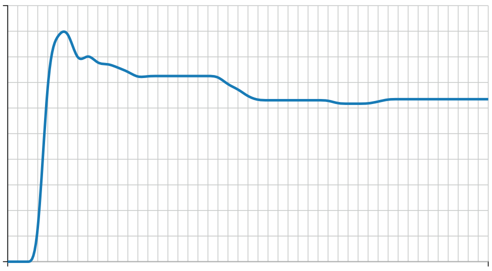


(a) Attendance Shown as Cumulative Totals; the y-axis represents the number of sessions.

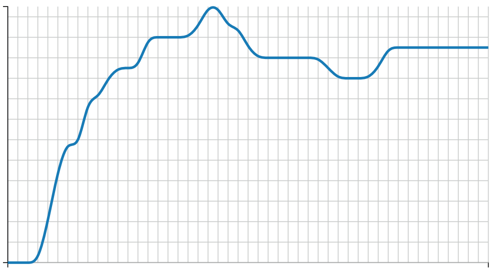


(b) Attendance Shown as Cumulative Totals with Missed Sessions as Negative values; the y-axis represents the number of sessions.

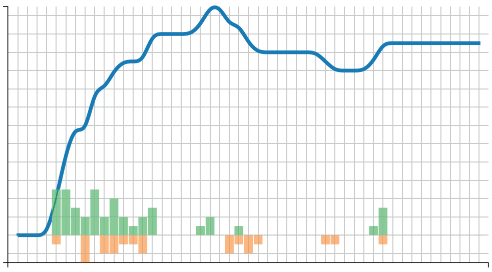
Figure 4.2(a) and Figure 4.2(b) show each event attended and missed for each of scheduled events, tutor meetings and custom events as separate lines.



(c) Attendance Shown as Ratio of Sessions Attended; the y-axis represents a percentage as a decimal between 0 and 1.



(d) Attendance Shown as Engagement Metric; the y-axis represents the metric score.



(e) Attendance Shown as Engagement Metric with Attended/Missed Bar Overlay; the y-axis of the line represents the metric score, the bars are counts of sessions attended and missed.

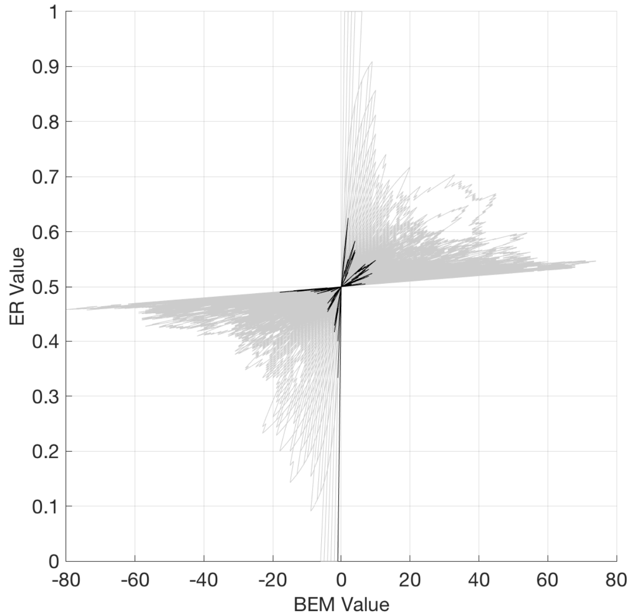
Figure 4.2: Alternative Visualisations and Measures for Student Attendance Data. In all charts the x/horizontal axis represents time, with a data point summarising each week of a full academic year.

Carlo [211] trials. For these trails, each z_i is a Bernoulli random variable taking the value 1 with $p = 0.5$. All z_i values are sampled independently. For each virtual individual, there are $k = 1000$ observations. A set of 200 virtual individuals are created per run. For each individual, and at each interval i, the ER and BEM values are calculated using their previously defined equations. Figure 4.3 shows the resulting scatter-plot when every individual's ER is mapped to the BEM at each observation. One individual is highlighted in black. If there were a direct relationship between these two measures, a single (over-plotted) line or curve would be produced. These plots are based on random data, so a tight band would also be indicative of a direct relationship between the metrics.

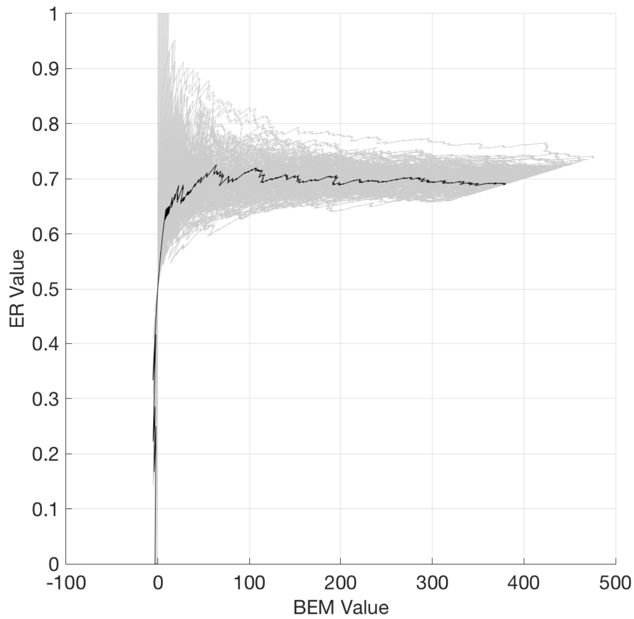
A random sample of 500 students was taken from the 2016/'17 academic year. This includes all 39 teaching and assessment weeks. From this data, each student's BEM and ER metric was calculated. A new scatter-plot showing this actual (rather than simulated/random data) is shown in Figure 4.4. Again, a random individual has been highlighted in black.

Comparison of the two sets of scatter-plots show the simulated data to be in line with actual results, and that neither set shows a direct relationship between the metrics. The difference is starker when comparing a 10-bin histogram for linearly sampled weeks for a random set of 500 students. This is shown in Figure 4.5. The salient point of the comparison of distributions is that the BEM displays an approximately normal distribution, and the ER does not. As a result, empirical testing of both metrics will be required when completing the ML experiments.

However, there is a flaw which affects both the BEM and ER. There are two situations which can produce the same values for adjacent entries. These are; no monitored sessions occurred between the two points, and that the student missed as many monitored sessions as they attended. These issues will be evident when using any form of summarisation of the values; i.e. whenever the set z does not contain every distinct event as a separate value. The risk of these two situations is proportional to the summarisation length.



(a) Attendance Probability $p = 0.5$



(b) Attendance Probability $p = 0.7$

Figure 4.3: Scatter-plots comparing the Engagement Ratio (ER) and Bangor Engagement Metric (BEM) values for 200 virtual individuals with $k = 1000$ random observations. A random individual is highlighted in black. Plot (b) at first glance appears to show a correlation with generated data, the perturbations in the line disprove this theory.

The more values that are included in the summary reading the more likely one or both situations will be encountered.

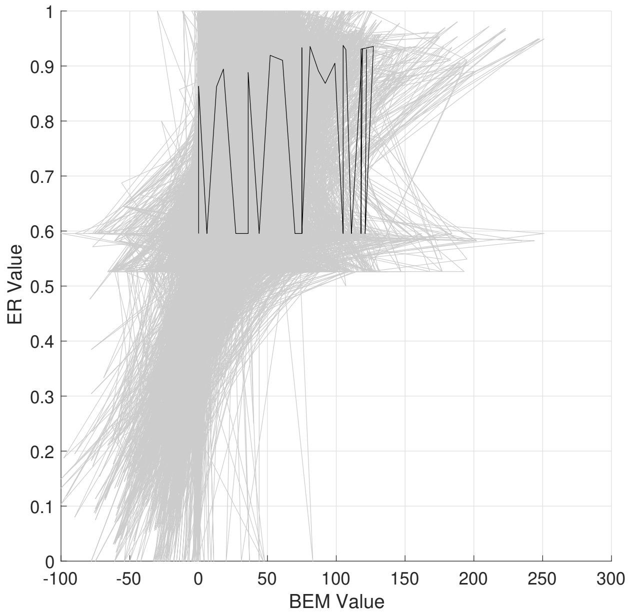


Figure 4.4: Scatter-plots comparing the Engagement Ratio (ER) and Bangor Engagement Metric (BEM) values for 500 student records in the 2016/'17 academic year. A random individual is highlighted in black. This plot confirms that contrary to the idealised plot in Figure 4.3 (b), the BEM and ER have no correlation between them for observed (rather than generated) attendance data.

4.4 Engagement Traces & Insights

Using the new summary statistic (the BEM), the project progressed to the third stage of the Data Science Pipeline, exploration. These initial explorations involved simply plotting various combinations of the metrics against time.

These have been termed 'Engagement Traces'. On each plot, the x-axis, representing time, always covers the same thirty[1] weeks of the academic year. The y-axis represents the BEM value. To aid comparison, both within the plot and between plots, a standard range for the BEM needs to be selected. An appropriate starting point would be between the positive total number of sessions and the negative total number of sessions. This recognises that a perfect attendance score would be at the top end of the scale, and no attendance at the bottom. In practice, very few (if any) students will achieve the maximal and minimal values. It will also

[1]Comprised of 2 x 12-week semesters, 2-week January assessment period, and 4-week May assessment period.

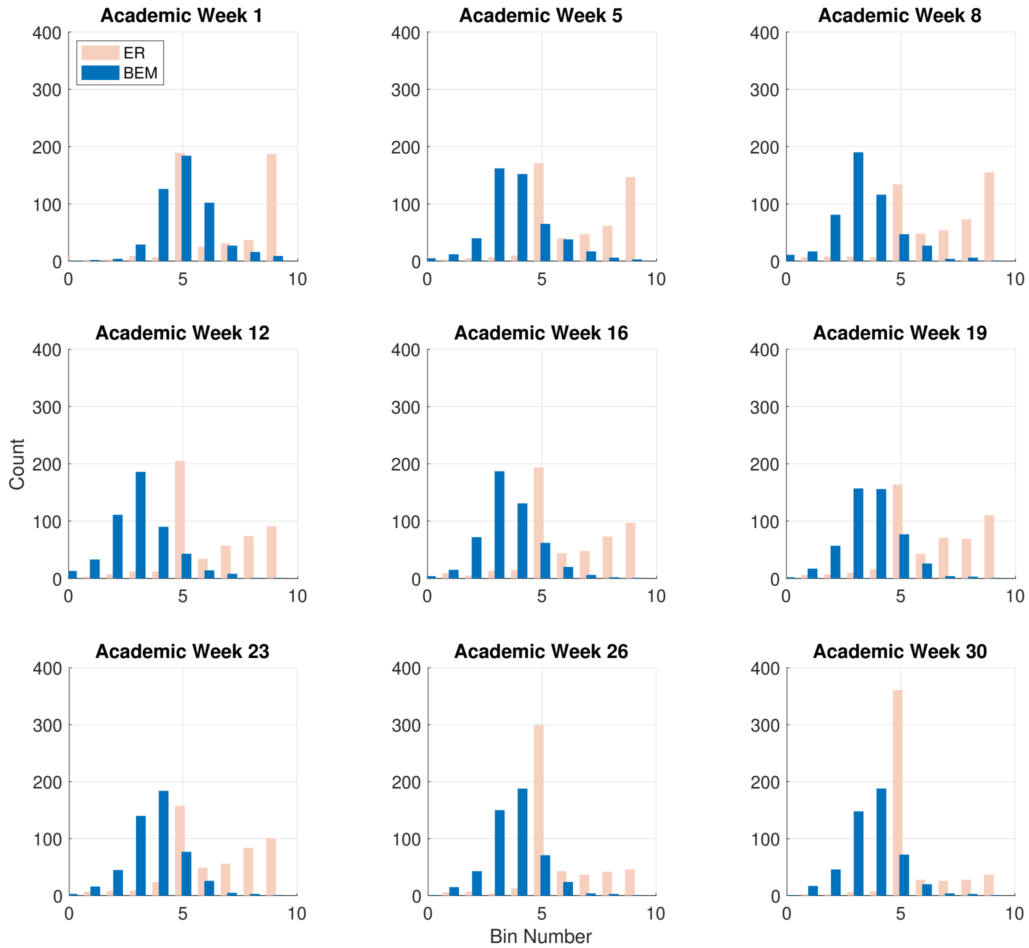
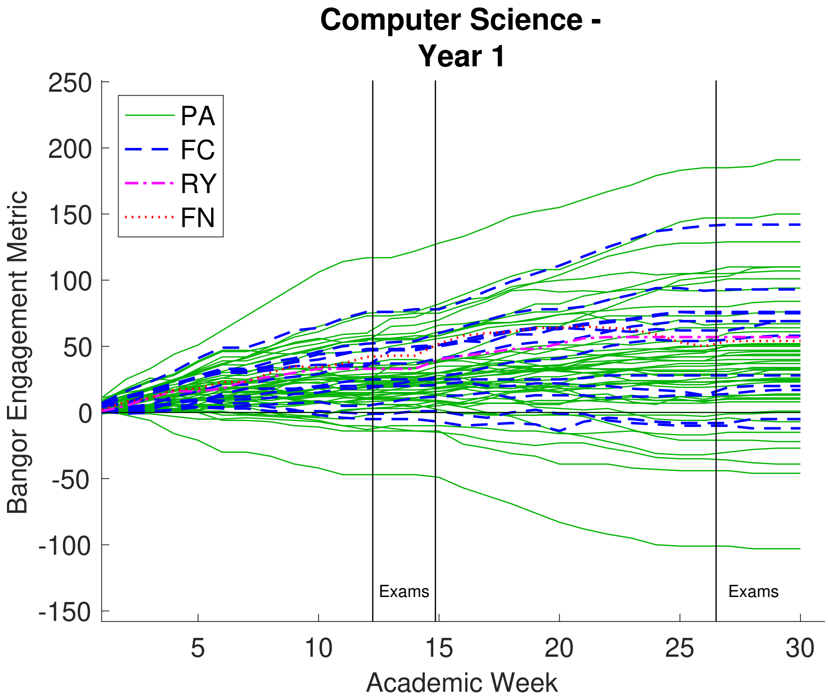


Figure 4.5: 10-Bin Histogram comparing a random sample of 500 students' BEM and ER values for linearly sampled weeks throughout the academic year. While the magnitude and skew alter in some weeks, the columns show that the BEM maintains an approximately normal distribution. This suggests a more appropriate measure and stronger validity under statistical analysis. The ER population does not exhibit any normal tendencies.

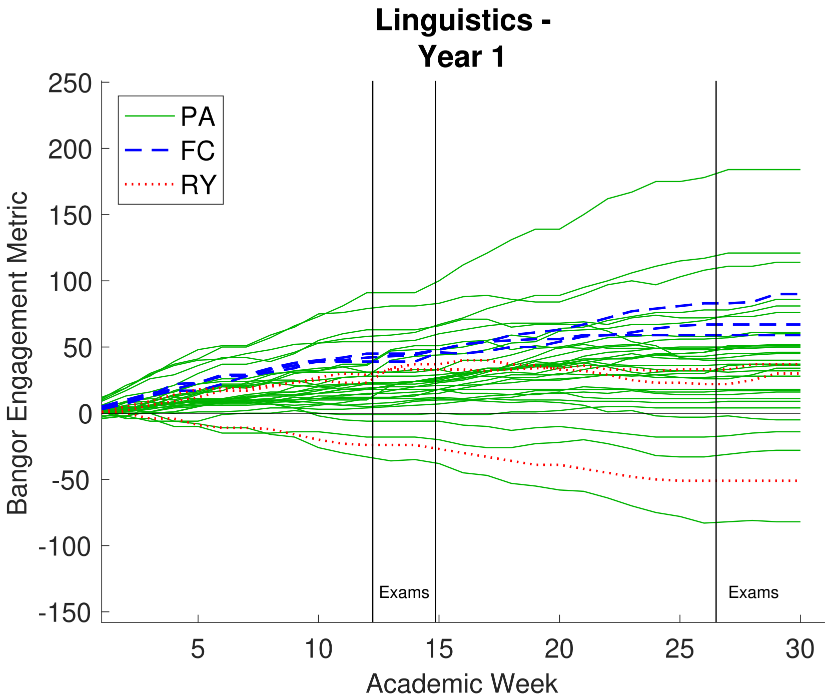
lead to comparable gradients between students. However, care must be taken to avoid direct comparisons between cohorts/programmes that have significantly different attendance or monitoring regimes.

Example Engagement Traces are shown in Figure 4.6, plotting attendance of two 2016/'17 First-year cohorts at Bangor University. This data was collected using Bangor's established process, ID card scanning in labs and lectures. This collection occurred before consideration of its use in LA. For these reasons, there should be no additional bias (positive or negative) present in the data. 'Additional bias' is used here because every student knows that they should be carrying their student ID at all times and attend every session.

However; not all do, and different modules/lecturers have slight variations in the application of school policy.



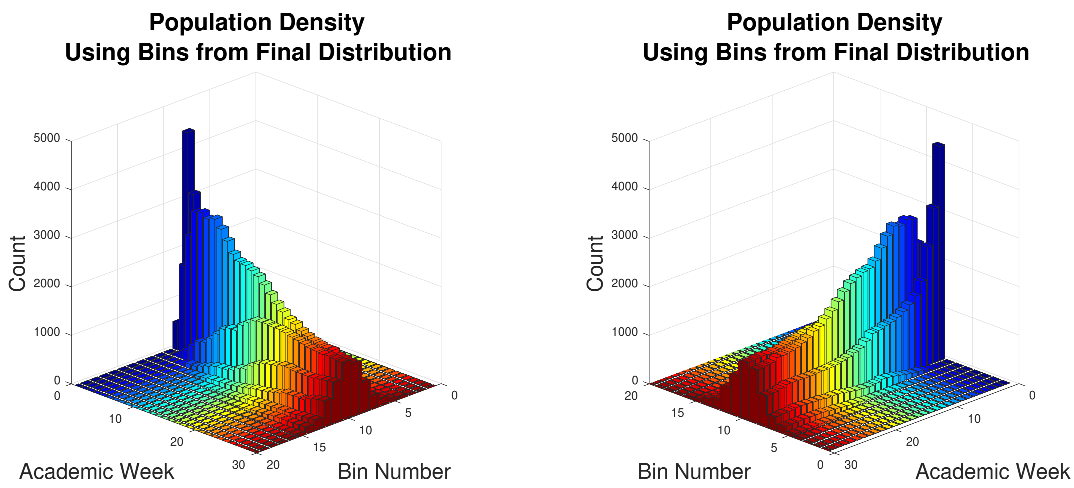
(a) Engagement Traces for the Year 1 Computer Science Cohort of 2016/'17.



(b) Engagement Traces for the Year 1 Linguistics Cohort of 2016/'17.

Figure 4.6: Each line in a plot represents one student. Each line is decorated to represent the Academic Standing of that student at the end of the 2016/'17 year. See Academic Standing Code in the Definitions section for the full definition and meaning of the codes.

The example plots show some common themes evident in many, if not all, cohorts. Firstly, those students that end the year with a negative BEM tend



(a) Projection 1 $(+45°)$ **(b)** Projection 2 $(-45°)$

Figure 4.7: 3D Histogram (from two angles) showing relative population density across academic weeks. Each week is divided into 20 linearly spaced bins calculated from the maximal range of the Bangor Engagement Metric. The bars are coloured by academic week.

to start disengaging early in Semester 1. Next, allowing for some minor perturbations, no student drastically alters their trajectories. These two observations provide critical insight for retention and tutor interventions. Early identification and swift action will be required to prevent these students from falling further behind or becoming more disengaged.

Separately, in first-year cohorts, there are always more extreme outliers. This effect is lessened in subsequent years. There are two potential causes for this. These outlying students may not be able to keep up their activity level in the face of more difficult or numerous tasks in later years — potentially through failure. Alternatively, the ability of their classmates has changed to lessen the difference between the group.

While the skew and mean of each weeks' population altered, Figure 4.7 shows that they remain an approximately normal distribution throughout the year. Semester 2, academic weeks 14-26, shows the most substantial diffusion. This may be due to the Easter/spring break, or it may be learning burnout [289] owing to the longer perceived spring semester.

4.5 Probabilistic Model

Having defined the BEM, it is possible to create an initial model showing the likelihood of an 'average' student attending sessions in each week. As students will either attend or not each session, attendance can be thought of as a Bernoulli process. Each decision to attend is mathematically independent and in essence random. Although, this is an assumption as for many students the planning of their time does depend on other sessions [306] or outside activities.

In order to calculate the expected value, first, the raw BEM value must be re-scaled to a value in the range of $0...k$ from the original range of $-k...0...k$. In these scales, k is the maximum number of sessions a student could attend. The BEM does not encode a maximum, so an assumption is made that the highest value is the maximum. The scale can be thought of as a line between two points: $A = (0, -k)$ and $B = (k, k)$. A is the minimum of the two scales, and B is the maximum.

The linear transformation is therefore obtained using the line equation between two points, derived in Equation (4.3). In these equations, x represents the Binomial variable scale, and y represents the BEM scale.

$$\begin{aligned} \frac{x-0}{B_1-A_1} &= \frac{y-A_2}{B_2-A_2} \\ \Rightarrow \frac{x-0}{k-0} &= \frac{y-(-k)}{k-(-k)} \\ \Rightarrow 2x &= y+k \end{aligned} \tag{4.3}$$

With the assumption that x is a Binomial variable, the probability of success is $P_{k\ trials}(attend)$. The expected value of x is, therefore, $x \sim B(k, P)$ and simplifies to $x \sim Pk$. The expected value of y, from the line equation, is

$y = 2x - k$. When the simplification of x is substituted, Equation (4.4) can be derived in terms of the transformed BEM scale $\bar{y}$.

$$
\begin{aligned}
\bar{y} &= 2Pk - k \\
\Rightarrow \bar{y} &= k \cdot (2P - 1) \\
\Rightarrow \frac{\bar{y}}{k} &= 2P - 1 \\
\Rightarrow P &= {}^{1}\!/_{2} \cdot \left(\frac{\bar{y}}{k} + 1\right)
\end{aligned}
\tag{4.4}
$$

Applying this formula, to the available data-sets, produces a result for each week and school. Each data point represents the probability that an 'average' student attends any session during that week. Figure 4.8 shows the full set of data. Figure 4.9 shows representative patterns from a selection of Schools. The representative patterns for 2015/'16 are more chaotic. These reflect a general trend toward lower attendance by the end of the semester. Each school has peaks during different weeks, most likely reflecting an aspect of the course design, such as assessment.

There is a glaring anomaly in the 2016/'17 plot, where students in the School of Welsh/Cymraeg appear to have returned en masse at Week 9. This is probably more accurately explained in terms of attendance monitoring policy. Most likely a change between taking attendance at the very start of a session (when students may not have all arrived), to the end where a higher proportion will be recorded. In the same vein, Bangor introduced new pan-University monitoring policies for the start of the 2017/'18 year. These have had the effect of normalising monitoring which shows as a more tightly banded plot.

Once the monitoring policy had bedded in (the 2017/'18 academic year), the patterns still vary in the exact probability however, they show a more consistent set of patterns. There are two exceptions, English Literature and Education. Students in the School of Education are often on placement during the school year, therefore only the students returning are recorded as being present on campus. It is hypothesised that the pattern shown by English Literature reflects the importance of scheduled lectures towards the end of the semester when texts are discussed.

It is also noteworthy that at a certain point multiple schools show no change in P between two weeks. (These are between weeks 5 and 6 in ’17/’18; 6 and 7 in ’16/’17 and ’15/’16.) This indicates that the mean of the whole-school population attendance and the maximum number of sessions either varied in the same way or did not change. In an educational context, the latter is far more likely. This provides data-driven insight into course structure; implying that no activity occurred in those weeks.

4.6 Attendance Feature Selection

The first stage of developing an ML-based predictive model is to investigate the predictive power of the metrics and data available. Prior work has already investigated possible options (see Sections 4.2 and 4.3 for more details) and selected those to be further investigated. This section evaluates their predictive power, both individually and in combinations.

To succeed in the primary goal of identifying students at-risk as early as possible, the values of the chosen metrics must be correlated with the resulting Academic Standing Codes. Ideally, this correlation needs to be as close to the beginning of Semester 1 as possible while remaining accurate. This task is one of feature selection. For each student, their Engagement Ratio and BEM metrics were summarised weekly. This provides the initial feature space; two sets of 30 readings in addition to the home school of the student, programme the student is following, the resulting Academic Standing Code (as the class label) and the student ID. The student ID should never be selected by the algorithms but is included to provide a unique reference for each instance.

The initial experiment utilised Sequential Forward Selection (SFS) and the Nearest Neighbour (1-NN) classifier, an approach followed by Kudo and

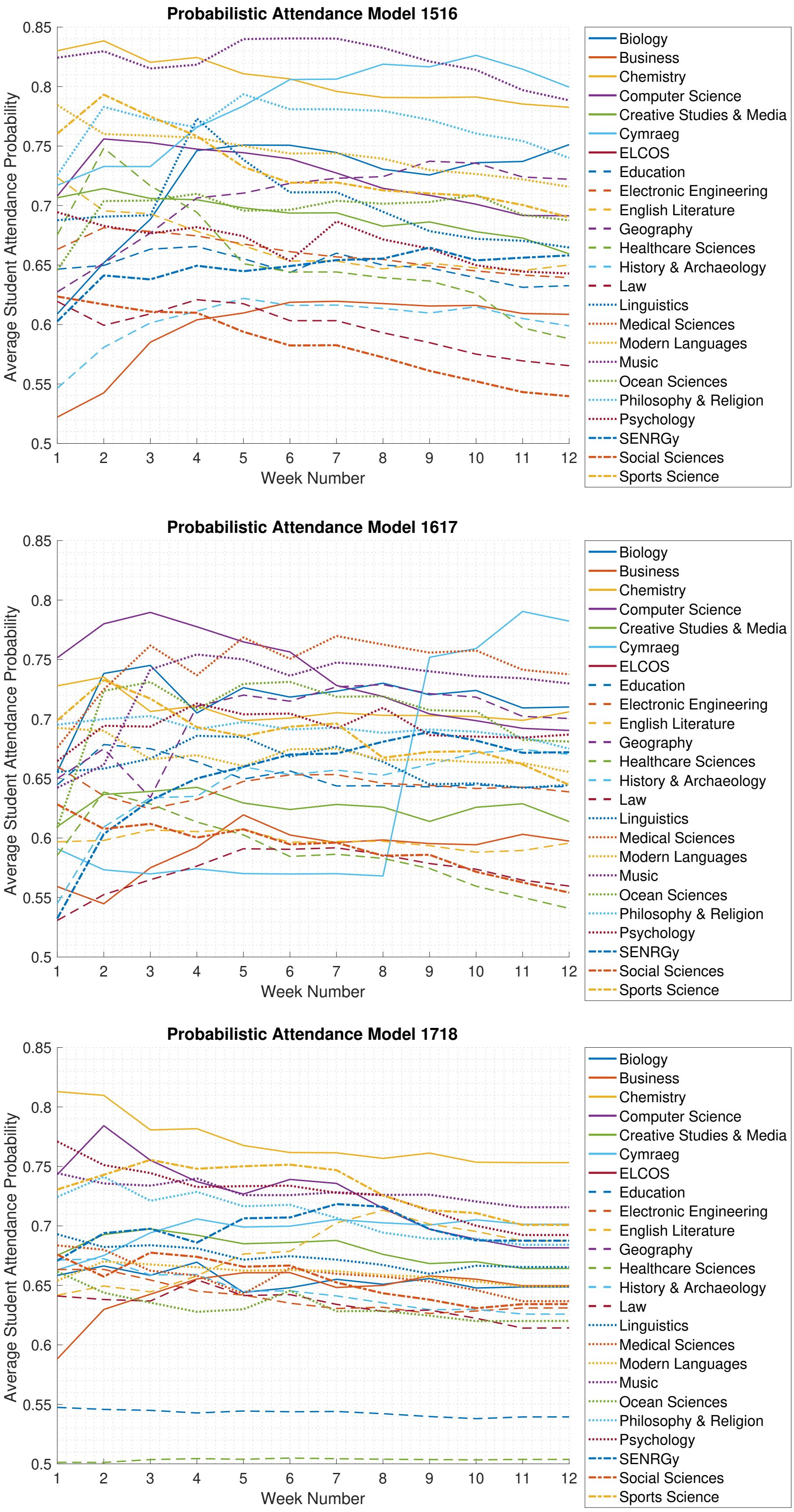


Figure 4.8: Reference plots showing the likelihood of an average student attending sessions in each week of Semester 1 in 2015/'16, 2016/'17 and 2017/'18. (School acronyms: ELCOS - English Language Course for Overseas Students, SENRGY - School of Environment, Natural Resources and Geography.)

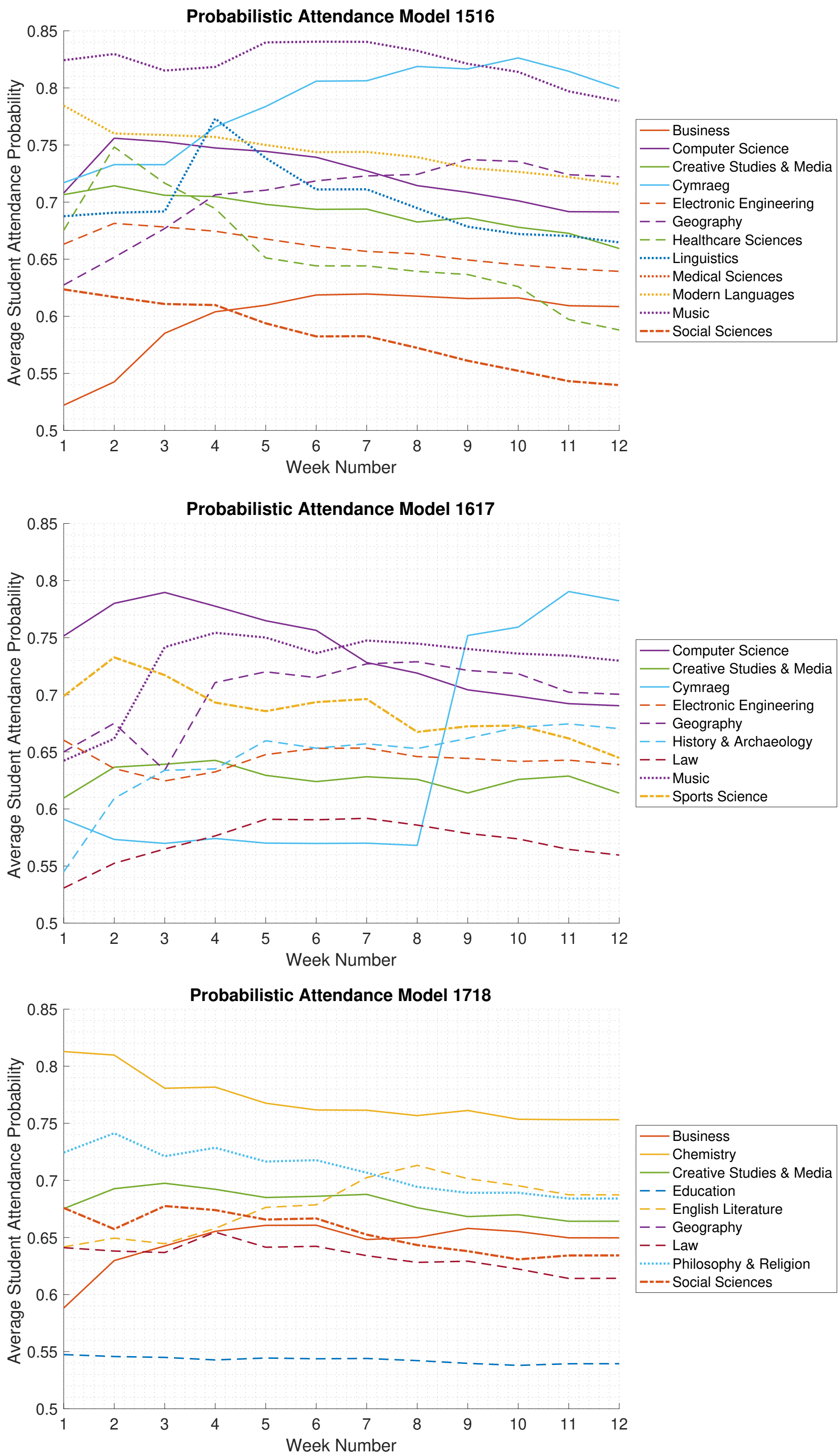


Figure 4.9: Selected plots highlighting specific attendance patterns across 2015/'16, 2016/'17, and 2017/'18. These patterns are chaotic, reflecting individual course structures in 2015/'16. There is then a normalisation during 2016/'17 where mandatory attendance monitoring was introduced. This results in a similar patterns in 2017/'18, with the notable exception of Education where students are most often on placement.

Sklansky [163]. The 1-NN classifier labels a new/test instance with that of its closest neighbour, using the Euclidean distance in the n-dimensional feature space [57] All of these experiments are conducted using the Weka Machine Learning Toolkit [128]. The Pass/PA class drastically outnumbers other classes; therefore, the F-Measure statistic is used to measure performance. As this is a multi-class problem, multiple F-Measures will be produced. By default, Weka uses a weighted average of these. The weighting is a proportion based on the number of instances in the respective class, with the entire total divided by the total number of instances. This method can cause an anomaly where the value is no longer between the precision and the sensitivity for the class. It also assumes that every class is equally important to the classification task.

The top four features, from the BEM set and in order, were found to be School, Week 4, Week 5, and Week 3. However, the top four from the ER set were found to be Week 29, Week 27, Week 6, and Week 19. These divergent results are the first indication that the two metrics hold different predictive power. A check of the BEM algorithmic result was made, utilising the 1-NN classifier and the selected features. Table 4.1 shows the results of this experiment. The headline result is a LOO-CV accuracy of 81.10%, and a weighted F-Measure value of 0.789. However, these results also show very low F-Measure values for the least represented classes.

Table 4.1: Experimental results using the 1-NN classifier and the top four algorithmically selected features. (TP = True Positive, FP = False Positive, Prec. = Precision, Sens. = Sensitivity, AUC = Area Under [ROC] Curve)

Class	TP Rate	FP Rate	Prec.	Sens.	F-Measure	AUC
Pass	0.931	0.814	0.873	0.931	0.901	0.596
Supplementary	0.088	0.052	0.150	0.088	0.111	0.588
Repeat Year	0.038	0.009	0.061	0.038	0.047	0.645
Fail	0.118	0.018	0.183	0.118	0.143	0.610
Repeat Semester	0.000	0.000	0.000	0.000	0.000	0.557
Weighted Avg.	0.811	0.703	0.770	0.811	0.789	0.596

This result (identifying weeks 3, 4, and 5) shows that early identification of potentially vulnerable students is possible. The accuracy of 80%, however, was less satisfactory. In order to ascertain which portion of the experiment was responsible, a second check was conducted using the C4.5 Pruned

Decision Tree classifier [240]. Decision Tree classifiers split the instances based on sets of rules. Only one feature is evaluated at each decision point, but there may be multiple branches for differing values/ranges of that feature. C4.5 Trees use pruning (removing of branches or sub-trees) to generalise the classifier. This process is known to reduce the accuracy on the training set, while hopefully increasing accuracy on test sets. The outcome of this experiment did produce better accuracy, to 84.85%. However, this classifier weakened the predictive power for students that would need to complete supplementary work, with this class' F-Measure dropping to 0.023.

As the algorithmic results pointed toward early weeks of Semester 1, the next set of experiments tracked the accuracy when progressively adding weekly data. These experiments used the same C4.5/LOO-CV combination; the results are shown in Table 4.2. The best overall accuracy achieved was 86.20%, which includes all of the weeks of Semester 1. This does not meet the stated goal of early identification of students. The second-best result, 86.10%, or just 0.10%/five students less, required only the first three weeks' values without the school or year of study included. Using this model, tutors would be in a position to potentially make appropriately targeted interventions from academic week 4.

Table 4.2: Results from testing successive week feature sets, using C4.5 Trees with Leave One Out Cross Validation. Rows are arranged by accuracy. See Academic Standing Code in Nomenclature for expansions of the class acronyms.

Weeks	School/Year Inc.	Accuracy %	Per-Class F-Measure				
			PA	FC	RY	FN	RS
1-12	Y	86.20	0.935	0.202	0.068	0.190	0.000
1-3	N	86.10	0.926	0.049	0.000	0.231	0.000
1-4	N	86.04	0.929	0.040	0.065	0.149	0.000
1-4	Y	85.77	0.929	0.151	0.019	0.178	0.000
1	N	85.75	0.923	0.000	0.000	0.000	0.000
1	Y	85.75	0.923	0.000	0.000	0.000	0.000
1, 2	Y	85.75	0.923	0.000	0.000	0.000	0.000
1-5	N	85.73	0.927	0.058	0.065	0.245	0.000
1-5	Y	85.71	0.929	0.141	0.075	0.211	0.000
1-6	Y	85.69	0.930	0.135	0.038	0.200	0.000
1-6	N	85.69	0.927	0.075	0.041	0.216	0.000
1-3	Y	85.65	0.928	0.139	0.000	0.110	0.000
1, 2	N	85.59	0.924	0.000	0.000	0.056	0.000
1-12	N	84.35	0.923	0.103	0.017	0.183	0.000

These experiments provided evidence that there is sufficient predictive power by using only the first three weeks' BEM values to provide a suitable model. It also shows that including discriminator features, such as year, school, or program increases some F-Measures. A further experiment is needed to determine which of these three discriminators is the most influential. This will result in a selected feature set of:

- Week 1 BEM Value
- Week 2 BEM Value
- Week 3 BEM Value
- School, or Program
- Academic Standing (Class Label)

When comparing the Engagement Traces of different Schools and Programs, there are noticeable clusters and patterns. A small sample is provided in Figure 4.10. For this reason, program and school are included in the set of features to be used in future experiments. This result is aligned with anecdotal evidence from educators and highlighted by Gašević [113].

4.7 Classifier Selection

Previous experiments failed to provide sufficient accuracy, overall nor for individual classes, with any combination of early availability features. Therefore, the next component to investigate is which classifier is used, with a standardised data-set.To this point, the success metric had been the overall classification accuracy, the number of students correctly matched to their actual outcomes. However,accurate classification into all five classes, while ideal, is not strictly required. Identification of potentially at-risk students is not contingent on which failure mode they may achieve, just that it is not expected to be a pass. With this extra constraint, it would be reasonable to reduce the problem to a two-class classification problem and re-categorise the student outcomes. However, it is still useful information for tutors to understand the severity of the potential outcome by mode.

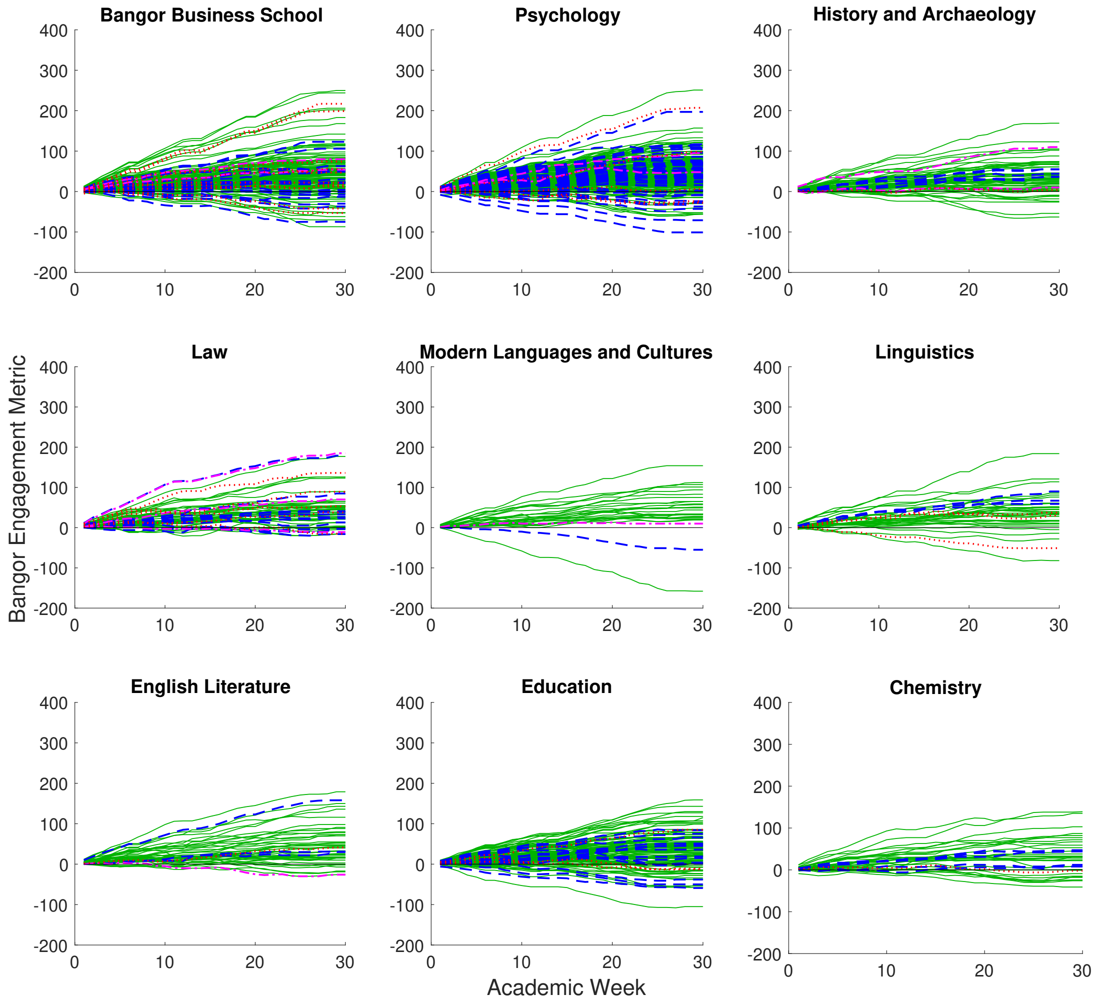


Figure 4.10: Engagement Traces of all First-Year Students by School. All axes are scaled identically to allow direct comparison. Note; each school has a distinct pattern/distribution of outcomes. A legend is not shown for clarity reasons, but coding is the same as previous Engagement Trace plots.

Confusion Matrices are the tool for reporting summary output in classification tasks, a simple table showing true/actual labels as rows and predicted labels as the columns. Each cell contains a count of instances/objects that fall at that intersection. The perfect outcome would be correct counts on the main diagonal showing that the true and predicted labels match. All counts elsewhere in the matrix indicate a classification error. There are two separate regions; above the diagonal showing false negatives, Type I errors, and below showing false positives known as Type II errors. Table 4.3 shows an example confusion matrix from the C4.5 Feature Selection experiment.

Table 4.3: Confusion Matrix from a C4.5 Feature Selection Experiment. Shaded cells represent 'problematic' classifications where a poor outcome would be missed.

↓ Actual / Predicted →	PA	FC	RY	FN	RS
Pass (PA)	4218	23	2	19	0
Supplementary[1] (FC)	425	28	2	12	0
Repeat Year (RY)	68	3	0	7	0
Fail (FN)	117	10	3	31	0
Repeat Semester (RS)	2	0	0	0	0

[1] A student achieving an FC/Supplementary Assessment, or more formally Conditional Fail, status would need to undertake supplementary assessments to pass their modules. In the U.S. system, this would be equivalent of Summer school.

The shaded cells in the table show classification results where students are identified as having a positive outcome when in fact, they would achieve a negative one. By re-framing the training of classifiers to minimise Type II errors, these inaccurate and misleading errors will be minimised as well. This adjustment was made as it is better to intervene with a student that may well succeed on their own than to misclassify a weak student as passing. Instead of minimising the raw Type II count, maximising the F-Measure score of all classes will tend toward the 'perfect' classifier.

One final set of experiments evaluated suitable classifiers using the initial data-set. Each experiment ran a combination of classifier algorithm, protocol (Resubstitution, or Leave-One-Out), and a cohort discriminator. This discriminator was either the Academic School or Degree Program. The use of the School and Year was examined in a previous experiment, but ultimately the Year did not provide any additional predictive power.

Selecting the final combination, became a multivariate optimisation problem. The F-Measure for all classes needs to be maximised, along with the final accuracy rate; while minimising the difference between the different protocols. These goals are set to create the best classifier while resisting over-fitting on the data-set. The complete results of this benchmark can be found in Table C.1.

All of the classifiers benchmarked have good performance with the majority class (PA). This is unsurprising as a pass outcome applies to 85.75% of the instances. Neither are any of the classifiers able to differentiate the two instances for the RS (Repeat Semester) class. The selection criteria must, therefore, revolve around the accurate prediction of the other failure modes. The F-Measure provides a surrogate for individual accuracy; therefore, the sum of the failure mode F-Measures can be used as a comparative metric, in the range $[0...5]$. The top six classifiers (ranked by this metric) achieve between 36% and 47% of the maximum score (1.878/5 to 2.342/5).

The top three classifiers all use the degree program as the cohort discriminator. These three also have more significant differences between the resubstitution and the LOO-CV protocols. This leads to the conclusion that 'degree program' is less appropriate when determining the patterns within a cohort and those of the preceding or successive cohorts. Excluding program combinations, leave Random Tree and Random Forest using the School discriminator. The classification of failure modes is within 1% of each other, whereas Random Tree has an 11% improvement with the pass class. On that basis, the Random Tree with School will be used as the classifier of choice. The McNemar statistical test (as applied to classifiers [165]), proves that there are statistically significant differences between the errors made by these classifiers at $p = 0.05$. This does not validate the choice, only that the differences are significant.

We believe that classifiers with a stochastic element can overcome local maxima/minima during training. This ability allows the classifier to provide a more rounded approach to the data-set. When examining the classification regions and trees produced by other candidate algorithms; the areas, and numbers of instances they represent, become too small to avoid misclassification.

4.7.1 Alternative Metric Experiment

As previously noted, utilising a percentage of sessions attended is a plausible alternative to the BEM. Using the same classifier and parameters, we conducted a companion experiment but using the proportion data-set instead. The results were gathered using both the resubstitution and LOO-CV protocols from the first three teaching weeks. When using resubstitution, the classifier achieves 98.28% accuracy with only three students mischaracterised under the 'on-mission' metric. However, when using n-Fold Cross Validation, it fairs 12% worse (85.90%) with 290 students misclassified. We can conclude from these results, that the classifier is prone to overfitting when using the proportion. It also achieves 7.66% less in overall accuracy. This would lead to more potentially unnecessary student interventions. While intervening is preferred instead of letting a poor outcome continue unimpeded, it may cause undue stress for passing students that did not require intervention.

4.8 Trial 1 — Backward Prediction

Accepting the parameters from the previous sections and experiments, a new classifier instance was trained using all student data from the 2016/'17 academic year. This data-set contains $n = 4970$ instances, with the same $c = 6$ features (ID, Week 1 BEM, Week 2 BEM, Week 3 BEM, School, and Academic Standing Code. These instances represent $C = 5$ classes, the full suite of initial Academic Standing Codes[1]. Note; that by training the classifier on the full set, rather than using the LOO-CV protocol, there may be differences in the resulting decision tree.

However, the testing of this new classifier is not completed on the same data-set (or any subset). Instead, the testing was completed with a previously unseen data-set from the 2015/'16 academic year. This was the first year that Bangor University had instituted more stringent attendance monitoring policies, therefore the earliest set that a reflective BEM value could be measured from. As this data was collected before the start of the LA work, there was no possibility of bias - intentional or otherwise.

The data has been processed, coded, and formatted in the same manner as the initial set. It contains; $C = 5$ classes, $c = 6$ features, and $n = 4877$ students. The proportions of the FN, RY, and RS classes remained roughly

[1] These are described as 'initial' as supplementary assessment can result in additional codes such as 'FS' for students that fail that work.

equivalent. However, the split between PA and FC was significantly different. This is due to the point in time where the 2016/'17 data was extracted, early July 2017. At that stage, the supplementary work had not been completed by students; therefore the FC class is over-represented in the data for 2017/'18.

The exact accuracy of the model on the 2015/'16 data dropped to 84.79% (4135/4877 instances), a difference of 8.77% from the train/test on the original data. This is understandable given the prior noted data gathering effect as well as cohort effects, and potentially a graduating class introducing different patterns that will no longer be present. These effects can also be caused by changes in lower education, filtering their way through the system. This would mean that the incoming first-year cohorts do not act in the same way as their predecessors.

However, when the model is evaluated against its primary goal, the correct identification of a potentially poor outcome, the accuracy rises to 97.33%. While the model will still classify students into one of the five Academic Standing Codes; for the evaluation, the outcome is reduced to a two-class problem - pass or some form of failure. By using this form of evaluation, 130 students within the institution would not be identified, and 583 (11.95%) students that would have passed unaided are flagged as potentially requiring intervention. This was deemed to present too high a potential risk to otherwise capable students. As a result, the experiment was repeated with an alternative classifier. The original evaluation discounted the RandomForest classifier as too unstable with changing data; however, it offers a significantly lower Type II error rate as well a marginal improvement in overall accuracy.

A new model, using the RandomForest classifier but keeping all other parameters the same, was created from the 2016/'17 data and again tested on 2015/'16 results. This model yielded significantly better results. The overall accuracy improved to 88.05% (+3.26%/159 instances), while the misclassification rate for otherwise passing students dropped to 8.65% (-3.3%/164 instances). Crucially, the model only marks an additional two students as passing when they would not, bringing the total to 132 not identified. This constitutes a 0.041% drop in effective accuracy.

4.9 Trial 2 — Forward Prediction

Having previously proven that the model created can achieve significant levels of accuracy, it was then applied to a genuinely predictive task. The task was to predict, at week 4, the outcomes of students at the end of the same year. This experiment generated three models of potential at-risk students, utilising the three combinations of data available. The first was trained solely on the data from the 2015/'16 academic year, the second using only 2016/'17, and the third trained on the combined data-set 2015/'17. All three models used the same set of parameters and protocols as in previous experiments and were tested against the unseen 2017/'18 academic outcomes.

The forward data-set contains $n = 5943$ instances, with the established $c = 6$ features and $C = 5$ classes. This is an increase in instances over prior experiments due to the increasing number of students at the University. All of the data has been coded and formatted as with the previous experiments. The proportions of each class were similar to those in the testing phase, as the data extract occurred after supplementary work had been completed. This is due to a change in process imposed by the institution independent of this study.

The exact match accuracy from all three models was consistent with that of Trial 1. When using our established 'On-Mission' accuracy, these rates improve into the low-90% range in correctly identifying students that would benefit from an intervention. The exact values are shown in Table 4.4. The combined model misidentified 345 students.

These results, when combined with the previous trial, highlight a potential cohortal effect. The longer a given model is in use, the more the anomalies/differences in student behaviour will distort the future

Table 4.4: Results of Testing using the various previously generated models with the forward prediction data-set taken from the 2017/'18 academic year.

	2015/'16 Model	2016/'17 Model	Combined '15-'17 Model
Exact Match	5132/5943	4872/5943	5025/5943
Accuracy	*86.35%*	*81.98%*	*84.55%*
'On-Mission'	5542/5943	5549/5943	5598/5943
Accuracy	*93.25%*	*93.37%*	*94.19%*

classifications. From Table 4.4 we can infer that the 2017/'18 cohort had more in common with the student body of 2015/'16 than the subsequent year. Extrapolating logically; the graduating class of 2015/'16 and the entering class for the same year share behaviour with 2017/'18's group. Upon progression, the graduating and the new entrants for 2016/'17 are the majority components, thus lowering the accuracy of the model. This result indicates that models will need to be retrained at a minimum of bi-annually to adjust to the shifting patterns/behaviours of the student body.

4.10 Validation

As a secondary validation exercise, the predictions from Trial 2 were provided to Directors of Student Engagement in each school. The lists showed all students from each school that were predicted one of the failure modes (FC, FN, RY, or RS) in any of the three models examined as part of the trial. The complete list comprised 425 students.

Three schools – Computer Science, Psychology, Sport & Exercise Science – replied to the request for comments and details of their independent identification efforts. In all three cases, the predicted list was a proper super-set of students identified by the school.

4.11 Summary

This work/chapter ostensibly answers research questions 1-3 (see Section 1.3). In order to best describe student attendance, it developed the Bangor Engagement Metric (BEM). This metric has been shown to remain a normal distribution in real-world data, lending greater statistical validity. The work has also shown that despite the BEM's simplicity, it shares no correlation with a simple attendance ratio.

By utilising this new metric, the author was able to derive a probabilistic model of student attendance from actual data. This model is consistent with anecdotal evidence from educators and allows data-driven decisions to be made on timing and course structure. The notable result is that several schools showed no change in the probability of attendance at critical points in the semester. This insight is fully explored in Chapter 6.

This BEM has been proven to have a 12% greater predictive power over other logical choices, notably the ration of sessions attended which is termed the Engagement Ratio (ER) in this work. Through a series of experiments, this chapter has shown that significant accuracy can be achieved using just the BEM values for the first three weeks of Semester 1. Further experiments tested different classifier algorithms to find the most appropriate to the attendance problem. There is a clear difference between deterministic algorithms and the stochastic algorithms which ultimately were proven to yield better results. The accuracy achieved is between 93.25 and 97.33%, tested with two sets of unseen data (predicting both past and future achievements). This only improves by using the full data of Semester 1, which is too late for educators to stage an intervention.

Chapter 5

Degree Pictures and the Student Journey

> *Data is a precious thing and will last longer than the systems themselves.*
>
> — **Sir Tim Berners Lee**

Previous chapters have shown how predictions can be made about a student's progress, and how even relatively small disruptions to student's patterns can produce significant effects. These operate in a microcosm, wrapped up in the specific situations occurring at that time. Students and examiners also need to reason globally about their academic situation and likely outcomes. Most often these discussions are supported by dense tabulations of the raw component and/or module marks. Therefore, a more intuitive system should be sought to present this information to both audiences.

The work has been published as; C. C. Gray and D. Perkins, 'Visualising the University Degree Journey', presented at the Computer Graphics & Visual Computing (CVCG) Conference, Poster Paper, Swansea, UK, Sep. 2018. DOI: 10.5281/zenodo.1475858 and accepted for publication as; C. C. Gray, D. Perkins and P. D. Ritsos, 'Degree Pictures: Visualizing the University Student Journey', *Assessment & Evaluation in Higher Education*, 2019, Invited paper. DOI: 10.1080/02602938.2019.1676397.

5.1 Degree Pictures

Within the Education and Behavioural Psychology fields, there is a well-established practice known as Precision Teaching. Pioneered by Ogden R. Lindsley [183] and developed over the next 13 years [182], this practice examines quantitative measures to evaluate and improve teaching interventions. The primary tool is known as a Standard Celeration Chart. This is an often hand- drawn graph measuring both positive and negative behaviour frequencies on a logarithmic scale. Lindsley's work gave rise to the term 'Learning Pictures', describing common patterns observed in these charts. In his (unpublished) Master's thesis, Pat All – supervised by Lindsley – described 17 unique learning pictures [8]. Each picture demonstrated an established pattern and an accompanying intervention.

Figure 5.1 shows the set of Learning Pictures describing students that are improving. This can mean an increase in positive/correct behaviours, a reduction in negative behaviour, or both. The most desirable of these is the 'Jaws' picture where both the conditions occur. In this case, All suggests the teacher monitor the situation and continue whatever practices have led to this situation. However, with the 'Uphill' picture the advice alters with educators being asked to teach specifically those aspects the student struggled with. This is because the student is already improving the desirable aspects on their own.

This work aims is to produce a symbolic and iconic overview of a student's progress. This view does not necessarily need to be accurate to the exact numeric percentage or mark but must be representative and comparable between students. These representations are then abstracted to form a category or similar patterns in the same way Lindsley, and All's Learning Pictures have been.

There are two use-cases envisioned for these representations. The first is during Boards of Examiners meetings, where the representation will assist by displaying achievement graphically. Following the acclaimed visualisation

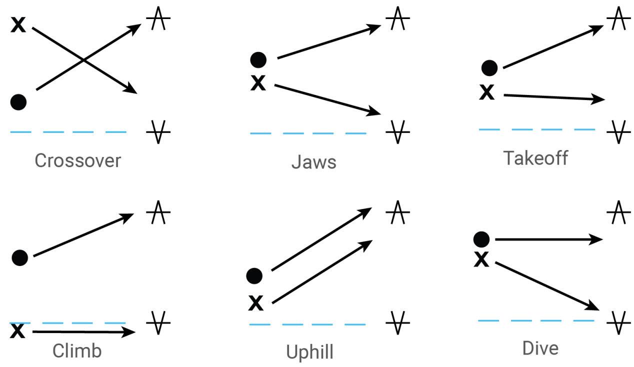


Figure 5.1: The 'Improving' set of Learning Pictures, Lindsley/All. Lines annotated with circles represent 'good' behaviours, and 'x' those that are undesirable. The arrows to the right of each chart, show the intended direction of the lines. Of these pictures, an educator would make more impacting interventions with an Uphill picture, as the undesirable behaviours do not move in the intended direction — image credit: Rick Kubina, Chartlytics.

mantra "overview first, zoom and filter, then details on demand" [279], the Degree Pictures will be coupled with an additional representation of the component marks and augmented by any other (usually textual) relevant detail. The second use-case is intended to assist with Pastoral Care systems. Personal Tutors are often called upon to offer advice and guidance to students, without much support on how to do so. This often results in a divide in the quality of responses due to varying factors, such as experience level. The Degree Pictures are expected to assist by allowing standardised responses and interventions. This should end the potential discrepancies of care, benefiting all students.

As a first stage, a formal requirements analysis exercise was completed. Input was integrated from relevant stakeholders, all experienced lecturers from within the School of Computer Science at Bangor. These requirements are all compatible with the aim and intent of this work. The stakeholders noted that they wished to see at-a-glance two points; the general standing of a student, and an indicator of any regulatory or special circumstances that needed to be discussed and considered.

As special circumstances can be dealt with as a highlight, the authors concentrated on providing the summary representation first. Following a process similar to the one suggested by [213], we generated a series of prototypes utilising simulated student achievement data. Most of these used the raw achievement percentages for each module and/or component. However, this was deemed too visually overwhelming for a summary representation. An initial round of testing, based on informal interviews with a subset of the interested stakeholders, showed that the most popular design drew inspiration from the UK Degree Classification system. This sees the institutional honours split into four classes; First, Upper Second, Lower Second, and Third. (There are other non-honour exit awards possible, but these are used as a last resort.)

5.1.1 Devising the Degree Pictures

Using a simple line-chart, we devised a set of fifteen potential paths through these classifications. These are shown in Figure 5.2, where each colour represents a majority or final classification. This set was created as an educated best-guess from the authors' experience of student achievement patterns. The x-axis represents time; there are six data points to each picture corresponding to the weighted averages at the end of each of the six semesters (for a 3-year undergraduate program). The y-axis represents their achievement level split into the classifications, or 'F' meaning fail.

We quickly came to realise that the y-axis (different achievement) was not the only 'movement' in the pictures that could occur. Students may enter or complete a pattern at different times as well, in effect sliding the picture left or right. This effect is incidental when examining an individual's performance for discussion and/or comparison purposes, however when deciding on an intervention - it can confound the issue. Therefore, the patterns were abstracted to relative rises and falls in a sequence. Borrowing from Lindsley's vocabulary, we termed these sequences/patterns 'Degree Pictures'.

To ensure that there were no edge or corner cases had been inadvertently missed, we constructed the charts from the data of Computer Science students' achievement over two years. This set was reduced from 118 students/individual patterns to 53 unique patterns, shown in Figure 5.3.

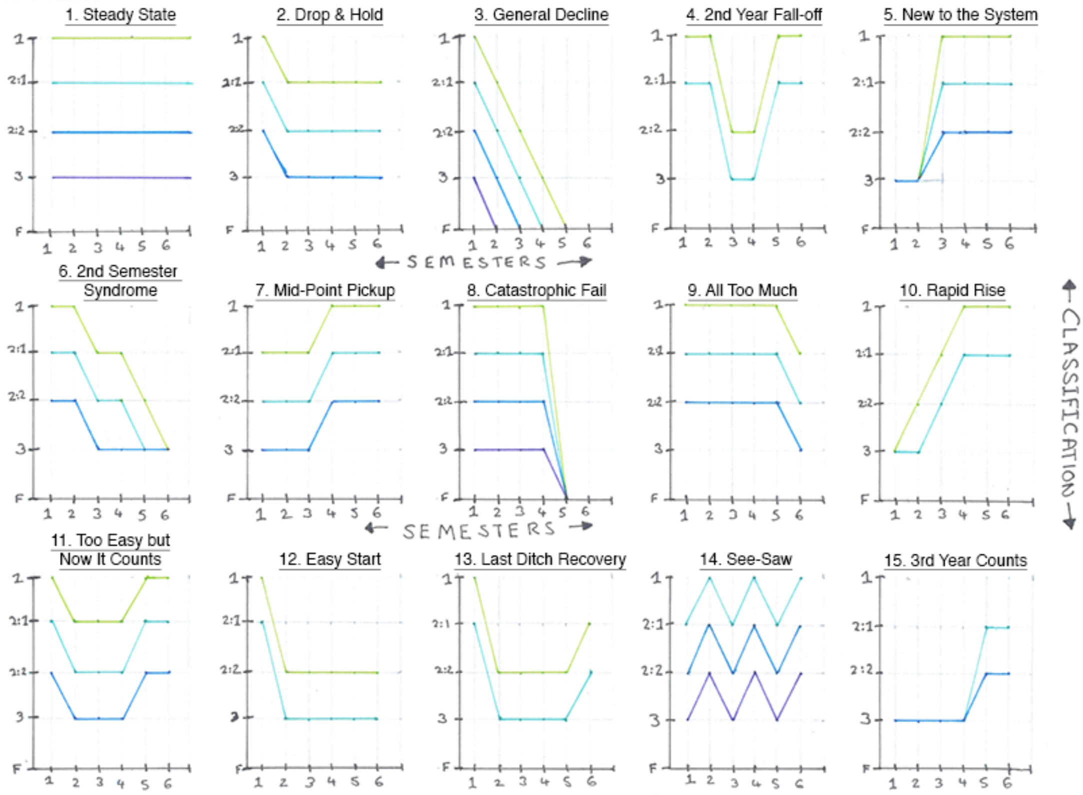


Figure 5.2: Simple line-chart representations of the initial set of patterns based on educator's intuition. Each line represents a similar pattern scaled by achievement.

These 53 were then grouped by eye, creating the 16 descriptions below from the common features.

1. **Steady State**

   The student, from the outset in Semester 1, achieves a level and remains at that level for their entire degree.

2. **New to the System**

   The student establishes a flat trajectory and remains there by the end of Semester 2; there may be a delayed achievement of this level.

3. **One-Step**

   The student achieves a given level, maintains that for at least one semester, then improves maintaining the higher level until the end of their program.

3b. **One-Step with Drop-off**

   As with a standard One-Step; however, the final year full average drops by one classification.

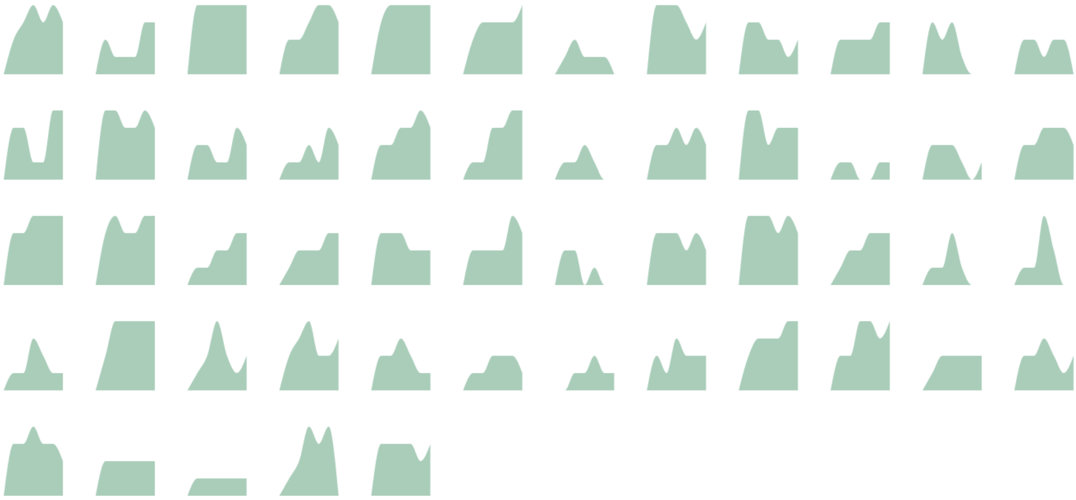


Figure 5.3: The 53 unique versions of the charts from all students in Computer Science over two years. These unique views do not account for translation (occurring earlier or later on the x-axis), nor magnitude (a larger or smaller rise/drop in achievement/on the y-axis).

4. **Two-Step**

   The student increases their achievement each year creating three flat aspects. This occurs as Semester 1 and 2 of each year will be at the same level.

4b. **Two-Step with Drop-off**

   As with Two-Step but includes a one classification drop in the final semester.

5. **Single Spike**

   At one point during their studies, a single semester has an upward spike. Their results return to the previous level the following semester.

6. **Spike and Drop**

   The student shows a single semester upward spike in their results; however, the following semesters drop further than the level before the spike. There may or may not be a flat section where the lower result is held.

7. **One Dip**

   The student achieves a given level and maintains it, except for a single semester where their achievement decreases.

8. **Unbalancing Dip**

   The student achieves a given level and maintains it, and a decrease occurs for one semester. The student recovers but only to a maximum of one classification below the previous level.

9. **Wake-Up Call**

   The student establishes and maintains a level, but then their performance decreases for a semester. Following the decrease, the student then establishes a new level higher than the previous one.

10. **Last Ditch Effort**

    This Degree Picture is typified by an at least two classification drop, with or without sustain and a final semester increase of one classification.

11. **Oscillating**

    The student appears to be on a boundary between two classifications, where there is a repeated pattern of at least two rises and falls.

12. **Step Down**

    The student starts at whichever level but will decrease by a single classification. They will then sustain that level for at least a semester. This pattern may repeat, as long as they do not drop far enough to fail their program.

There are two additional Degree Pictures, which were in the initial ideas but did not figure in the actual data. Both of these deal with failure modes.

13. **General Decline**

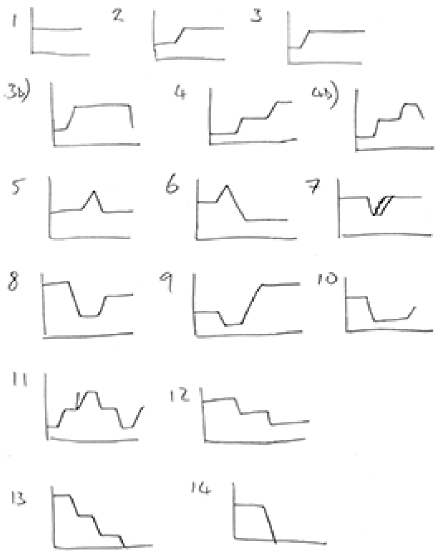
    The student begins a gradual decreasing trajectory, falling by one classification per semester until finally reaching a failing level. There may not be a sustaining step of one or more semesters.

14. **Catastrophic Fail**

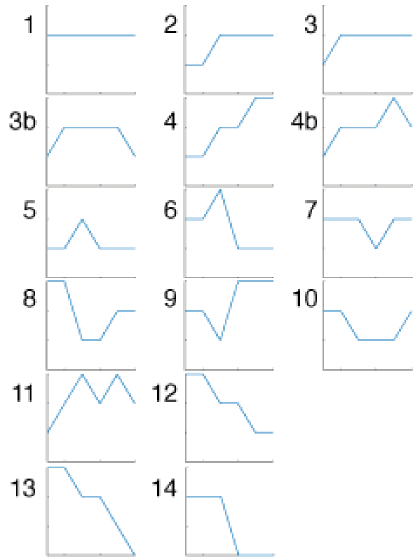
    Another picture would otherwise describe students in this category; however, at a single point, their achievement immediately falls to a failing level.

5.1.2 Validation of the Degree Pictures

As part of the definition of the Degree Pictures, we undertook a validation exercise using lecturers within the School of Computer Science, who were unfamiliar with the work. They were set the task to sketch a version of the Degree Picture from the description alone. All, but one (Oscillating), were correctly reproduced. Figure 5.4 shows a sample response compared with the idealised equivalent. Upon reflection, while the response does differ from the picture we originally intended, the sketch of the Oscillating picture is still valid and reasonable.



(a) A sample response from a lecturer in the School of Computer Science.



(b) Ideal representations.

Figure 5.4: Output of the validation exercise for the descriptions of all the Degree Pictures from Section 5.1.1. A prospective user was asked to sketch a graph from each of the Degree Picture definitions; they were told that the x-axis would represent time and the y-axis achievement/classification. These sketches (a) are then compared to the ideal representations (b).

Once these definitions had been decided, the unique set of pictures were extracted from a different discipline (Chemistry). Again, these unique pictures were grouped by eye. The same set of sixteen archetypal patterns emerged.

5.2 The Student Journey

Almost all academic assessment relies on the subjective judgements of the educator; there are, however, notable exceptions such as in mathematics. Educators continuously strive to reduce the impact of this to promote fairness. They employ rubrics, mark schemes, multiple-choice assessment, and other mechanisms in the pursuit of an objective individual grade for each student [63].This process, while potentially time-consuming, can work on an individual assessment or module basis. As students progress further through their programmes, more and more subjectivity creeps in [34]. In UK Higher Education (HE), examiners are granted broad powers to make necessary adjustments when making progression (end-of-year) and award (end-of-programme) mark confirmations [42]. These adjustments are indented to provide a balance to mitigate exceptional circumstances, such as pastoral crises or missteps during delivery or assessment. Each HE institution has its particular regulations, practices, customs, and methods when considering final grades, adjustments and award outcomes.

Visually, extensive tables of similar data do little to aid the viewer to distinguish rows and columns. This a recognised problem in the Information Visualisation field, with solutions already proposed [242]. Banding, division lines, colours, or more exotic methods add emphasis to assist the reader, but fundamentally they are left to interpret the data on their own. These measures are possible but often lack these elements due to printing costs (these reports are usually printed as low resolution black and white) or lack of support for the features within the software used. Where these tools are available, the support for visualisations is limited to simple charts [28]. Furthermore, any commercial offering would undoubtedly lack the customisation needed to encode all nuances of the institutional environment, regulations, etc.

There is a move toward progress tracking within Learning Analytics applications. Examples include Mastery Grids created by Loboda et al. [184] and Study Paths created by Busler and Semmler [47]. Most visualisation remains basic, using bar charts to show achievement [89] on an assessment-by-assessment basis. (A more complete evaluation of techniques in use can be found in Section 3.2.3.) A University of Tennessee student proposed in their thesis [81] a method of tying some elements of a course and student data together. DeCotes' method, however, focuses on unifying student cohorts rather than individual achievement.

Without any systemic support from tools, each examiner must follow a similar process; assimilate the data being presented, reason around it, apply (sometimes abstract) regulations and arrive at a final grade. While disagreement, with the correct intent, is healthy; this individual process will vary in focus, effort, and conviction [337]. To fully consider any student's effort, examiners may need to refer to the full set of courses/modules rather than just those in the final year. The assembled quorum of examiners relies on a limited form of crowdsourcing to smooth out individual variances to arrive at a sound decision [62].

5.2.1 Design Constraints

Following consultation with select staff within the School of Computer Science and Bangor, mixed with the authors own experiences of Boards of Examiners; the following design constraints were defined. This is the set of initial criteria and can be further refined where necessary.

There are two main events this visualisation/system has to support. The first is known as the 'Progression Board' meeting. This meeting is a formal meeting of all teaching staff responsible for awarding marks to students in all but the final year of their programmes. The function of this board is to confirm the marks awarded and check that each student has met the institutional requirements to progress to the next stage of their studies. The second is the 'Final Award Board of Examiners'. In this meeting, academics involved at any stage of the programme confirm grades and confer awards. This meeting is where the majority of regulations are examined against each student case. At both of these meetings, abbreviated details of the exceptional circumstances are presented. Usually, this is a severity grading and the period the situation spanned. Besides, examiners would need to be able to recall historical performance to judge whether adjustments would be required.

In both cases, there will be a significant number of students to review and pass judgement on. This necessitates a symbolic, obvious and intuitive display, enabling examiners to make quick initial judgements.

However, in more difficult cases the full set of results must be easy to access to facilitate further discussion. Additionally, how student data is organised becomes more important the more extensive the cohort population becomes. Examiners will need to be able to select groups of students to deal with in addition to singling out individuals.

The level of detail will also need to be split; between an overview for high-level consideration, and a full-detail view when required. As any discussion could be fast and free-flowing, the switch between the two would need to be as efficient as possible. As most of the detail will lose some of its meaning without the overall context, the overview should still be visible when examining the detail.

5.2.2 Data Dimensionality

There are multiple sets of parallel data that this system will need to be able to handle. Most obvious is the students themselves; the system will need to be able to present individual students. Given the sorts of decisions that need to be made by examiners, it would also be useful to present groupings of students. These groupings may be by programme, cohort, achievement or an arbitrary set made by interested parties.

The data describing a student's accomplishments also exist in at least three levels of detail: the macro – overall achievement; mid— a single module; and micro – an individual component or assessment. The two different user groups (progression and final award), will be interested in different elements of these levels of detail. Therefore, both should be available through the system.

Each of the accomplishments/marks also has a weighting component, how much does that individual data point contribute to the given module, programme, and overall. These multiple levels of contribution information also need to be represented to allow examiners to reason about the implications of any issues they wish to address or changes they may wish to make.

The entire process hinges on local regulations, codes of practice, and guidelines. While much of these are homogenised across the UK HE sector, there will still be quirks bespoke to any implementing institution. These include but are most likely not limited to classification boundaries, pass marks and credits, numerical rounding, supplementary assessment, and special/mitigating circumstances. Each of these will need to be present in any proposed system, and ideally, be easily customisable.

5.2.3 Design/Ideation Process

Initial design efforts were focused on how to best visually depict the dimensionality of the data necessary. This work took the form of hand-drawn sketches, with notations of the ideas behind them (see Appendix D). Broadly the techniques are split into four categories by purpose: grouped paths, hierarchical, multivariate, and single paths. These designs, while perfectly valid from the Information Visualisation view, fundamentally misrepresented the intended focus of the tool. As a result, the process was revisited to clarify the aims and requirements. The basic premise is to represent clearly and graphically the relevant information about students when making progression and/or award decisions.

The Degree Pictures plot ideally represents the longitudinal view of the student's overall progress. This can be used as the situating overview, appealing to System 1 thought (see Section 2.3). The low amount of visual detail allows the shape to be easily absorbed and a rough judgement made. Coupled with the Brushing technique [46], which allows the user to select or highlight an area of interest, the interaction completes the remainder of Shneiderman's mantra – 'filter and details on demand'.

One possible analogy for the weighting of modules and a student's performance within it is fluid flowing through a pipe. A larger pipe representing a more highly weighted module, with the amount of flow through the pipe representing the achievement. While non-traditional, for education, this analogy is intuitive. There are two classes of visualisation that are designed to represent flow or movement (in a non-geographic setting). These are Sankey Diagrams [247], and Parallel Sets [26]. As Parallel Sets can appear more cramped with multiple levels, it was decided to use Sankey Diagrams as a basis.

The remaining factor to be considered is the inclusion of regulatory aspects of the decision. This includes whether the student gains enough credits, has special circumstances to be considered, or would require supplementary assessment. Combined with the student's performance, these aspects can require deliberation and action from the Board of Examiners. Any tool should also provide these prompts.

A prototype was created using these principles and comments solicited from expert users. These users were members of the School of Computer Science with particular expertise in visualisation or the institution's regulations. The comments received surrounded the use of the Degree Picture curve. The users were concerned that a single continuous line could misrepresent the nature of the underlying performance. In response, an additional block overlay was added. This uses straight lines with any step after the data point to show the lag of results until the end of the next period.

5.2.4 The Student Journey Tool

As a result, the final tool becomes a combination of three different visualisations. The interaction with the overview bar links these three, which makes the tool an implementation of Multiple Coordinated Views [251]. Figure 5.5 shows the final version of the tool.

The user selects the student concerned using the three combo-boxes at the top. These adapt to the previous selections, narrowing down possible selections to avoid overwhelming the user. Once selections are made, the user would then press the Find button (the icon of a magnifying glass). Once details are retrieved from campus data sources, only the overview Degree Picture is drawn. It will include circular markers providing an annotation wherever special circumstances exist that the Board would need to consider.

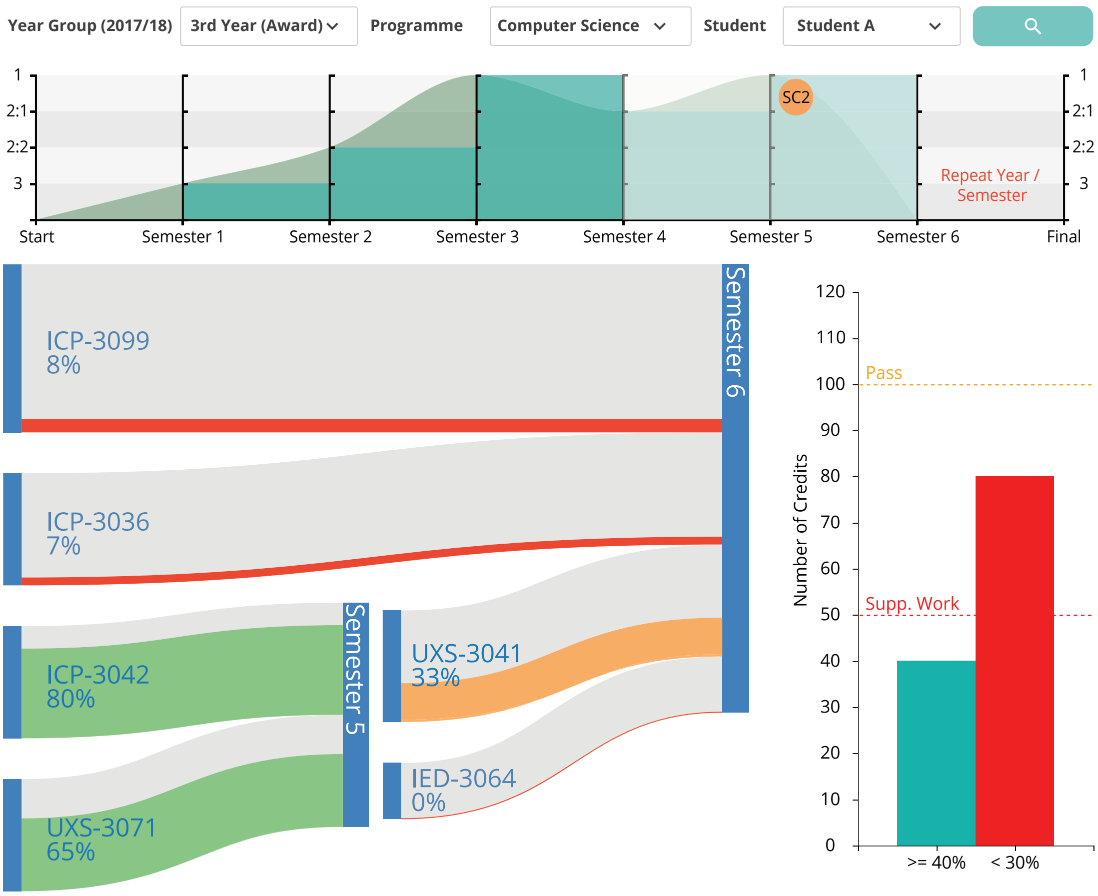


Figure 5.5: The final layout of the Student Journey Tool, including the three co-ordinated views and controls to allow users to select which students to view. The student's third year results are selected for detailed view.

The far right section of the overview chart contains the final weighted average. For final year students, this will show their overall degree classification; for other years it shows the recommendation to the board regarding the progression of the student. In both cases, the regulation logic is built into the recommendation. For example, Bangor's regulations include processes for considering border-line cases. These cases are when the student's average is within 2% of a classification boundary — 38%, 48%, 58%, and 68%. The Board are instructed to award the higher classification when certain criteria are met, for example when $^2/_3$ of final year module scores are in the higher range. Without this logic, examiners would need to be fully aware of all applicable regulations and manually apply them. Adding the logic stops the Board from having to rely on a member spotting these situations to avoid potentially making detrimental decisions.

In the event of further deliberations, the user can select any academic year by clicking on the overview chart at the desired point. This action reveals the two coordinated views, the one on the left showing all constituent modules and scores within that year.

The main view is a modified Sankey diagram. This shows the flow of achievement towards the common end of the academic year. Each bar is scaled according to the number of credits that the module is worth. This indicates the relative importance of a module to the result. Within each bar is an additional coloured section representing the actual achievement of the student. These bars have three primary colours; red for scores which are considered an outright fail, orange for results considered ‘condonable’ (usually 30% to 39%), and green for passing scores. A bolder hue of each colour is selected for scores within the border-line of 2%, indicating results that may be significant to these further deliberations.

The view on the right shows the regulatory situation in terms of credits required. In Bangor University’s case, there are two such applicable regulations. Those that govern the necessary credits to progress from level to level and that govern eligibility for supplementary assessment. As these regulations set out thresholds, the most appropriate graphical representations are simple column charts with thresholds marked. Through the use of colour, the simple view can provide an instant iconic view of what options are available to examiners.

When combined with the Small Multiples Degree Pictures view, this tool presents a powerful, visual analytics platform allowing every examiner to reason about progression and award decisions at multiple levels of detail. These range from the cohort to individual performances in each module.

5.2.5 Evaluation

A preliminary evaluation was completed using the standard System Usability Scale (SUS) [40]. A total of 16 examiners, familiar with the institution’s practices but not all regulations, were surveyed. This specific form of questions, using a Likert scale arrives at a score out of 40, but is commonly multiplied by 2.5 to achieve a score out of 100, but is not a percentage. Subsequent testing of the scale (using over 500 trials) has established that the average score is 68 [266].

Figure 5.6 shows the overall result of this evaluation, with Table 5.1 showing the detailed breakdown. The SUS questions alternate in their forms, with odd-numbered questions requesting a positive affirmation and even ones a negative. The Student Journey visualisation scored an average of 79.65 / 100 on the SUS. This score places it on the Good/Excellent boundary [19]. The cut off for excellent is 80 / 100. Examining those results ranking the tool below average ($n = 5$), the average score was 63 / 100. This places the tool in the marginal section of the scale.

Table 5.1: Results of the SUS Evaluation of the Student Journey Visualisation. Each cell lists the number of respondents selecting that option for each question. Odd-numbered questions have a positive phrasing, meaning 'strongly agree' is best. Even-numbered questions are negatively phrased, therefore disagree is the desired answer.

Question	Strongly Disagree	Somewhat	Neither Agree nor Disagree	Somewhat	Strongly Agree
1	1	1	2	6	6
2	7	5	3	1	
3			3	4	9
4	8	2	3	2	2
5			1	10	5
6	8	5	3		
7			3	4	9
8	9	7			
9			4	3	9
10	6	3	6		1

Respondents were also given the opportunity to provide free-form comments. Most of the negative-leaning remarks were a request for additional training/materials rather than suggestions or complaints. One positive comment states that this tool was able to show the situation with a student which matched, almost precisely, examiners' intuitions.

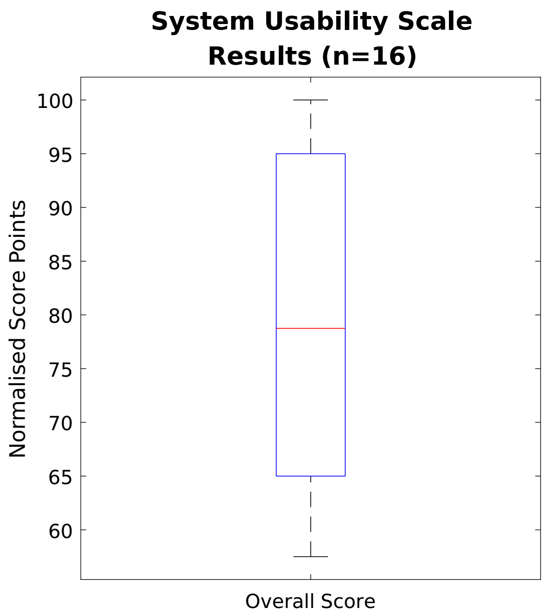


Figure 5.6: Box Plot showing the overall, normalised total score for the Student Journey tool using the SUS methodology.

5.3 Summary

This chapter presents a new iconic method for presenting student performance. These Degree Pictures follow on from Learning Pictures used in Precision Teaching. This work has defined sixteen differing patterns (when magnitude and delay are abstracted), that encompass the progress of a student. These Degree Pictures provide for a standardised benchmark of performance by time, allowing students to be compared directly, irrespective of course or tenure. While the initial set of pictures were derived from educator intuition, this work found that they correlated with actual data strongly. The final set was then refined using the patterns uncovered from real student performance data. The descriptions of the patterns have been validated using a blind test. A tutor within the School of Computer Science, unfamiliar with the pictures, was asked to sketch a graph for each description. In all but one case, the sketches exactly match the reference picture.

These standards could be used to develop rule-of-thumb pastoral care responses to best aid students. This would end the lottery students unwittingly participate in when being assigned a personal tutor. Institutions would be able to provide a baseline level of pastoral care, derived from many interventions but still applying to a given student – based on their Degree Picture.

The Degree Pictures methodology was then incorporated into a tool for use by Boards of Examiners when making progression and award decisions. It provides a longitudinal overview of student attainment as part of the Student Journey tool. This tool provides cohortal, student, academic year and module levels of detail of achievement. There are specially designed visualisations to appeal to System 1 thinking, where the viewer makes a snap decision based on low fidelity information. The detail is then filled in allowing System 2, or deep reasoning, about what the visualisation is representing. Paired with details of Special Circumstances, that may impact students, this tool can provide recommendations on what action a Board should take.

This work adds a new dimension to Research Question 5, although not specifically about attendance. The provision of a visual analytics tool for achievement can be coupled with the BEM, Engagement Traces, and the predictive model to provide insight into how a set of results has been achieved. This tool is primarily designed to provide an example of how technology can assist educators in making the best decisions possible. Evaluation of this tool using the System Usability Scale proved that it could assist educators, providing one answer to Research Question 6.

Chapter 6

Disengagement Insight and Nudge Behaviour Change

> *The greatest moments are those when you see the result pop up in a graph or in your statistics analysis - that moment you realise you know something no one else does and you get the pleasure of thinking about how to tell them.*
>
> — **Emily Oster**

During the development of the predictive model, patterns were observed that did not appear to follow expectations. Existing University engagement systems still flagged students, that registered strong opening attendance patterns, as at risk. The work presented in this chapter investigates this situation and the resulting insight into student behaviour as well as the inferred structure of the affected cohorts.

This work has been published as; C. C. Gray, 'Don't Disturb the Student - Investigating Pattern Disturbances with Student Attendance', Presented at AdvanceHE Surveys Conference, May 2018, [Online]. Available: `https://www.heacademy.ac.uk/download/surveys-conference-2018-programme`.

6.1 Methodology

The work in this chapter attempts to use a data-first approach to isolate and explain the cause of this discrepancy. Instead of attempting to find a known event in the data-set, this work attempts to identify candidate events within the data, then explain them as a secondary activity. This work will be based on the predictive model findings. The work will begin to answer research question 5; 'Can this model be used to provide additional insight into student attendance patterns?'

This will involve identifying appropriate numerical and statistical analysis methods and tools. Each candidate will need to be correlated to an academic, social, or another physical event. From there, the impact of the event will be calculated by determining which student groups should be naturally affected, excluding them and identifying the remaining set. If possible, the driving factors and motivations will be identified, and suggestions made as to how the collateral impact can be lessened or removed.

6.2 Problem Description

Following on from the Forward Prediction Data Experiment (see Section 4.9), a set of 62 Computer Science students were identified as having poor engagement[1] by the end of Semester 1. None of these students were identified by any version of the ML-based predictive model. The questions raised are;

- Why were these students not identified?
- What characteristics do these set share?
- Can the model be refined to take these into account?
- What do the characteristics reveal about the student population in general?

[1] Poor engagement is defined as attending fewer than 60% of the required sessions by Bangor University.

6.3 Initial Investigations

After identifying the students concerned, the first step was to retrieve their Engagement Traces. These are produced in the same manner as for the example students in Figure 4.6. These new traces can be seen in Figure 6.1.

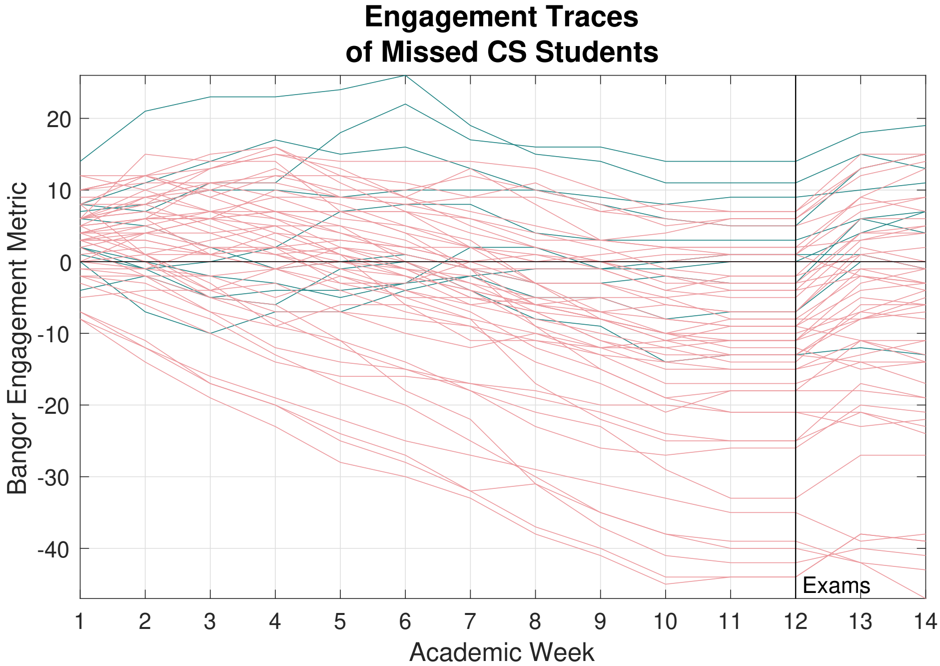


Figure 6.1: Engagement Traces of Semester 1 2017/'18 for the 62 students not flagged as potential risks but ended with a low engagement measure. Each line represents one student, teal traces show a positive gradient at week 5 and pink ones a negative gradient.

Visually, viewers can discern a general downwards (less engaged) trend in all but a few cases at week 5. These are plotted in pink in the figure (for highlighting purposes only). This observation provides a starting point for interrogating the data. While this is a small sample from just one School, the reasoning follows that an event or combination of events has caused this change in behaviour. Next, the entire data-set for the academic year needs to be checked for the same pattern.

Each student will naturally develop individual patterns of attendance. From the visual inspection, the week five disengagement appears to be an anomaly from the prior pattern. Investigations began around deviations from this pattern. The entire data-set is split by academic school, then the second

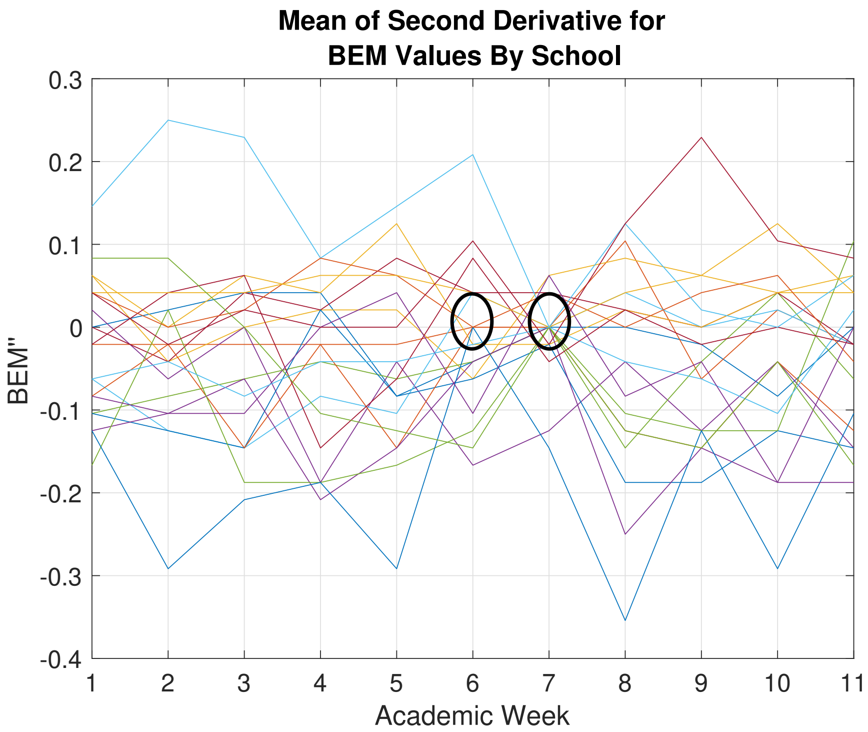


Figure 6.2: Plot showing the mean change in change (2nd derivative) of BEM values by academic school. The ellipses highlight two clusters of potential inflexion points. For this illustration, the colours identifying individual schools are not significant.

derivative is calculated. The mean of each school's weekly values is taken to produce a summary. These values are then used to produce the plot shown in Figure 6.2.

When a second derivative equals zero, there is a possible inflexion point. This would indicate that the mean weekly differences begin to alter direction (from positive to negative or the inverse). Clusters of these events, when summarised by school, indicate a more substantial disturbance in student behaviour as the effect is felt across multiple schools. The timing also coincides with the observations made from the 62 disengaging students. While still circumstantial, these findings provided enough grounds to proceed with further investigation.

6.4 Disturbances Due to Reading Weeks

The previously identified disturbances in student attendance, coincide with the Reading Weeks held at Bangor during Semester 1 of the 2017/'18 academic year. Some of the Reading Weeks at Bangor are fixed, i.e. occur in the same week each year; but others occur in different weeks due to external factors such as ocean tides. In 2017/’18 the main Reading Week was Week 7, which explains the larger cluster of zero-crossings for this week. The secondary Reading Week, observed by the sciences rather than the arts, occurs the week before (week 6).

The next aim, having demonstrated an observable disruption from the data, is to quantify how large an effect it has on the student body. The descriptive statistic ‘variance’, or the squared difference to the mean – measures relative spread in a population at different measurement points. This study measures the population (students’ BEM) at weekly resolution, so therefore our independent variable is time. Just as found with the ML model, each School has different patterns; as a result, the population has been subdivided by School.

The study hypothesises that with all other variables being equal the proportion of above and below average students should remain approximately constant. Therefore, the mean BEM value is calculated for each subpopulation and a count made of students above and below that mean. The result is shown in Figure 6.3.

From there, it is a trivial transformation to calculate the mean and variance of both above and below average counts. The cumulative result is shown in Figure 6.4. In order to smooth any extreme outliers, a three-week rolling mean has also been included on the plot.

This figure allows us to draw some conclusions without further analysis. Focusing on the initial weeks, the pattern of variance stabilises for weeks 4 and 5. This further lends credence to the assertion that the end of week three sets attendance patterns. Comparing the initial weeks (Weeks 1-5) and the final weeks (Weeks 8-12), there is a lasting change in the variance, i.e. it does not return to the pre-disturbance levels. In between these ranges, there is an evident change around week 6 and 7 Reading Weeks. Based on the

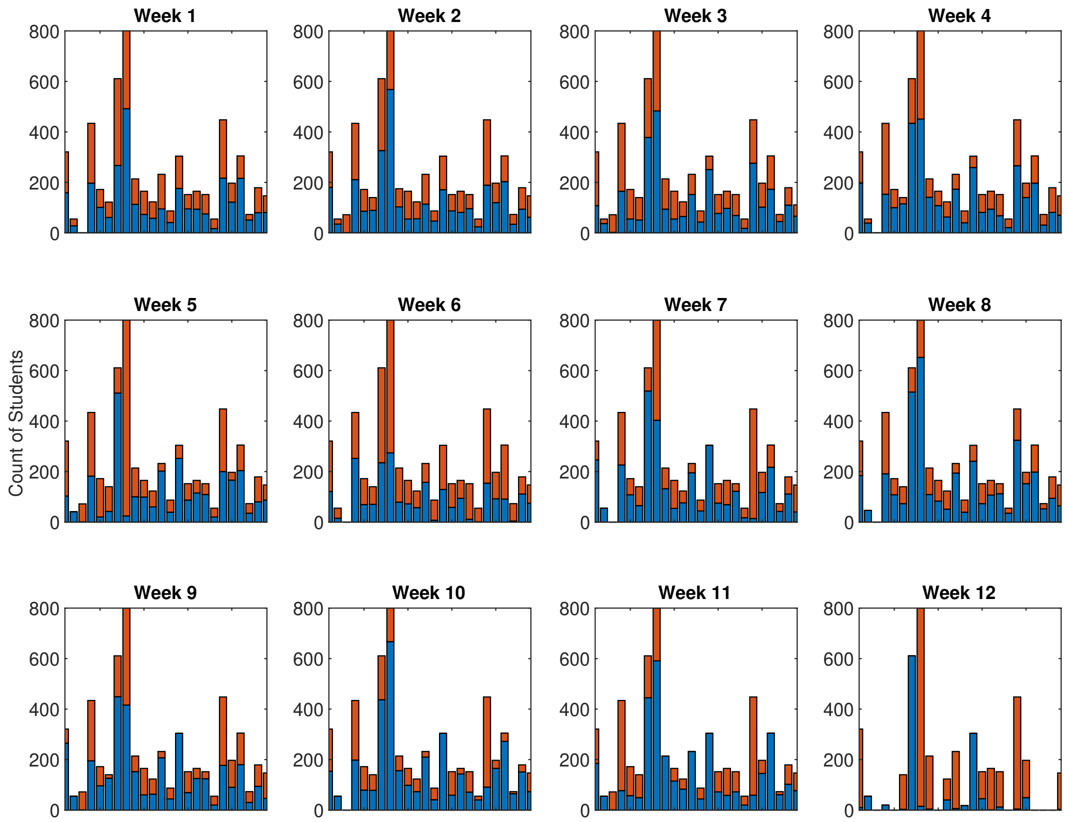


Figure 6.3: Plot showing relative counts of above and below average attendance by week and school. Each stacked column shows a single academic school. Blue bars represent above average students, red below average.

relative values, it is also possible to conclude that above average attending students are affected less than their less well-attending colleagues.

The data-set was split into two groups, using local knowledge, those schools that observe Reading Weeks and those that do not. The same statistical process was followed to produce Figures 6.5(a) and 6.5(b).

These figures show two fundamentally different student behaviours. Students from schools that do not have Reading Weeks show a spike in variance at week 3, before stabilising. Unlike their colleagues that do observe Reading Weeks, these below average students show a higher likelihood of variability from week four onward. This effect is caused by a larger variance in attendance of below average students in the last weeks of the semester.

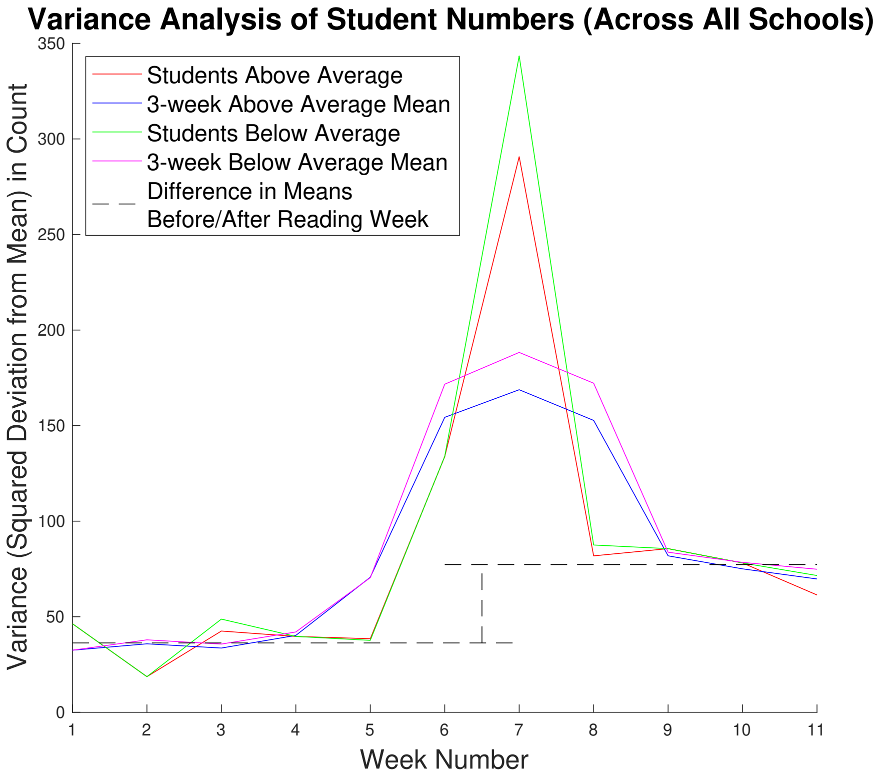
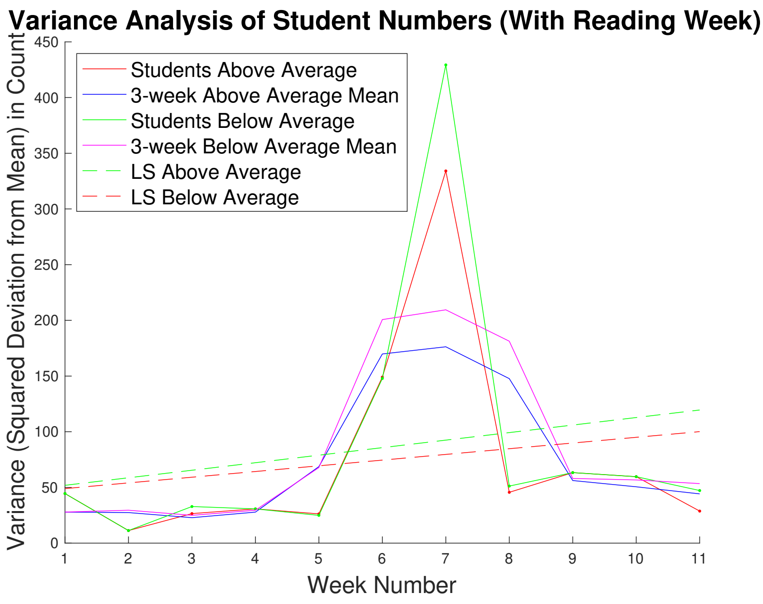


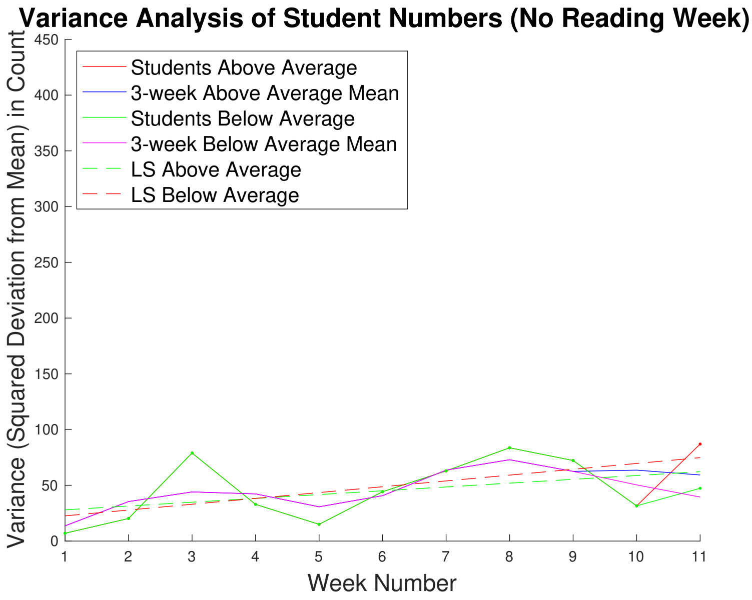
Figure 6.4: Plot showing the calculated variance between the counts of above average and below average attending students, across all schools. Note the step in variance levels before and after the disturbance (shown by the black dash lines). This indicates that more student patterns are changing after the disturbance than before.

Both observing and non-observing schools experience more variability toward the end of the semester as shown by the gradients of the trend lines. However; the effect of disruption, in weeks 6-8, is longer-lived. The variance metric does not return to near-prior levels until week 10. After which they alter due to the end of the semester. Therefore, there is a 'short and sharp' alteration in patterns where Reading Weeks occur, with affected students understanding that they are expected to return. Students that do not have Reading Weeks are affected by a smaller degree, but the effect drags on for significantly longer.

This can be seen in a comparison of the Least Square trend lines, as shown in Figure 6.6. The gradients of both below average cohorts remain approximately equal. This indicates that Reading Weeks have little effect on students that were already engaging poorly. However, this is not the



(a) With Reading Weeks



(b) Without Reading Weeks

Figure 6.5: Plots showing the variance values in the count of students above and below average at each week of Semester 1. Also included is the Least Squares trend line for each data-set.

same for their above average classmates. Reading Weeks, where they exist, introduce more variation in above average students.

These findings mirror those from a longitudinal study tracking 748 students through 22 modules [216] at the University of Glamorgan, UK. Their study set out to investigate attendance directly, whereas this work is focused on detecting the anomalies in the data. The Glamorgan study uses descriptive analytics, only stating what occurred. The author's work is more aptly

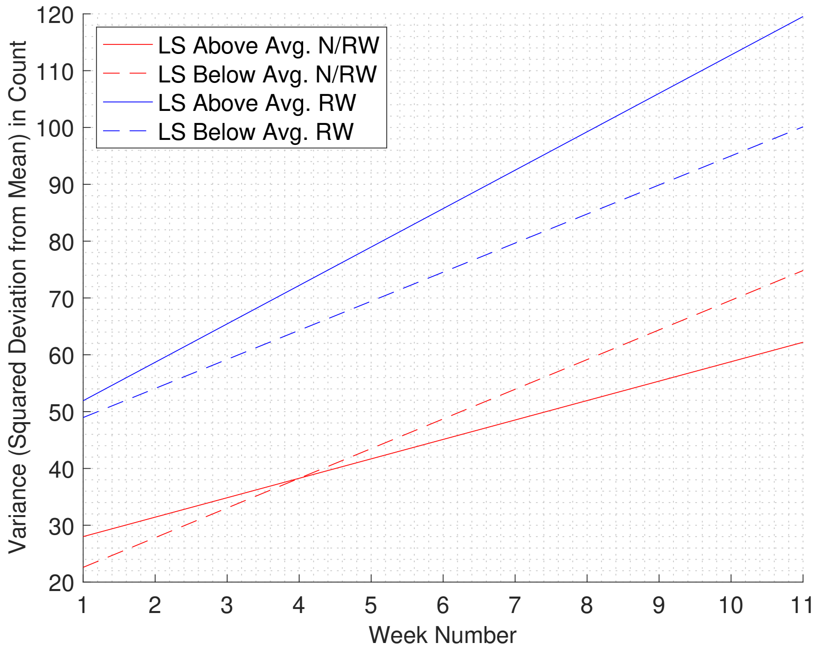


Figure 6.6: Plot showing the comparison of the Least Squares Trend Lines for Schools with (blue) and without (red) Reading Weeks. Solid lines show variance in the Above Average set of students, dotted lines the below average students.

described as insight analytics, as it attempts to explain why a particular event occurred.

6.5 Negative Nudges

Nudge Theory has most recently been popularised by Thaler and Sunstein's book 'Nudge: Improving Decisions about Health, Wealth and Happiness' [299]. The lessons from this book have been applied in the UK [325], the US [186], and the EU [187]. With each forming a group to study how public opinion and behaviour can be meaningfully altered using Nudge Theory. These groups are more formally referred to as 'Behavioural Insight Teams', with each having a nickname within the respective societies; the 'Nudge Unit' in the UK and 'Nudge Squad' in the US.

Classically; a nudge, derived from behavioural economics, has been held to mean positive reinforcement given when an individual performs an indirectly suggested action. There have been two high profile, and reasonably successful nudges. The first was introduced at Amsterdam's Schipol Airport [96]. A small picture of a housefly was etched into each urinal, the aim to cut down on cleaning costs of 'misses'. The study cites that 60-70% of users aim for the fly preventing splashing. The second, from Stockholm, added an audio system to stairs alongside an escalator at the exit of a metro station [151]. The escalator study reported a 66% increase in usage of the stairs. In order to qualify as a nudge, as with the examples provided, the desired activity must be easy to avoid in all senses. An intervention that places a tax, penalty, or extra cost on people that do not select the desired behaviour cannot be considered a nudge.

Intuitively, there is also the opposite phenomena, interventions and activities that encourage undesirable behaviour choices. These have been termed 'Negative Nudges' [70]. Negative nudges have been taken to include situations that place an undue penalty on the target when choosing the desired choice. This concept needs to be distinguished from 'Dark Nudges' [215]. Dark Nudges are sometimes considered a corruption of Nudge Theory, where the same motivations are abused to cause targets to make a desired but wrong, and sometimes harmful, choice. For example, online retailers add a ticking clock, or an erroneous 'stocks limited' notice to encourage purchases there and then.

There is little quantitative data available to analyse why students choose to attend or not attend class. Van Blerkom attempted to explain some of the reasons for non-attendance in a Psychology course [312]. This study found that most often extrinsic factors and rewards determined attendance or participation. Logically, we can infer that the same is true with the behaviour exhibited in this case. If all factors were intrinsic to the course, module, or student group then there would be no contamination in other areas of Bangor University. It is also widely acknowledged that one of the most influential external factors influencing student behaviour is peer impression/pressure [284].

If it is assumed that this will hold for attendance generally, there is a potential explanation. Bangor is a non-campus university; its buildings are spread geographically across the city, with most students needing to travel between their accommodations and school. Before the 2017/‘18 academic year, the University would only guarantee places in Halls of Residence to first-year students. This meant that continuing students would band together and rent private accommodation. These groups, anecdotally, tend to be centred around social groups rather than necessarily courses or disciplines. As a result, students are exposed to differing schedules as part of their domestic lives. When one part of the household no longer maintains a routine, another part may lack the motivation to continue with theirs. The negative nudge may not be as conscious as the previous example; it may be due to the student no longer being woken up by the general morning noise in the house. As a result, the student oversleeps and misses a scheduled activity.

6.6 Attendance as a Habit

A daily routine is, in most senses, the most ingrained of social habits. If the routine of attending work, lectures, laboratories, or any other event, is treated as such these insights are supported by a body of Psychology research. The book Psycho-Cybernetics [195] first popularised the idea that any habit takes 21 days to be formed and continued as automatic. When this idea was revisited in 2009, Lally et al. [170] found the situation was not as definitive. Their investigations showed that the interval to establish a habit is far longer; the shortest time, in their study, was 18 days.

This finding is supported by the LA data presented in this chapter. Figure 6.4 shows the variance in student numbers above and below average stabilising at week 4 with little change in week 5. After this, the pattern is lost due to the disturbance. This coincides with the time-frame that, the Psychology literature states, is prime for habit formation.This may be as a result of the prospective memory task [92], intending to attend. As this action has not become an automatic conditioned response yet, the intention can be forgotten.

Research conducted by the sports social network Strava [22] suggests that New Year’s resolutions surrounding physical activity are most likely to be

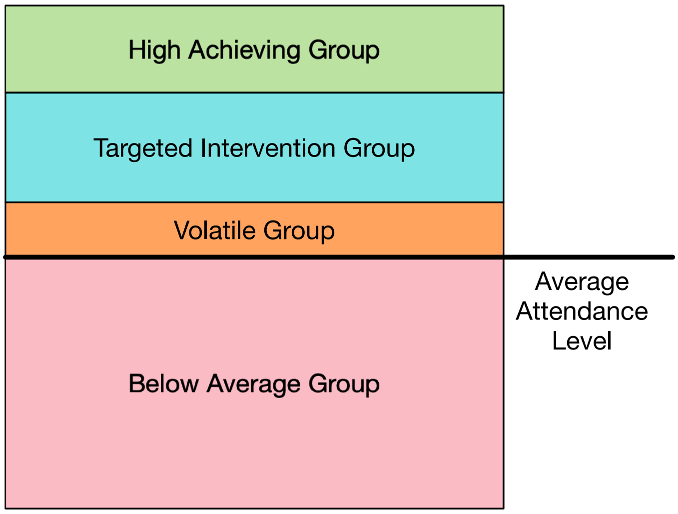


Figure 6.7: Graphical representation of hypothesised the structure of student cohorts, based on responses to disturbances in session patterns.

given up 11-14 days after the pattern is started. As a result, the second Friday in January is now colloquially known as ‘Quitters Day’. If this same timescale is applied to student attendance, this also fits with the initial observation that attendance begins to change from Week 5 – two weeks after the attendance pattern is established in Week 3.

6.7 Cohort Structure

It is possible to imply a structure within the student body by using the trends and data uncovered. This structure is not necessarily a complete or entirely accurate model of all students or cohorts. It should, however, be representative enough to allow educators and administrators to make informed judgements on their practices. Figure 6.7 shows these divisions graphically.

First, all of the ‘below average attendance’ groups show variance trends of approximately the same gradients. This suggests that these students have already chosen not to fully engage/disengage. Therefore, these students are unaffected by changes to class schedules caused by Reading Weeks or any other factor. This is a generalised assertion, but subsequent investigation shows that there were timetable changes (such as a change of venue) and/or one-off cancelled sessions affecting at least one module for most if not all students.

Second, anecdotally and based on results analysed in other sections of this thesis, there is a ‘attends everything’ group in all cohorts. Educators affectionately remember these students as ‘the good ones’. These students are hallmarked by close to, if not perfect engagement; higher than average marks; and intra-session engagement, such as answering questions. Due to their personalities [75], it is unlikely that any temporary disruption or change would influence their engagement. This group are those likely to have made attendance at all sessions a habit, and unlikely to forget an intention to attend.

This leaves the set of students that are between the average up to just below the high achievers. This group is most affected by disruptions and changes to their attendance patterns. Due to the two different gradients of for the LS Trend of the above average group of students, it is hypothesised that there are two sub-groups. One group more closely aligns with the high achieving group and the other with the below average group. The distance between the subgroups grows over the semester. This effect accounts for the increasing variance for above-average students observed in schools with Reading Weeks.

6.8 Potential Responses

Previous sections have explored three potential explanations; less extrinsic motivations/rewards, some form of negative nudge, and the habit of attendance not being fully established. Anecdotally, and based on the insight gained with this data, it is hypothesised that for the Targeted Intervention Group motivation is the correct response. This is because these students are largely engaged but stumble at some point in the module. Whereas the Volatile Group is more likely to be affected by nudges as their habit has not formed.

Educators must, therefore, respond differently to correct the undesired (non-attendance) behaviour. A simple pastoral meeting with a member of the Targeted Intervention Group providing a ‘pep talk’ is most likely sufficient motivation. However, with a Volatile student, interventions to assist with time and life planning may be more successful. The author acknowledges that this is a generalised response and that there will be intricacies/complications with every individual. The key is determining which group each student belongs to.

The insight revealed by this work is not to be taken as automatic that reading weeks are bad or harmful. Rather, what this reveals is that any alteration to the established pattern can cause issues for a subset of the cohort. However, if the change is integrated into a course, then it may well serve as the much-needed extrinsic motivation to keep that same subset interested and engaged.

6.9 Limitations

The insight available for this sort of analysis is only as valid as the input-data-set. At the time of this work, noone was aware of the analysis possible which rules out the potential for tampering or bias. However, the point still stands in other ways. There are two potential reasons for no entries being recorded for a student; either no activity occurred, or there were no events with attendance taken. If the second option occurs, then the BEM for the affected cohort remains static.

When this work was presented to the Directors of Teaching & Learning across the institution, several raised objections that they did hold sessions during Reading Weeks, most of these sessions were field trips, placements, visits, etc. and were attendance monitored. In these cases, this analysis would not flag as much of an effect, if any, for those schools. This is due to the method of detection. If the input BEM values remain the same, then the number above and below the mean value will, by definition, also remain the same. This leads to the variance of the counts remaining the same.

Two schools within Bangor University do not exhibit these traits. They are the School of Chemistry and the School of Psychology. The analysis does not reveal any quantitative reasons for this. Using anecdotal local knowledge, the School of Psychology uses their Reading Week for an activity called Psychology Oral Presentation Practice Sessions (POPPS). In this activity, each student is expected to prepare a topic then present it to the others in their group. While this is a difference in the pattern, students are still engaged with their course. This may serve to minimise negative nudges for this cohort. There is less obvious evidence to explain the exception of the School of Chemistry. The cohort is significantly smaller than with other schools (116 total students, 46 postgraduate and 70 undergraduate) which may make it a more insular community. A closer-knit community would mean that the routine is reinforced including any changes. It would also help resist any other outside influences.

6.10 Summary

This chapter presents an anomaly from Chapter 4, where a set of students were not identified as a result of the model. In an effort to determine if there was a flaw with the model, data-centric methods were used to identify the source of the discrepancy. This identified a critical pattern that there was a significant disturbance to these students' attendance patterns at week 5 of Semester 1. The initial discovery did not explain where this disturbance originated; therefore the investigation was widened.

Using the statistical variance (squared difference from the mean) in numbers of students above and below the average attendance level, and the second derivative (or change in change) of the BEM, this work identified significant events during weeks 6 and 7 of the semester. Again, this does not explain what the event was - however integrating local knowledge these were identified as Reading Weeks.

The subsequent exploration of schools within Bangor University that both do and do not observe a Reading Week showed that there were quantifiable effects on both sets of students, leading to further insights on student reactions to these disturbances. As a result, this work answers research question 5 (‘Can this model be used to provide additional insight into student attendance patterns?’). The methods used may also provide additional insight when provided with a different question/premise to start with.

By making deductive inferences and applying the results of additional studies/research, this work also presents new ideas for educators to consider when designing and delivering their modules. These do not make any judgements on the appropriateness or benefit of including Reading Weeks within a curriculum. There is evidence that inclusion without a structure or set of activities does harm student engagement. Based on the data observed, this work has been able to project a structure of the student bodies. This structure reveals where educators can best target their efforts. It also aligns with anecdotal beliefs from educators but is now supported by empirical data.

This work has also demonstrated that these effects can spread wider than the observing schools, contaminating engagement of other schools and programmes. These observations are one potential answer to research question 6, and this chapter also offers examples of potential changes/responses educators can make.

Chapter 7

Conclusions

> *People do not like to think. If one thinks, one must reach conclusions. Conclusions are not always pleasant.*
>
> — **Helen Keller**

The original aim was to create a single overarching Learning Analytics (LA) system to provide both the identification of at-risk students and further insight. While completing this work, it became apparent that a single system is much too large an undertaking, requiring far more components than this project was able to create. Instead, the work created four components that could be combined into a cohesive system at a later stage. These are:

1. a probabilistic model for attendance by school and week;
2. a Machine Learning (ML) model to identify at-risk students from early in Semester 1;
3. an iconic, generalised, graphical representation of student performance throughout their studies;
4. a tool for examiners to review student progress, gain insight, and receive recommendations as to the likely decision for a student building in institutional regulations.

As a result, the overall hypothesis has not been definitively proven. However, all six research questions have been answered, this work has met the seven objectives set. The following sections will highlight key findings for each part of this work that provides those answers and potential avenues for future work.

7.1 Answers to Research Questions Posed

Section 1.3 set out six research questions that were needed to test the overall hypothesis. Through the three pieces of linked work presented in Chapters 4 to 6, these questions have been answered, with some limitations, however. This section is presented in a narrative (rather than numeric) order.

1) What elements of attendance data can be used to predict student outcomes? 2) Can a suitably accurate predictive model be built and maintained from this data? 3) How close to the start of an academic year can the model yield sufficiently high accuracy rates (e.g. > 90%)?

The Machine Learning (ML) experiments conducted in Chapter 4 found that student attendance can be used to predict their outcomes. This conclusion is supported by current literature as well. This work examined how early these predictions could be made through a series of trials. The results show that the highest prediction rates are obtained using the entire set of attendance readings for Semester 1. However, as this will only be available after students have left for their Christmas holidays, tutors would be powerless to intervene. The evidence suggests that a student facing difficulties will probably be less likely to return after Christmas. This chapter has shown that using a synthetic metric, the Bangor Engagement Metric (BEM), 90%+ accuracy can be achieved using only the first three weeks of attendance data. This affords educators a far larger window to plan and execute interventions to recover students that may otherwise have been a retention or failure statistic. This method has been proven with current, future, and past data. There is a 5% range of accuracy across these data-sets, indicating that cohortal effects have influenced the result. As a result, the work recommends a two-year lifespan for any iteration of the model. It is freely acknowledged that these results were obtained for Bangor University only. However, the method was generalised enough to work across academic subjects. The author would expect similar results to be obtained in other institutions subject to the same method, metric and ML parameters being used.

However, there is a question that arises from the development of the BEM; ‘why is this metric more powerful than a simple ratio of sessions attended?’ This work proved an empirical 12% improvement using the BEM over the Engagement Ratio (ER), but has not provided the numerical analysis and mathematical theory as to why this may be. Such additional work is out of this project’s scope but may indicate future directions for other synthetic metrics and analytics possibilities.

5) Can this model be used to provide additional insight into student attendance patterns?

Chapter 6 presents insights that were uncovered during the validation/testing of the ML model that forms the primary body of this thesis. The results of the model were compared with the previous method of flagging students employed by the University. This comparison showed 62 students not identified by the ML model. These students exhibited a similar trait of disengagement from week 5 of semester 1. This work applied a data-first approach to trying to isolate potential causes and/or explanations for this effect. Utilising the variance in the count of students above and below each School’s average BEM value in each week, a pattern of large variances was observed between weeks 5 and 8 of Semester 1.

This pattern entirely coincides with when some parts of Bangor University observe reading weeks. In the year analysed, 2017/’18, there was a two consecutive reading week window as the Arts and Humanities week did not overlap with the Science week. The pattern observed highlighted several key findings. First, those attendance patterns stabilise (indicated by the lowest variance in above and below average students) in week 3 or 4. This confirms the finding from the ML experiments that week 3 is a significant point in the semester. Second, attendance starts to vary the week before a reading week and continued for a week after showing a prolonged disruption to students' attendance. Third, that after the reading week period was over the variance did not return to the prior (lower) level. This indicates a continuing effect within the student body. This work did not seek to categorise any effect as good or bad, which would require a longitudinal study with that express aim. However, it proves that additional analysis of attendance data can reveal new insight.

4) What results need to be conveyed to educators to best place them to intervene and in what form do they need to be presented? 6) How can the findings be used to aid educators in deciding the best course of any subsequent action?

Arguably the two most challenging aspects of education are assessment and care for students. This is in the most part due to the subjective and individual nature of these efforts. Systems and processes exist to try to limit this variability, but it is not possible to erase it. LA can provide recommendations to limit the impact further, mainly as these recommendations will be entirely based on the data available. The work contained in Chapter 5 has presented two related visualisations that summarise and present student achievement with time.

Based on the work on Learning Pictures, from Precision Teaching, Degree Pictures are a set of sixteen archetypal plots representing the possible achievement paths at each semester. These pictures can be used to make initial judgements using the 'fast mind'. Being both symbolic and standardised, educators would be able to craft best practices for intervening with students exhibiting deteriorating patterns. When used as the basis of the Student Journey tool and combined with individual module results, more guidance can be offered to educators. By building in details of institutional regulations, this advice is of particular use to examiners. The tool can provide a definitive recommendation on the course of action for every student. This may be to recommend the type of award or intervention based on special circumstances.

This work has presented more advanced forms of visualisation; namely correlated views, small multiples, and the Sankey chart. Using these methods has enabled more dimensions from the underlying data to be shown in a single picture. Having more of the dimensions depicted allows more complicated visual reasoning to be performed without switching between views. A seamless environment also means that the viewer does not need to remember information from one view to the next. Eventually, these new visualisations would be integrated into a single unified Learning Analytics system. Such a system would present the historical - performance; the current - attendance; and potential future - the prediction from a ML model.

7.2 Implications for Higher Education Practice

There are three main implications that Learning Analytics (LA) will have on the practice of teaching in HE. These are:

1. using data to drive all aspects of decision making within HE, from admissions through to performance;
2. changing student behaviour due to metrics recorded and actions taken;
3. the risk of infantilising students and damaging their ongoing resilience.

These implications will need to be managed carefully, and on an institutional level to ensure fairness and compliance with the regulatory environment.

7.2.1 Data Driven Decisions

The thesis argues that using data to drive decisions is a positive, stating that it can ensure consistency. However, this is just a single, and remarkably easy, step away from automatic decision making. Automatic decision making may seem to be the Utopian destination of analytics, removing human fallibility from the process. However, it is fraught with challenges. An automatic decision isn't able to display sympathy, empathy or other emotional intelligence. (A critical element in almost every pastoral care issue.) Neither is it able to take non-data-described elements into consideration.

There is an unavoidable question of blame when poor achievement is discussed in an educational setting. There are four potential parties that may be at fault: the student, parents, educators, and the institution/system [230]. Reflective practice encourages educators to constantly evaluate their work [212]. With the addition of data and analytics there may be a push to abuse the findings and shift the blame more towards the student.

While LA may provide a reasonable basis for this argument, HE will need to look beyond individual cases and draw higher-level conclusions. This fact will tie-in with the emerging Curriculum Analytics field, and allow data to drive module and course development rather than blame. Achievement analytics may also provide an insight to the previously subjective determination of ‘good’ and ‘bad’ teachers. The data could instead be used as an input to the Professional Development Review (PDR) process. This would maintain the spirit of Reflective Practice with the empirical support of data and analytics.

The metrics and data collected by a LA programme will most likely provide information and insight influencing strategic decisions affecting the entire institution. However, viewed from the institutional level, all data loses the context of either department, programme, or student. This leads to potentially dangerous possibilities as the context does matter. Examples could include misuse of the work presented in this thesis to permanently exclude students that are flagged as likely to fail. The early nature this work gives means that the institution would not need to report these students as part of the TEF metrics and cynically improve their score. Recent research supports this student-centred view [323].

7.2.2 Student Behaviour Change

As HE institutions develop their data portfolio about students, the analytical output will also develop. The developments will be based on the questions asked of the data. Educators will need to become aware of how the collection of data to satisfy their inquiry can change the behaviour of their students. Most will recognise this as ‘gaming the system’, where the student attempts to collect any positive outcome with the least amount of effort [18]. (This is a phenomenon also discussed in Section 3.4.2.) The effect is magnified as more automatic or non-individualised actions are taken.

In order to minimise this effect, all interventions as a result of an alert from LA systems must be handled by an educator. The action should not be presented as a fait accompli and as a result of the alert, but considered as part of the overall pastoral care of the student. It is precisely this element of indirection that will disassociate the analytic result from the action and thus reduce the compulsion to bias any analytics.

7.2.3 Risk of Infantilisation and Harm to Resilience

Many of the efforts currently being deployed on university campuses geared toward improving engagement have been identified and in some cases demonised as infantilising the student body [194]. If the collective media is to be believed, the problem extends far beyond these programmes and into all aspects of the university experience [329].

Whether or not ‘students as consumers’, ‘safe spaces’, ‘trigger warnings’, or institutions looking to administrators to provide one-size-fits-all does in fact infantilise students, there is a profound change in HE. The utilisation of data provide either rationalisation for or document the need for these initiatives. Data and analytics, including LA, are not intrinsically damaging. However, these powerful forces can certainly be used to damaging ends.

The increasing use of analytics, and crucially backed by interventions, can remove another set of life skills that graduates arguably should possess before being sent out into the world. Despite the good intentions (and possible cynical and self-centred ones), the end result can be a graduate that is less resilient on their own. Automated suggestions, tutor interventions and highly directed environments are what some fear [14] as the zenith of the ‘spoon-fed educational system’. In many ways these critics are correct; replacing critical self-evaluation with the analytics output does remove a skill from the student. However, it could be argued that this is a fault of the implementation of the analytics system or the interventions instead. This dichotomy must be carefully considered as part of the planning phase of any LA deployment.

7.3 Recommendations for Higher Education Institutions

As part of this thesis, the following recommendations are made for any HE institution that is implementing, or considering implementing LA. While the recommendations will apply universally, the author expects differing levels of initial adoption in different institutions.

7.3.1 Data Quality

As the quality of all analytics products and processes directly rely on the input data, the single biggest recommendation for any HE institution is to ensure the highest level of data quality within their LMS and/or campus information systems. This advice becomes ever more pertinent with the proliferation of LA systems. As educators begin to rely on the data and (eventually) recommendations presented, the quality of interventions and advice they pass on to students will be compromised. The level of compromise will vary in sequence with the amount of low-quality data that is fed into the models. Or, phrased more colloquially ‘garbage-in, garbage-out’.

Therefore, a critical aspect before a large-scale deployment of analytics will be to ensure that the institution’s data is of a sufficiently high quality. This will affect every field used in analytical models, and is supported by other studies [133, 168]. The author is not suggesting blame or malice is creating this situation. More likely this situation arises due to restructuring efforts and staff losses [80]. When institutions ask administrators to achieve more with less time and resources, errors are almost guaranteed to be made.

On a purely functional-level, data quality can be improved by improving work-flow tools. In addition to systems enforcing change as part of process, the support and training for those using these tools will need to be improved.

The chief training concern is that all appropriate members of staff need to appreciate of the reasoning for the process. This lowers the risk of wilful circumvention or at least causes staff to think before making changes in unapproved ways. In the more holistic view, data literacy (that is the skills necessary to produce, handle, and make decisions with data) among staff will need to be improved. In the ‘Industrial Revolution 4.0’ age, these skills are as necessary as basic computer literacy [27].

7.3.2 Student Data Literacy and Awareness

As more millennials and digital natives progress through their academic careers attitudes and awareness of personal data are changing. Past generations did not grow up with social media, and thus their lives were somewhat more private. The millennials and digital natives conversely document their lives widely on various platforms. Few are aware of the amount of meta-data this also reveals about themselves [238]. These generations would benefit from a course describing the potential vulnerability leaking this information can lead to. As a result, a more informed debate may be held between students, educators, and their institutions over the use, or not, of personal data for LA. It should be incumbent on the HE institutions to initiate this conversation with their student bodies, before deploying LA systems. This would afford the students the ability to make an informed decision, with a valid comparison to the amount of data they already make public.

The author would recommend that these issues should be raised as early as possible with each new cohort of students. The ideal opportunity would be during Welcome Week, although this runs the risk of being drowned out with ‘information overload’. Therefore; the topic should be integrated with the ‘soft skills’ curriculum of every programme to reinforce the concepts at regular intervals. An added side-benefit would be that this restarts the conversation with students at several points in their studies. This gives them the opportunity to develop their own understanding and preferences at multiple points in time.

7.3.3 Transparency

As discussed in Section 3.4, there are difficult and unavoidable ethical dimensions to all analytics. None more so than transparency of decisions. Deployments of LA can be tailored for one of three primary audiences (educator, students, or institution) or a hybrid of these. These different viewpoints will place premiums on different sets of information. However, this should not be taken as an opportunity to hide information from any one of them. This is most critical when discussing the student viewpoint. Any data that is worthwhile to display to an educator about a student, should be treated as equally worthwhile for the student.

Transparency can also be the answer to prevent students attempting to manipulate the results of the model. When institutions are open about what data is used, how it is used, what the responses are likely to be, and the motivations for putting them into place students will be less motivated to unfairly influence them. This recommendation can also be a remedy to the perceived infantilisation of the student by restoring their agency over whatever area of student life that is being examined.

7.3.4 Data Use Agency/Opt-out

Withholding data from predictive (or even just analytic) models compromises the effectiveness and accuracy of them. In addition, the legal landscape is changing underneath LA as more territories update their data protection provisions. At the modelling stage there is no need to include personally identifiable data, therefore complying with data protection provisions and laws. The recommendation arising from this thesis would therefore be to include the processing of data for analytics into the institution's standard student 'contract', terms, charter or other instrument. Alongside the required legal definitions and terms, institutions should provide accessible (both in terms of language and delivery) materials informing their students of the intent and purpose of LA. In order to allow a full understanding institutions should refrain from making these materials glossy marketing style pamphlets. Crucially, the text must include the benefits for the institution rather than cynically hiding them. A study completed in 2001 by Ghosh, Wipple, and Bryan found that openness and integrity contribute to a student's trust in their university [115]. There is already evidence that students are suspect about online assessment practices [53], and it would be logical that this would be extended to online data processing such as LA. The second part of the agency students must be afforded is whether or not each wish to have personalised interventions as a result of analytics. This is both a technically possible response and consistent with the ethical principals examined. Students are then contributing to an analytical equivalent of herd immunity without necessarily having automatic or automated decisions made concerning them.

This recommendation does raise an interesting side-discussion, should students be able to opt-out of tutor deduced/observed interventions as well? In most cases there would be fewer objections to a educator-generated intervention as there is more faith that the educator has the student's best interests in mind. However, in this sense the educator is simply following the same kind of pattern recognition that automatic systems follow, with the added risk of human-error.

7.3.5 Ongoing Model Development

As highlighted in Section 4.9, LA models will not remain static and must be retrained and refreshed on average every two years. This presents a large opportunity to augment the model. As student cohorts change, whether through changes in lower education or as part of the commoditisation of HE; new factors may become more important to accurately assess students. Therefore, institutions need to be receptive to the idea that these models and systems have to evolve along with the students.

Allowing this to occur will require an ongoing commitment in terms of resources and personnel. However, as a result institutions' investment in LA, they should be able to improve the student experience, strengthen safeguards against undesirable outcomes, provide early warnings for mental health and other concerning pastoral issues, and support educators. This support should lead to improving the curriculum and teaching, and pastoral responses. Initially, the investment will be relatively large (with most being dispersed on staffing costs - either directly or in opportunity cost in diverting resources), but the eventual return on investment is a compelling case given the results presented in this thesis.

7.4 Future Work

As Learning Analytics is presently in vogue, there are multiple future lines of enquiry. There are three main thrusts in this potential work.

First, the classification task championed in Chapter 4 should be expanded. This work has explored the potential of using one derived predictor metric; it can be extended to include others from various campus information sources. Also, this work is a prime target for the use of Classifier Ensembles. This technique uses multiple classifiers, each trained in different aspects of the problem domain. When combined, using a suitable method, the result can be more powerful and accurate than a single classifier alone. This technique is based on the same underlying theory as crowd-sourcing. There are significant unexplored sets of demographic data (see Appendix A for example candidates), that may be able to add insight to LA results.

This work has focused on the identification of potentially failing students. Degree Pictures offer a standard representation of these journeys in terms of achievement. As noted in Chapter 5, this common description could allow standardised, base responses to assist educators in helping their students. The work actively admits that any standardised approach must be flexible enough to deal with the varying degrees of special circumstances encountered in Education. Development of these approaches is a sorely needed aspect in the LA ecosystem. Work will need to be completed in the Social Sciences, Psychology and Education arenas to determine what that standardised response should be.

As a result of the work presented in this thesis, the possibility of designing or deriving a ML-based model that is able to advise educators on the appropriate or best intervention becomes closer. However, this would need to be a distinct model but possibly taking input from the predictive model designed in Chapter 4.

In the author's view, the future lies with an ensemble of different ML models each trained to a different portion of the student experience and academic measures. Each of these results would be mediated through a weighted oracle [340]. These relative weights (a measure of importance/contribution) would then be derived at each institution during the training events. Due to the large diversity of entrance criteria, course structures, teaching methods, student opinion and cultures, and many other factors; there is simply no way to create a single predictive model.

A relatively new sub-field of Learning Analytics is emerging, Curriculum Analytics. This applies similar theory, techniques, and tools to questions about curricula. For example; whether the design of the curriculum causes patterns observed in traditional LA or how modifications to curricula may impact on student learning and well-being. When appropriately implemented, Curriculum Analytics could inform educators about such concerns as overassessment, coverage of intended learning outcomes, and potentially how to structure entire courses.

Taking Curriculum Analytics to its logical end, there is the possibility that combining traditional formative assessment techniques and the analytics institutions will be able to offer entirely personalised programmes. This formative assessment would evaluate the students' existing capabilities and knowledge of key curriculum areas. It would then make recommendations about which courses would be required to match them with the required programme learning outcomes. This could offer students a more appropriate experience where skills they are already proficient in are assessed without instruction and those needing to be developed are focused on with lecture, laboratory, or other support.

As a result of LA efforts, there will be new questions from the patterns and observations identified. This work has already given rise, within Bangor University, to an important question surrounding the impact of reading weeks on student behaviour and learning. The question is whether the impact improves or disrupts student efforts. Chapter 6 has offered some potential links with theory from other disciplines, but any characterisation will need to be proven empirically with an appropriate scientific basis in theory.

References

[1] D. Abbott, *Applied predictive analytics: Principles and techniques for the professional data analyst*. John Wiley & Sons, 2014, ISBN: 1118727967 (p. 16).

[2] M. Abdous, H. Wu and C.-J. Yen, 'Using data mining for predicting relationships between online question theme and final grade', *Journal of Educational Technology & Society*, vol. 15, no. 3, p. 77, 2012 (p. 36).

[3] R. L. Ackoff, 'From data to wisdom', *Journal of applied systems analysis*, vol. 16, no. 1, pp. 3–9, 1989 (p. 4).

[4] Acrobatiq, LLC. (2017). Learning analytics, [Online]. Available: `http://acrobatiq.com/products/learning-analytics-3` (visited on 12th Jun. 2017) (p. 46).

[5] Adobe Systems Inc. (2017). Captivate prime features, [Online]. Available: `http://www.adobe.com/uk/products/captivateprime/features.html` (visited on 12th Jun. 2017) (p. 46).

[6] Á. F. Agudo-Peregrina, S. Iglesias-Pradas, M. Á. Conde-González and Á. Hernández-García, 'Can we predict success from log data in vles? classification of interactions for learning analytics and their relation with performance in vle-supported f2f and online learning', *Computers in human behavior*, vol. 31, pp. 542–550, 2014 (p. 38).

[7] D. W. Aha and R. L. Bankert, 'A comparative evaluation of sequential feature selection algorithms', in *Learning from data*, Springer, 1996, pp. 199–206 (p. 20).

[8] P. All, 'From get truckinto jaws, students improve their learning pictures', *Unpublished masters thesis, University of Kansas, Lawrence*, 1977 (p. 92).

[9] J. W. Alstete and N. J. Beutell, 'Performance indicators in online distance learning courses: A study of management education', *Quality Assurance in Education*, vol. 12, no. 1, pp. 6–14, 2004 (p. 38).

[10] C. Ames, 'Classrooms: Goals, structures, and student motivation.', *Journal of educational psychology*, vol. 84, no. 3, p. 261, 1992 (pp. 28, 54).

[11] T. Anderson, C. Whittington and X. J. Li, 'Classes to passes: Is class attendance a determinant of grades in undergraduate engineering subjects?', in *AAEE2016 CONFERENCE Coffs Harbour, Australia*, 2016 (pp. 55, 65).

[12] F. Araque, C. Roldán and A. Salguero, 'Factors influencing university drop out rates', *Computers & Education*, vol. 53, no. 3, pp. 563–574, 2009. DOI: `https://doi.org/10.1016/j.compedu.2009.03.013` (p. 38).

[13] K. E. Arnold and M. D. Pistilli, 'Course signals at purdue: Using learning analytics to increase student success', in *Proceedings of the 2nd international conference on learning analytics and knowledge*, ACM, 2012, pp. 267–270 (pp. 46, 49).

[14] K. E. Arnold and N. Sclater, 'Student perceptions of their privacy in leaning analytics applications', in *Proceedings of the seventh international learning analytics & knowledge conference*, ACM, 2017, pp. 66–69 (p. 131).

[15] J. L. Arvai and A. Froschauer, 'Good decisions, bad decisions: The interaction of process and outcome in evaluations of decision quality', *Journal of Risk Research*, vol. 13, no. 7, pp. 845–859, 2010 (p. 14).

[16] R. S. Baker, D. Lindrum, M. J. Lindrum and D. Perkowski, 'Analyzing early at-risk factors in higher education e-learning courses.', *International Educational Data Mining Society*, 2015 (p. 38).

[17] R. S. Baker, A. T. Corbett, K. R. Koedinger and A. Z. Wagner, 'Off-task behavior in the cognitive tutor classroom: When students game the system', in *Proceedings*

of the SIGCHI conference on Human factors in computing systems, ACM, 2004, pp. 383–390 (p. 54).

[18] R. S. Baker and P. S. Inventado, 'Educational data mining and learning analytics', in *Learning analytics*, Springer, 2014, pp. 61–75 (p. 130).

[19] A. Bangor, P. T. Kortum and J. T. Miller, 'An empirical evaluation of the system usability scale', *Intl. Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008 (p. 106).

[20] R. Barber and M. Sharkey, 'Course correction: Using analytics to predict course success', in *Proceedings of the 2nd international conference on learning analytics and knowledge*, ACM, 2012, pp. 259–262 (pp. 38, 56).

[21] M. M. Barbieri, J. O. Berger *et al.*, 'Optimal predictive model selection', *The annals of statistics*, vol. 32, no. 3, pp. 870–897, 2004 (p. 17).

[22] S. Barr, 'Quitters' day: People most likely to give up new years resolutions today', *The Independent*, 2018. [Online]. Available: https://www.independent.co.uk/life-style/quitters-day-new-years-resolutions-give-up-fail-today-a8155386.html (p. 119).

[23] A. S. Bataineh, R. Mizouni, M. E. Barachi and J. Bentahar, 'Monetizing personal data: A two-sided market approach', *Procedia Computer Science*, vol. 83, pp. 472–479, 2016. DOI: 10.1016/j.procs.2016.04.211 (p. 16).

[24] S. Beattie, C. Woodley and K. Souter, 'Creepy analytics and learner data rights', in *Rhetoric and reality: Critical perspectives on educational technology.* Proceedings of ASCILITE Conference, 2014, pp. 421–425 (pp. 52, 57, 64).

[25] C. Beer, K. Clark, D. Jones *et al.*, 'Indicators of engagement', *Curriculum, technology & transformation for an unknown future. Proceedings ascilite Sydney*, pp. 75–86, 2010 (p. 65).

[26] F. Bendix, R. Kosara and H. Hauser, 'Parallel sets: Visual analysis of categorical data', in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, IEEE, 2005, pp. 133–140 (p. 103).

[27] A. Beneová and J. Tupa, 'Requirements for education and qualification of people in industry 4.0', *Procedia Manufacturing*, vol. 11, pp. 2195–2202, 2017 (p. 133).

[28] J. Bergin, K. Brodie, M. Patiño-Martínez, M. McNally, T. Naps, S. Rodger, J. Wilson, M. Goldweber, S. Khuri and R. Jiménez-Peris, 'An overview of visualization: Its use and design: Report of the working group in visualization', in *ACM SIGCSE Bulletin*, ACM, vol. 28, 1996, pp. 192–200 (p. 99).

[29] Y. Bergner, 'Measurement and its uses in learning analytics', in *Handbook of Learning Analytics*. Society for Learning Analytics Research, 2017, pp. 35–48 (p. 36).

[30] J. Bertin, 'Semiology of graphics: Diagrams, networks, maps', 1983 (pp. 24, 39).

[31] S. Beynon and C. Lawrence, Eds., *OfS adopts TEF and reappoints Prof Chris Husbands*, Office for Students Press Releases 31st Jan. 2018. [Online]. Available: https://www.officeforstudents.org.uk/news-blog-and-events/news-and-blog/ofs-adopts-tef-and-reappoints-prof-chris-husbands/ (visited on 30th Mar. 2018) (p. 2).

[32] Blackboard Inc. (2016). Blackboard predict data sheet, [Online]. Available: http://www.blackboard.com/resources/pdf/datasheet%20-%20predict%20-%20rev20161213.pdf (visited on 12th Jun. 2017) (p. 46).

[33] P. Blikstein, 'Using learning analytics to assess students' behavior in open-ended programming tasks', in *Proceedings of the 1st international conference on learning analytics and knowledge*, ACM, 2011, pp. 110–116 (p. 38).

[34] S. Bloxham, 'Marking and moderation in the uk: False assumptions and wasted resources', *Assessment & Evaluation in Higher Education*, vol. 34, no. 2, pp. 209–220, 2009 (p. 99).

[35] R. Bodily, J. Kay, V. Aleven, I. Jivet, D. Davis, F. Xhakaj and K. Verbert, 'Open learner models and learning analytics dashboards: A systematic review', in *Proceedings of*

the 8th International Conference on Learning Analytics and Knowledge, ACM, 2018, pp. 41–50 (p. 39).

[36] R. Bodily and K. Verbert, ‘Review of research on student-facing learning analytics dashboards and educational recommender systems’, *IEEE Transactions on Learning Technologies*, vol. 10, no. 4, pp. 405–418, 2017 (p. 39).

[37] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward and M. Chen, ‘Glyph-based visualization: Foundations, design guidelines, techniques and applications.’, in *Eurographics (STARs)*, 2013, pp. 39–63 (p. 40).

[38] R. Bose, ‘Advanced analytics: Opportunities and challenges’, *Industrial Management & Data Systems*, vol. 109, no. 2, pp. 155–172, 2009 (p. 39).

[39] K. W. Brodlie, L. Carpenter, R. Earnshaw, J. Gallop, R. Hubbold, A. Mumford, C. Osland and P. Quarendon, *Scientific visualization: techniques and applications*. Springer Science & Business Media, 2012 (p. 23).

[40] J. Brooke *et al.*, ‘Sus-a quick and dirty usability scale’, *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996 (p. 105).

[41] R. Brooks, K. Te Riele and M. Maguire, *Ethics and education research*. Sage, 2014 (p. 31).

[42] R. Brown, *Quality assurance in higher education: The UK experience since 1992*. Psychology Press, 2004 (p. 99).

[43] A. S. Bryk, L. M. Gomez and A. Grunow, ‘Getting ideas into action: Building networked improvement communities in education’, in *Frontiers in sociology of education*, Springer, 2011, pp. 127–162 (p. 31).

[44] S. Buckler and P. Castle, *Psychology for teachers*, 2nd ed. London, United Kingdom: SAGE Publications Ltd, 2018 (p. 28).

[45] J. P. Buerck, ‘A resource-constrained approach to implementing analytics in an institution of higher education: An experience report’, *Journal of Learning Analytics*, vol. 1, no. 1, pp. 129–139, 2014 (p. 33).

[46] A. Buja, J. A. McDonald, J. Michalak and W. Stuetzle, ‘Interactive data visualization using focusing and linking’, in *Visualization, 1991. Visualization’91, Proceedings., IEEE Conference on*, IEEE, 1991, pp. 156–163 (p. 102).

[47] P. Buser and K.-D. Semmler, ‘Study paths, riemann surfaces and strebel differentials’, *Journal of Learning Analytics*, vol. 4, no. 2, pp. 62–75, 2017, ISSN: 1929-7750 (p. 99).

[48] L. Byron and M. Wattenberg, ‘Stacked graphs–geometry & aesthetics’, *IEEE transactions on visualization and computer graphics*, vol. 14, no. 6, 2008 (p. 40).

[49] L. Cai and Y. Zhu, ‘The challenges of data quality and data quality assessment in the big data era’, *Data Science Journal*, vol. 14, p. 2, 2015. DOI: `10.5334/dsj-2015-002` (p. 12).

[50] A. Calder, *EU GDPR: a pocket guide*. IT Governance Publishing Ltd, 2018 (p. 50).

[51] S. K. Card, J. D. Mackinlay and B. Shneiderman, ‘Using vision to think’, in *Readings in information visualization*, Morgan Kaufmann Publishers Inc., 1999, pp. 579–581 (p. 24).

[52] R. M. Carini, G. D. Kuh and S. P. Klein, ‘Student engagement and student learning: Testing the linkages’, *Research in higher education*, vol. 47, no. 1, pp. 1–32, 2006 (p. 56).

[53] D. Carless, ‘Trust, distrust and their impact on assessment reform’, *Assessment & Evaluation in Higher Education*, vol. 34, no. 1, pp. 79–89, 2009 (p. 135).

[54] D. Carr, *Professionalism and ethics in teaching*. Routledge, 2005 (p. 27).

[55] J. Cassell and A. Young, ‘Why we should not seek individual informed consent for participation in health services research’, *Journal of medical ethics*, vol. 28, no. 5, pp. 313–317, 2002 (p. 51).

[56] P. Castle and S. Buckler, *Psychology for Teachers*, 2nd ed. Sage, 2018 (p. 28).

[57] C.-L. Chang, 'Finding prototypes for nearest neighbor classifiers', *IEEE Transactions on Computers*, vol. C-23, no. 11, pp. 1179–1184, Nov. 1974. DOI: 10.1109/T-C.1974.223827 (p. 80).

[58] T. Y. Chen, F.-C. Kuo and R. Merkel, 'On the statistical properties of the f-measure', in *Quality Software, 2004. QSIC 2004. Proceedings. Fourth International Conference on*, IEEE, 2004, pp. 146–153 (p. 23).

[59] Q. Chengzhi, Z. Chenghu and P. Tao, 'Taxonomy of visualization techniques and systems–concerns between users and developers are different', in *Asia GIS Conference*, vol. 35, p. 37 (p. 39).

[60] E. H.-h. Chi, 'A taxonomy of visualization techniques using the data state reference model', in *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, IEEE, 2000, pp. 69–75 (p. 39).

[61] M. T. Chi and R. Wylie, 'The icap framework: Linking cognitive engagement to active learning outcomes', *Educational Psychologist*, vol. 49, no. 4, pp. 219–243, 2014 (p. 56).

[62] C.-M. Chiu, T.-P. Liang and E. Turban, 'What can crowdsourcing do for decision support?', *Decision Support Systems*, vol. 65, pp. 40–49, 2014 (p. 100).

[63] E. Choinski, A. E. Mark and M. Murphey, 'Assessment with rubrics: An efficient and objective means of assessing student outcomes in an information resources class', *portal: Libraries and the Academy*, vol. 3, no. 4, pp. 563–575, 2003 (p. 99).

[64] K. L. S. Clair, 'A case against compulsory class attendance policies in higher education', *Innovative Higher Education*, vol. 23, no. 3, pp. 171–180, 1999 (p. 54).

[65] K. Clayton, F. Blumberg and D. P. Auld, 'The relationship between motivation, learning strategies and choice of environment whether traditional or including an online component', *British Journal of Educational Technology*, vol. 41, no. 3, pp. 349–364, 2010 (p. 28).

[66] D. Clow, 'An overview of learning analytics', *Teaching in Higher Education*, vol. 18, no. 6, pp. 683–695, 2013 (p. 1).

[67] L. Cohen, L. Manion and K. Morrison, *Research methods in education*, 8th ed. London, UK: Routledge, 2011, ISBN: 9780415583350 (p. 30).

[68] C. Colvin, S. Dawson, A. Wade and D. Gaevi, 'Addressing the challenges of institutional adoption', in *Handbook of Learning Analytics*. Society for Learning Analytics Research, 2017 (p. 12).

[69] J. Concato, N. Shah and R. I. Horwitz, 'Randomized, controlled trials, observational studies, and the hierarchy of research designs', *New England Journal of Medicine*, vol. 342, no. 25, pp. 1887–1892, 2000 (p. 31).

[70] F. Cordeiro, D. A. Epstein, E. Thomaz, E. Bales, A. K. Jagannathan, G. D. Abowd and J. Fogarty, 'Barriers and negative nudges: Exploring challenges in food journaling', in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 2015, pp. 1159–1162 (p. 118).

[71] Cornerstone OnDemand Ltd. (2017). Analytics features, [Online]. Available: https://www.cornerstoneondemand.co.uk/analytics (visited on 12th Jun. 2017) (p. 46).

[72] L. Corrin and P. de Barba, 'Exploring students interpretation of feedback delivered through learning analytics dashboards', in *Proceedings of the ascilite 2014 conference*, 2014, pp. 629–633 (p. 29).

[73] L. Corrin, G. Kennedy, P. de Barba, A. Bakharia, L. Lockyer, D. Gaevi, D. Williams, S. Dawson and S. Copeland, 'Loop: A learning analytics tool to provide teachers with useful data visualisations', presented at the 32nd Australasian Society for Computers in Learning in Tertiary Education Conference (ASCILITE '15), Perth, Australia, 2015, pp. 409–413. [Online]. Available: https://www.researchonline.mq.edu.au/vital/access/services/Download/mq:43896/SOURCE1 (p. 46).

[74] J. W. Crampton, 'Interactivity types in geographic visualization', *Cartography and geographic information science*, vol. 29, no. 2, pp. 85–98, 2002 (p. 25).

[75] M. Credé, S. G. Roch and U. M. Kieszczynka, 'Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics', *Review of Educational Research*, vol. 80, no. 2, pp. 272–295, 2010 (p. 121).

[76] D2L Corporation. (2017). Brightspace insights, [Online]. Available: `https://www.d2l.com/en-eu/products/insights/` (visited on 12th Jun. 2017) (p. 46).

[77] C. Daassi, L. Nigay and M.-C. Fauvet, 'A taxonomy of temporal data visualization techniques', *Information-Interaction-Intelligence*, vol. 5, no. 2, pp. 41–63, 2005 (p. 40).

[78] B. Daniel, 'Big data and analytics in higher education: Opportunities and challenges', *British journal of educational technology*, vol. 46, no. 5, pp. 904–920, 2015 (p. 49).

[79] W.-P. De Roever, K. Engelhardt and K.-H. Buth, *Data refinement: model-oriented proof methods and their comparison*. Cambridge University Press, 1998, vol. 47, ISBN: 0521103509 (p. 13).

[80] J. Dearlove, 'The deadly dull issue of university "administration"? good governance, managerialism and organising academic work', *Higher Education Policy*, vol. 11, no. 1, pp. 59–79, 1998 (p. 132).

[81] M. B. DeCotes, 'Data analytics of university student records', Master's thesis, University of Tennessee, 2014 (p. 99).

[82] S. Deterding, D. Dixon, R. Khaled and L. Nacke, 'From game design elements to gamefulness: Defining gamification', in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, ACM, 2011, pp. 9–15 (p. 47).

[83] P. Domingos, 'Bayesian averaging of classifiers and the overfitting problem', in *ICML*, vol. 2000, 2000, pp. 223–230 (p. 21).

[84] C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão *et al.*, 'Collinearity: A review of methods to deal with it and a simulation study evaluating their performance', *Ecography*, vol. 36, no. 1, pp. 27–46, 2013 (p. 20).

[85] J. Downs, R. Gilbert, R. D. Hayes, M. Hotopf and T. Ford, 'Linking health and education data to plan and evaluate services for children', *Archives of disease in childhood*, archdischild–2016, 2017 (p. 29).

[86] D. Draper and M. Gittoes, 'Statistical analysis of performance indicators in uk higher education', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 167, no. 3, pp. 449–474, 2004 (p. 38).

[87] L. P. Dringus, 'Learning analytics considered harmful.', *Journal of Asynchronous Learning Networks*, vol. 16, no. 3, pp. 87–100, 2012 (p. 58).

[88] G. T. Duncan and D. Lambert, 'Disclosure-limited data dissemination', *Journal of the American statistical association*, vol. 81, no. 393, pp. 10–18, 1986 (p. 51).

[89] E. Duval, 'Attention please!: Learning analytics for visualization and recommendation', in *Proceedings of the 1st international conference on learning analytics and knowledge*, ACM, 2011, pp. 9–17 (pp. 38, 99).

[90] A. L. Dyckhoff, D. Zielke, M. Bültmann, M. A. Chatti and U. Schroeder, 'Design and implementation of a learning analytics toolkit for teachers.', *Journal of Educational Technology & Society*, vol. 15, no. 3, 2012 (p. 48).

[91] C. Dziuban, P. Moskal, T. Cavanagh and A. Watts, 'Analytics that inform the university: Using data you already have.', *Journal of Asynchronous Learning Networks*, vol. 16, no. 3, pp. 21–38, 2012 (p. 65).

[92] G. O. Einstein and M. A. McDaniel, 'Prospective memory: Multiple retrieval processes', *Current Directions in Psychological Science*, vol. 14, no. 6, pp. 286–290, 2005 (p. 119).

[93] G. Ellis and F. Mansmann, 'Mastering the information age solving problems with visual analytics', in *Eurographics*, vol. 2, 2010, p. 5 (p. 25).

[94] Ellucian Inc., *Banner by ellucian*, 9.0, 2018. [Online]. Available: https://www.ellucian.com/emea-ap/Software/Banner-by-Ellucian/ (visited on 26th Oct. 2018) (p. 37).

[95] J. Elster, 'Altruistic behavior and altruistic motivations', *Handbook of the economics of giving, altruism and reciprocity*, vol. 1, pp. 183–206, 2006 (p. 31).

[96] B. Evans-Pritchard, 'Aiming To Reduce Cleaning Costs', *Works That Work*, no. 1, Dec. 2013. [Online]. Available: https://worksthatwork.com/1/urinal-fly (p. 117).

[97] W. Fan and A. Bifet, 'Mining big data: Current status, and forecast to the future', *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 2, pp. 1–5, 2013 (p. 12).

[98] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, 'From data mining to knowledge discovery in databases', *AI magazine*, vol. 17, no. 3, p. 37, 1996 (p. 13).

[99] R. Ferguson, 'Learning analytics: Drivers, developments and challenges', *International Journal of Technology Enhanced Learning*, vol. 4, no. 5-6, pp. 304–317, 2012 (pp. 1, 58).

[100] R. Ferguson and D. Clow, 'Where is the evidence?: A call to action for learning analytics', in *Proceedings of the seventh international learning analytics & knowledge conference*, ACM, 2017, pp. 56–65 (p. 58).

[101] E. Ferrance, *Action research*. LAB, Northeast and Island Regional Education Laboratory at Brown University, 2000 (p. 30).

[102] C. Ferri, J. Hernández-Orallo and R. Modroiu, 'An experimental comparison of performance measures for classification', *Pattern Recognition Letters*, vol. 30, no. 1, pp. 27–38, 2009 (p. 23).

[103] D. S. Fike and R. Fike, 'Predictors of first-year student retention in the community college', *Community college review*, vol. 36, no. 2, pp. 68–88, 2008 (pp. 55, 66).

[104] S. Finlay, *Predictive analytics, data mining and big data: Myths, misconceptions and methods*. Springer, 2014, ISBN: 1349478687 (p. 15).

[105] A. Fortenbacher, L. Beuster, M. Elkina, L. Kappe, A. Merceron, A. Pursian, S. Schwarzrock and B. Wenzlaff, 'Lemo: A learning analytics application focussing on user path analysis and interactive visualization', in *Intelligent data acquisition and advanced computing systems (idaacs), 2013 ieee 7th international conference on*, IEEE, vol. 2, 2013, pp. 748–753 (p. 39).

[106] J. R. Fraenkel, *How to design and evaluate research in education*, 8th ed. New York, N.Y.: McGraw-Hill Higher Education, 2012, ISBN: 9780071315180 (p. 30).

[107] S. de Freitas, D. Gibson, C. Du Plessis, P. Halloran, E. Williams, M. Ambrose, I. Dunwell and S. Arnab, 'Foundations of dynamic learning analytics: Using university student data to increase retention', *British Journal of Educational Technology*, vol. 46, no. 6, pp. 1175–1188, 2015 (pp. 38, 56).

[108] M. Friendly, 'A brief history of data visualization', in *Handbook of data visualization*, Springer, 2008, pp. 15–56 (p. 39).

[109] M. Friendly and D. J. Denis, 'Milestones in the history of thematic cartography, statistical graphics, and data visualization', *URL http://www. datavis. ca/milestones*, vol. 32, p. 13, 2001 (p. 43).

[110] J. Fritz, 'Using analytics at umbc: Encouraging student responsibility and identifying effective course designs', in *Research Bulletin*, Educause Center for Applied Research, 2013. [Online]. Available: https://library.educause.edu/resources/2013/4/using-analytics-at-umbc-encouraging-student-responsibility-and-identifying-effective-course-designs (p. 46).

[111] A. Gandomi and M. Haider, 'Beyond the hype: Big data concepts, methods, and analytics', *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015 (p. 17).

[112] D. Gaevi, S. Dawson, T. Rogers and D. Gaevi, 'Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success', *The Internet and Higher Education*, vol. 28, pp. 68–84, 2016 (p. 38).

[113] D. Gaevi, S. Dawson and G. Siemens, ‘Lets not forget: Learning analytics are about learning’, *TechTrends*, vol. 59, no. 1, pp. 64–71, 2015 (pp. 2, 33, 58, 82).

[114] L. Getoor and B. Taskar, *Introduction to statistical relational learning*. MIT press Cambridge, 2007, vol. 1, ISBN: 0262072882 (p. 18).

[115] A. K. Ghosh, T. W. Whipple and G. A. Bryan, ‘Student trust and its antecedents in higher education’, *The Journal of Higher Education*, vol. 72, no. 3, pp. 322–340, 2001 (p. 135).

[116] M. Gluchmanova, ‘Non-utilitarian consequentialism and its application in the ethics of teaching’, in *Proceedings of the XXII World Congress of Philosophy*, vol. 37, 2008, pp. 67–75 (p. 27).

[117] B. Goldacre, ‘Building evidence into education’, 2013 (p. 31).

[118] P. J. Goldstein and R. N. Katz, *Academic analytics: The uses of management information and technology in higher education*. Educause, 2005, vol. 8 (p. 33).

[119] P. Gomis-Porqueras, J. Meinecke and J. A. Rodrigues-Neto, ‘New technologies in higher education: Lower attendance and worse learning outcomes?’, *Agenda: A Journal of Policy Analysis and Reform*, pp. 69–83, 2011 (p. 55).

[120] L. E. Gonzalez, M. S. Brown and J. R. Slate, ‘Teachers who left the teaching profession: A qualitative understanding.’, *Qualitative Report*, vol. 13, no. 1, pp. 1–11, 2008 (p. 28).

[121] K. Gordon, ‘What is big data?’, *Itnow*, vol. 55, no. 3, pp. 12–13, 2013 (p. 15).

[122] M. Grabe and K. Christopherson, ‘Optional student use of online lecture resources: Resource preferences, performance and lecture attendance’, *Journal of Computer Assisted Learning*, vol. 24, no. 1, pp. 1–10, 2008 (p. 56).

[123] S. Graf, C. Ives, L. Lockyer, P. Hobson and D. Clow, ‘Building a data governance model for learning analytics’, in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, ACM, 2012, pp. 21–22 (p. 51).

[124] T. M. Green, W. Ribarsky and B. Fisher, ‘Building and applying a human cognition model for visual analytics’, *Information visualization*, vol. 8, no. 1, pp. 1–13, 2009 (p. 26).

[125] W. Greller and H. Drachsler, ‘Translating learning into numbers: A generic framework for learning analytics’, *Journal of Educational Technology & Society*, vol. 15, no. 3, p. 42, 2012 (pp. 1, 50).

[126] K. P. Gunn, ‘A correlation between attendance and grades in a first-year psychology class.’, *Canadian Psychology/Psychologie canadienne*, vol. 34, no. 2, p. 201, 1993 (p. 55).

[127] I. Guyon and A. Elisseeff, ‘An introduction to variable and feature selection’, *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003 (p. 19).

[128] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, ‘The weka data mining software: An update’, *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009 (p. 80).

[129] L. Hamilton, R. Halverson, S. S. Jackson, E. Mandinach, J. A. Supovitz, J. C. Wayman, C. Pickens, E. S. Martin and J. L. Steele, ‘Using student achievement data to support instructional decision making’, 2009. [Online]. Available: `http://repository.upenn.edu/gse_pubs/279` (p. 53).

[130] D. T. Hansen, *Exploring the moral heart of teaching: Toward a teacher’s creed*. Teachers’ College Press, 2001, ISBN: 0807740934 (p. 26).

[131] B. L. Harrison, R. Owen and R. M. Baecker, ‘Timelines: An interactive system for the collection and visualization of temporal data’, in *Graphics Interface*, 1994, pp. 141–141 (p. 40).

[132] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani and S. U. Khan, ‘The rise of big data on cloud computing: Review and open research issues’, *Information Systems*, vol. 47, pp. 98–115, 2015 (p. 14).

[133] B. T. Hazen, C. A. Boone, J. D. Ezell and L. A. Jones-Farmer, ‘Data quality for data science, predictive analytics, and big data in supply chain management:

An introduction to the problem and suggestions for research and applications', *International Journal of Production Economics*, vol. 154, pp. 72–80, 2014 (p. 132).

[134] C. Heaton-Shrestha, S. May and L. Burke, 'Student retention in higher education: What role for virtual learning environments?', *Journal of Further and Higher education*, vol. 33, no. 1, pp. 83–92, 2009 (p. 49).

[135] C. Herodotou, B. Rienties, A. Boroowa, Z. Zdrahal, M. Hlosta and G. Naydenova, 'Implementing predictive learning analytics on a large scale: The teacher's perspective', in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, ser. LAK '17, Vancouver, British Columbia, Canada: ACM, 2017, pp. 267–271, ISBN: 978-1-4503-4870-6. DOI: 10.1145/3027385.3027397. [Online]. Available: http://doi.acm.org/10.1145/3027385.3027397 (p. 49).

[136] T. Hoel and W. Chen, 'Implications of the european data protection regulations for learning analytics design', *Proceedings of CollabTech*, pp. 14–16, 2016 (p. 51).

[137] G. Huiskamp, 'Minority report on the bush doctrine', *New Political Science*, vol. 26, no. 3, pp. 389–415, 2004 (p. 52).

[138] Information Commissioner's Office. (2006). Information Commissioners - Data protection register - entry details (Z7439647), [Online]. Available: https://ico.org.uk/ESDWebPages/Entry/Z7439647 (visited on 27th Jan. 2016) (p. 65).

[139] S. M. Jayaprakash, E. W. Moody, E. J. Lauría, J. R. Regan and J. D. Baron, 'Early alert of academically at-risk students: An open source analytics initiative', *Journal of Learning Analytics*, vol. 1, no. 1, pp. 6–47, 2014 (p. 38).

[140] H. C. G. Johnsen and R. Normann, 'When research and practice collide: The role of action research when there is a conflict of interest with stakeholders', *Systemic Practice and Action Research*, vol. 17, no. 3, pp. 207–235, 2004 (p. 32).

[141] B. Johnson and B. Shneiderman, 'Tree-maps: A space-filling approach to the visualization of hierarchical information structures', in *Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on*, IEEE, 1991, pp. 284–291 (p. 42).

[142] M. V. Joshi, 'On evaluating performance of classifiers for rare classes', in *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 2002, pp. 641–644. DOI: 10.1109/ICDM.2002.1184018 (p. 23).

[143] D. Kahneman and P. Egan, *Thinking, fast and slow*. Farrar, Straus and Giroux New York, 2011, vol. 1 (p. 24).

[144] I. Kant, *The metaphysical elements of ethics*. The Floating Press, 2009 (p. 26).

[145] D. A. Keim, 'Information visualization and visual data mining', *IEEE Transactions on Visualization & Computer Graphics*, no. 1, pp. 1–8, 2002 (p. 47).

[146] D. A. Keim, F. Mansmann and J. Thomas, 'Visual analytics: How much visualization and how much analytics?', *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 5–8, 2010 (p. 17).

[147] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer and G. Melançon, 'Visual analytics: Definition, process, and challenges', in *Information visualization*, Springer, 2008, pp. 154–175 (pp. 24, 25).

[148] S. Kemmis, R. McTaggart and R. Nixon, *The action research planner: Doing critical participatory action research*. Springer Science & Business Media, 2013 (p. 30).

[149] R. Kimball and M. Ross, *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons, 2011 (p. 15).

[150] W. B. Kippers, C. L. Poortman, K. Schildkamp and A. J. Visscher, 'Data literacy: What do educators learn and struggle with during a data use intervention?', *Studies in Educational Evaluation*, vol. 56, pp. 21–31, 2018, ISSN: 0191-491X. DOI: https://doi.org/10.1016/j.stueduc.2017.11.001 (p. 53).

[151] J. Kirkup, 'Musical stairs could nudge Britain to better heath', *The Telegraph*, Dec. 2010. [Online]. Available: https://www.telegraph.co.uk/news/health/news/8234077/Musical-stairs-could-nudge-Britain-to-better-heath.html (p. 118).

[152] D. Kirsh and P. Maglio, ‘On distinguishing epistemic from pragmatic action’, *Cognitive Science*, vol. 18, no. 4, pp. 513–549, 1994. DOI: https://doi.org/10.1016/0364-0213(94)90007-8 (p. 24).

[153] M. A. Kitzrow, ‘The mental health needs of today’s college students: Challenges and recommendations’, *NASPA journal*, vol. 41, no. 1, pp. 167–181, 2003 (p. 29).

[154] KlassData. (2017). Smartklass (tm) | learning analytics plugin, [Online]. Available: http://klassdata.com/smartklass-learning-analytics-plugin/ (visited on 12th Jun. 2017) (p. 46).

[155] R. Kohavi *et al.*, ‘A study of cross-validation and bootstrap for accuracy estimation and model selection’, in *International Joint Conference on Artificial Intelligence*, Montreal, Canada, vol. 14, 1995, pp. 1137–1145 (p. 21).

[156] R. Kohavi, N. J. Rothleder and E. Simoudis, ‘Emerging trends in business analytics’, *Communications of the ACM*, vol. 45, no. 8, pp. 45–48, 2002 (p. 17).

[157] J. Kohlhammer, T. May and M. Hoffmann, ‘Visual analytics for the strategic decision making process’, in *Geospatial visual analytics*, Springer, 2009, pp. 299–310 (p. 25).

[158] B.-J. Koops, ‘The trouble with european data protection law’, *International Data Privacy Law*, vol. 4, no. 4, pp. 250–261, 2014 (p. 50).

[159] S. B. Kotsiantis, I. Zaharakis and P. Pintelas, ‘Supervised machine learning: A review of classification techniques’, *Emerging artificial intelligence applications in computer engineering*, vol. 160, pp. 3–24, 2007 (p. 18).

[160] S. Kotsiantis, D. Kanellopoulos, P. Pintelas *et al.*, ‘Handling imbalanced datasets: A review’, *GESTS International Transactions on Computer Science and Engineering*, vol. 30, no. 1, pp. 25–36, 2006 (p. 20).

[161] Z. J. Kovacic, ‘Predicting student success by mining enrolment data’, *Research in Higher Education Journal*, vol. 15, p. 1, 2012 (p. 38).

[162] D. L. Krebs, ‘Altruism: An examination of the concept and a review of the literature.’, *Psychological bulletin*, vol. 73, no. 4, p. 258, 1970 (p. 27).

[163] M. Kudo and J. Sklansky, ‘Comparison of algorithms that select features for pattern classifiers’, *Pattern recognition*, vol. 33, no. 1, pp. 25–41, 2000 (p. 80).

[164] G. D. Kuh, ‘Assessing what really matters to student learning inside the national survey of student engagement’, *Change: The Magazine of Higher Learning*, vol. 33, no. 3, pp. 10–17, 2001 (p. 56).

[165] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004 (p. 85).

[166] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal and A. Wolff, ‘Ou analyse: Analysing at-risk students at the open university’, *Learning Analytics Review*, pp. 1–16, 2015 (p. 46).

[167] B. C. Kwon, H. Kim, E. Wall, J. Choo, H. Park and A. Endert, ‘Axisketcher: Interactive nonlinear axis mapping of visualizations through user drawings’, *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 221–230, 2017 (p. 43).

[168] O. Kwon, N. Lee and B. Shin, ‘Data quality management, data usage experience and acquisition intention of big data analytics’, *International journal of information management*, vol. 34, no. 3, pp. 387–394, 2014 (p. 132).

[169] A. Labrinidis and H. V. Jagadish, ‘Challenges and opportunities with big data’, *Proceedings of the VLDB Endowment*, vol. 5, no. 12, pp. 2032–2033, 2012 (p. 15).

[170] P. Lally, C. H. Van Jaarsveld, H. W. Potts and J. Wardle, ‘How are habits formed: Modelling habit formation in the real world’, *European journal of social psychology*, vol. 40, no. 6, pp. 998–1009, 2010 (p. 119).

[171] M. Landin and J. Pérez, ‘Class attendance and academic achievement of pharmacy students in a european university’, *Currents in Pharmacy Teaching and Learning*, vol. 7, no. 1, pp. 78–83, 2015, ISSN: 1877-1297. DOI: https://doi.org/10.1016/j.cptl.2014.09.013 (p. 55).

[172] H. E. Larkin, '''but they won't come to lectures..." the impact of audio recorded lectures on student experience and attendance', *Australasian journal of educational technology*, vol. 26, no. 2, 2010 (p. 56).

[173] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins and N. Kruschwitz, 'Big data, analytics and the path from insights to value', *MIT sloan management review*, vol. 52, no. 2, p. 21, 2011 (p. 17).

[174] R. D. Lawrence, S. J. Hong and J. Cherrier, 'Passenger-based predictive modeling of airline no-show rates', in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2003, pp. 397–406 (p. 17).

[175] J. Lawson, *Data Science in Higher Education*. Chico, CA, USA, 2015, ISBN: 1515206467 (pp. 12, 13).

[176] P. A. Legg, E. Maguire, S. Walton and M. Chen, 'Glyph visualization: A fail-safe design scheme based on quasi-hamming distances', *IEEE computer graphics and applications*, vol. 37, no. 2, pp. 31–41, 2017 (p. 41).

[177] R. Leitch and C. Day, 'Action research and reflective practice: Towards a holistic view', *Educational action research*, vol. 8, no. 1, pp. 179–193, 2000 (p. 30).

[178] D. Leony, A. Pardo, L. de la Fuente Valentín, D. S. de Castro and C. D. Kloos, 'Glass: A learning analytics visualization tool', in *Proceedings of the 2nd international conference on learning analytics and knowledge*, ACM, 2012, pp. 162–163 (p. 39).

[179] H. Levkowitz and M. C. F. de Oliveira, 'From visual data exploration to visual data mining: A survey', *IEEE Transactions on Visualization & Computer Graphics*, vol. 9, pp. 378–394, Jul. 2003, ISSN: 1077-2626 (p. 23).

[180] H. Li, J. Zhang and J. Sun, 'A visual analytics approach for deterioration risk analysis of ancient frescoes', *Journal of Visualization*, vol. 19, no. 3, pp. 529–542, 2016 (p. 43).

[181] F.-R. Lin, L.-S. Hsieh and F.-T. Chuang, 'Discovering genres of online discussion threads via text mining', *Computers & Education*, vol. 52, no. 2, pp. 481–495, 2009 (p. 38).

[182] O. R. Lindsley, 'Precision teaching: By teachers for children', *Teaching Exceptional Children*, vol. 22, no. 3, pp. 10–15, 1990 (p. 92).

[183] ——, 'What we know that ain't so.', in *Third Convention Midwestern Association for Behavior Analysis*, Chicago, IL: Invited Address, 1977 (pp. 8, 92).

[184] T. D. Loboda, J. Guerra, R. Hosseini and P. Brusilovsky, 'Mastery grids: An open source social educational progress visualization', in *European Conference on Technology Enhanced Learning*, Springer, 2014, pp. 235–248 (p. 99).

[185] J. Lodge and M. Lewis, 'Pigeon pecks and mouse clicks: Putting the learning back into learning analytics', *Future challenges, sustainable futures. Proceedings ascilite Wellington*, pp. 560–564, 2012 (p. 57).

[186] M. Lott, 'Govt knows best? white house creates nudge squadto shape behavior', *Fox News*, Jul. 2013. [Online]. Available: `https://www.foxnews.com/politics/govt-knows-best-white-house-creates-nudge-squad-to-shape-behavior` (p. 117).

[187] J. S. Lourenço, E. Ciriolo, S. R. Almeida and X. Troussard, 'Behavioural insights applied to policy: European report 2016', European Commission, KJ-NA-27726-EN-N, 2016. DOI: `10.2760/903938`. [Online]. Available: `http://publications.jrc.ec.europa.eu/repository/bitstream/JRC100146/kjna27726enn%5C_new.pdf` (p. 117).

[188] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis and V. Loumos, 'Dropout prediction in e-learning courses through the combination of machine learning techniques', *Computers & Education*, vol. 53, no. 3, pp. 950–965, 2009, ISSN: 0360-1315. DOI: `https://doi.org/10.1016/j.compedu.2009.05.010` (p. 38).

[189] A. Macaskill, 'The mental health of university students in the united kingdom', *British Journal of Guidance & Counselling*, vol. 41, no. 4, pp. 426–441, 2013 (p. 29).

[190] L. P. Macfadyen and S. Dawson, 'Mining lms data to develop an early warning system for educators: A proof of concept', *Computers & Education*, vol. 54, no. 2, pp. 588–599, 2010, ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2009.09.008 (p. 38).

[191] ——, 'Numbers are not enough. why e-learning analytics failed to inform an institutional strategic plan.', *Journal of Educational Technology & Society*, vol. 15, no. 3, 2012 (p. 58).

[192] B. Macfarlane, 'The surveillance of learning: A critical analysis of university attendance policies', *Higher Education Quarterly*, vol. 67, no. 4, pp. 358–373, 2013 (p. 57).

[193] ——, 'Why students are treated worse than customers', *University World News*, May 2016. [Online]. Available: https://eprints.soton.ac.uk/398379/ (p. 57).

[194] B. Macfarlane and M. Tomlinson, 'Critiques of student engagement', *Higher Education Policy*, vol. 30, no. 1, pp. 5–21, 2017 (pp. 58, 131).

[195] M. Maltz *et al.*, *Psycho-Cybernetics: A New Technique for Using Your Subconscious Power*. Wilshire Book Company, 1967 (p. 119).

[196] K. Malvern, 'Bristol university suicides were not a failure in care, says academic', *The Times*, Jun. 2018, ISSN: 0140-0460. [Online]. Available: https://www.thetimes.co.uk/article/bristol-university-suicides-were-not-a-failure-in-care-says-academic-hxfldngwg (p. 29).

[197] S. A. Martin, 'Early intervention program and college partnerships', *ERIC Digest*, 1999 (p. 3).

[198] H. Mason and C. Wiggins, Eds., *A Taxonomy of Data Science*, dataists 10th Sep. 2010. [Online]. Available: http://www.dataists.com/2010/09/a-taxonomy-of-data-science/ (visited on 6th Apr. 2018) (p. 12).

[199] U. bin Mat, N. Buniyamin, P. M. Arsad and R. Kassim, 'An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention', in *Engineering Education (ICEED), 2013 IEEE 5th Conference on*, IEEE, 2013, pp. 126–130 (p. 56).

[200] J. N. Matthews, *Introduction to randomized controlled clinical trials*. Chapman and Hall/CRC, 2006 (p. 32).

[201] B. Maxwell and M. Schwimmer, 'Seeking the elusive ethical base of teacher professionalism in canadian codes of ethics', *Teaching and Teacher Education*, vol. 59, pp. 468–480, 2016, ISSN: 0742-051X (p. 27).

[202] H. A. McAllister, 'Self-serving bias in the classroom: Who shows it? who knows it?', *Journal of Educational Psychology*, vol. 88, no. 1, p. 123, 1996 (p. 31).

[203] S. K. McCann, M. K. Campbell and V. A. Entwistle, 'Reasons for participating in randomised controlled trials: Conditional altruism and considerations for self', *Trials*, vol. 11, no. 1, p. 31, 2010 (p. 31).

[204] McGraw-Hill Education. (2017). Connect features for educators, [Online]. Available: http://www.mheducation.com/highered/platforms/connect/features-educators.html (visited on 12th Jun. 2017) (p. 46).

[205] A. M. McIntosh, R. Stewart, A. John, D. J. Smith, K. Davis, C. Sudlow, A. Corvin, K. K. Nicodemus, D. Kingdon, L. Hassan *et al.*, 'Data science for mental health: A uk perspective on a global challenge', *The Lancet Psychiatry*, vol. 3, no. 10, pp. 993–998, 2016 (p. 29).

[206] J. McKay and P. Marshall, 'The dual imperatives of action research', *Information Technology & People*, vol. 14, no. 1, pp. 46–59, 2001 (p. 31).

[207] J. H. McMillan and S. Schumacher, *Research in Education: Evidence-Based Inquiry, MyEducationLab Series.* Harlow, UK: Pearson Higher Education, 2010 (p. 53).

[208] B. Means, E. Chen, A. DeBarger and C. Padilla, *Teachers' ability to use data to inform instruction: Challenges and supports.* Office of Planning, Evaluation and Policy Development, US Department of Education, Washington, D.C., United States, 2011 (p. 53).

[209] Mentorix ApS. (2017). Eurekos for education, [Online]. Available: `https://www.eurekos.com/education` (visited on 12th Jun. 2017) (p. 46).

[210] A. Merceron, P. Blikstein and G. Siemens, 'Learning analytics: From big data to meaningful data', *Journal of Learning Analytics*, vol. 2, no. 3, pp. 4–8, 2016 (p. 1).

[211] N. Metropolis and S. Ulam, 'The monte carlo method', *Journal of the American statistical association*, vol. 44, no. 247, pp. 335–341, 1949 (p. 69).

[212] K. Morss and R. Murray, *Teaching at university: a guide for postgraduates and researchers*. Sage, 2011 (p. 130).

[213] T. Munzner, 'A nested model for visualization design and validation', *IEEE transactions on visualization and computer graphics*, vol. 15, no. 6, 2009 (p. 93).

[214] S. Nalchigar, E. Yu and R. Ramani, 'A conceptual modeling framework for business analytics', in *International Conference on Conceptual Modeling*, Springer, 2016, pp. 35–49 (p. 16).

[215] P. W. S. Newall, 'Dark nudges in gambling', *Addiction Research & Theory*, pp. 1–3, 2018. DOI: `10.1080/16066359.2018.1474206` (p. 118).

[216] L. NewmanFord, K. Fitzgibbon, S. Lloyd and S. Thomas, 'A largescale investigation into the relationship between attendance and attainment: A study using an innovative, electronic attendance monitoring system', *Studies in higher education.*, vol. 33, no. 6, pp. 699–717, 2008, ISSN: 0307-5079 (p. 116).

[217] Q. Nguyen, B. Rienties and L. Toetenel, 'Unravelling the dynamics of instructional practice: A longitudinal study on learning design and vle activities', in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, ser. LAK '17, Vancouver, British Columbia, Canada: ACM, 2017, pp. 168–177, ISBN: 978-1-4503-4870-6. DOI: `10.1145/3027385.3027409`. [Online]. Available: `http://doi.acm.org/10.1145/3027385.3027409` (p. 49).

[218] F. Nightingale, W. Farr and A. Smith, *A contribution to the sanitary history of the British army during the late war with Russia*. John W. Parker and Son, 1859 (p. 42).

[219] C. N. Noble, 'Normative ethical theories', *The Monist*, vol. 62, no. 4, pp. 496–509, 1979 (p. 26).

[220] A. L. Nolen and J. V. Putten, 'Action research in education: Addressing gaps in ethical principles and practices', *Educational Researcher*, vol. 36, no. 7, pp. 401–407, 2007 (p. 32).

[221] Oracle Inc., *Peoplesoft campus solutions*, 8.53, 2018. [Online]. Available: `https://docs.oracle.com/cd/E52319_01/infoportal/cs.html` (visited on 26th Oct. 2018) (p. 37).

[222] B. Orend, *War and international justice: a Kantian perspective*. Wilfrid Laurier Univ. Press, 2006 (p. 26).

[223] Y. E. Orgler, 'A credit scoring model for commercial loans', *Journal of money, Credit and Banking*, vol. 2, no. 4, pp. 435–445, 1970 (p. 17).

[224] F. K. Oser, 'Chapter 2: Moral perspectives on teaching', *Review of research in education*, vol. 20, no. 1, pp. 57–127, 1994 (p. 28).

[225] C. Palenzona, 'The concept of necessity in the laws and deontology', *Minerva ginecologica*, vol. 23, no. 21, p. 917, 1971 (p. 27).

[226] A. Pardo, N. Mirriahi, R. Martinez-Maldonado, J. Jovanovic, S. Dawson and D. Gaevi, 'Generating actionable predictive models of academic performance', in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ACM, 2016, pp. 474–478 (p. 38).

[227] M. Parry, 'Colleges mine data to tailor students experience', *Chronicle of Higher Education*, vol. 58, no. 7, A1–A4, 2011 (p. 32).

[228] T. Patil and T. Davenport, 'Data scientist: The sexiest job of the 21st century', *Harvard Business Review*, 2012 (p. 11).

[229] Pearson Education Inc. (2017). Pearson higher education | mylab mastering, [Online]. Available: `https://www.pearsonmylabandmastering.com/global/`

educators/features/reporting-dashboard/index.html (visited on 12th Jun. 2017) (p. 46).

[230] E. R. Peterson, C. M. Rubie-Davies, M. J. Elley-Brown, D. A. Widdowson, R. S. Dixon and S. E. Irving, 'Who is to blame? students, teachers and parents views on who is responsible for student achievement', *Research in Education*, vol. 86, no. 1, pp. 1–12, 2011 (p. 130).

[231] C. Petrini, 'Preemptive kidney transplantation: Ethical issues', *Ann Ist Super Sanita*, vol. 45, no. 2, pp. 173–7, 2009 (p. 52).

[232] A. G. Picciano, 'The evolution of big data and learning analytics in american higher education.', *Journal of Asynchronous Learning Networks*, vol. 16, no. 3, pp. 9–20, 2012 (p. 50).

[233] M. D. Pistilli and K. E. Arnold, 'Purdue signals: Mining real-time academic data to enhance student success', *About Campus*, vol. 15, no. 3, pp. 22–24, 2010 (p. 33).

[234] *Policy on ethical use of student data for learning analytics*, Open University, 2014. [Online]. Available: http://www.open.ac.uk/students/charter/essential-documents/ethical-use-student-data-learning-analytics-policy (visited on 1st Oct. 2019) (p. 51).

[235] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng and D. R. Webster, 'Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning', *Nature Biomedical Engineering*, vol. 2, no. 3, pp. 158–164, 2018. DOI: 10.1038/s41551-018-0195-0 (p. 18).

[236] L. A. Price, *Characteristics of Early Student Dropouts at Allegany Community College and Recommendations for Early Intervention*. Allegany Community College, 1993. [Online]. Available: http://files.eric.ed.gov/fulltext/ED361051.pdf (p. 3).

[237] P. Prinsloo and S. Slade, 'Student privacy self-management: Implications for learning analytics', in *Proceedings of the fifth international conference on learning analytics and knowledge*, ACM, 2015, pp. 83–92 (p. 52).

[238] ——, 'Student vulnerability, agency, and learning analytics: An exploration', *Journal of Learning Analytics*, vol. 3, no. 1, pp. 159–182, 2016 (p. 133).

[239] F. Provost and T. Fawcett, 'Data science and its relationship to big data and data-driven decision making', *Big data*, vol. 1, no. 1, pp. 51–59, 2013 (p. 11).

[240] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993, ISBN: 1-55860-238-0 (p. 81).

[241] F. Rahman and P. Devanbu, 'How, and why, process metrics are better', in *Software Engineering (ICSE), 2013 35th International Conference on*, IEEE, 2013, pp. 432–441 (p. 23).

[242] R. Rao and S. K. Card, 'The table lens: Merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 1994, pp. 318–322 (p. 99).

[243] G. Rebolledo-Mendez, B. Du Boulay and R. Luckin, 'Motivating the learner: An empirical evaluation', in *International Conference on Intelligent Tutoring Systems*, Springer, 2006, pp. 545–554 (p. 28).

[244] Renaissance Learning, Inc. (2017). Learnalytics, [Online]. Available: http://www.renaissance.com/learnalytics/ (visited on 12th Jun. 2017) (p. 46).

[245] Y. Rezgui and A. Marks, 'Information security awareness in higher education: An exploratory study', *Computers & Security*, vol. 27, no. 7-8, pp. 241–253, 2008 (p. 51).

[246] G. Richards, 'Measuring engagement: Learning analytics in online learning', *electronic Kazan*, vol. 2011, 2011 (p. 38).

[247] P. Riehmann, M. Hanfler and B. Froehlich, 'Interactive sankey diagrams', in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, IEEE, 2005, pp. 233–240 (p. 103).

[248] C. Riener and D. Willingham, ‘The myth of learning styles’, *Change: The magazine of higher learning*, vol. 42, no. 5, pp. 32–35, 2010 (p. 48).

[249] B. Rienties, S. Cross and Z. Zdrahal, ‘Implementing a learning analytics intervention and evaluation framework: What works?’, in *Big data and learning analytics in higher education*, Springer, 2017, pp. 147–166 (p. 53).

[250] P. D. Ritsos and J. C. Roberts, ‘Towards more visual analytics in learning analytics’, in *Proceedings of the 5th EuroVis Workshop on Visual Analytics*, 2014, pp. 61–65 (p. 48).

[251] J. C. Roberts, ‘State of the art: Coordinated & multiple views in exploratory visualization’, in *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV’07. Fifth International Conference on*, IEEE, 2007, pp. 61–71 (p. 103).

[252] G. G. Robertson, S. K. Card and J. D. Mackinlay, ‘Information visualization using 3d interactive animation’, *Communications of the ACM*, vol. 36, no. 4, pp. 57–71, 1993 (p. 40).

[253] J. D. Rodriguez, A. Perez and J. A. Lozano, ‘Sensitivity analysis of k-fold cross validation in prediction error estimation’, *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 569–575, 2010 (p. 21).

[254] I. Roll and P. H. Winne, ‘Understanding, evaluating, and supporting self-regulated learning using learning analytics’, *Journal of Learning Analytics*, vol. 2, no. 1, pp. 7–12, 2015 (pp. 29, 55).

[255] V.-A. Romero-Zaldivar, A. Pardo, D. Burgos and C. D. Kloos, ‘Monitoring student progress using virtual appliances: A case study’, *Computers & Education*, vol. 58, no. 4, pp. 1058–1067, 2012 (p. 38).

[256] C. Romero, M.-I. López, J.-M. Luna and S. Ventura, ‘Predicting students’ final performance from participation in on-line discussion forums’, *Computers & Education*, vol. 68, pp. 458–472, 2013 (p. 38).

[257] C. Romero, S. Ventura, P. G. Espejo and C. Hervás, ‘Data mining algorithms to classify students’, presented at the 1st International Conference on Educational Data Mining, Montreal, Québec, Canada, Jun. 2008, pp. 8–17 (p. 38).

[258] F. Rosenblatt, *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957 (p. 21).

[259] S. Rovira, E. Puertas and L. Igual, ‘Data-driven system to predict academic grades and dropout’, *PLoS one*, vol. 12, no. 2, e0171207, 2017 (p. 38).

[260] ——, ‘Data-driven system to predict academic grades and dropout’, *PLOS ONE*, vol. 12, no. 2, pp. 1–21, Feb. 2017. DOI: 10.1371/journal.pone.0171207. [Online]. Available: https://doi.org/10.1371/journal.pone.0171207 (p. 49).

[261] D. E. Rumelhart, G. E. Hinton and R. J. Williams, ‘Learning representations by back-propagating errors’, *Nature*, vol. 323, Oct. 1986. DOI: 10.1038/323533a0 (p. 18).

[262] S. Russell and J. Bohannon, ‘Artificial intelligence. fears of an ai pioneer.’, *Science (New York, NY)*, vol. 349, no. 6245, pp. 252–252, 2015 (p. 18).

[263] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Pearson Education, 2003, ISBN: 0137903952 (p. 18).

[264] Y. Saeys, T. Abeel and Y. Van de Peer, ‘Robust feature selection using ensemble feature selection techniques’, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, 2008, pp. 313–325 (p. 20).

[265] SAP SE. (2017). Successfactors product brochure, [Online]. Available: https://www.successfactors.com/content/dam/successfactors/en_us/resources/brochures-product/learning-analytics.pdf (visited on 12th Jun. 2017) (p. 46).

[266] J. Sauro, *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC Denver, CO, 2011 (p. 106).

[267] R. Schutt and C. O’Neil, *Doing data science: Straight talk from the frontline*. O’Reilly Media, Inc., 2013, ISBN: 1449358659 (p. 12).

[268] B. A. Schwendimann, M. J. Rodriguez-Triana, A. Vozniuk, L. P. Prieto, M. S. Boroujeni, A. Holzer, D. Gillet and P. Dillenbourg, ‘Perceiving learning at a glance: A systematic

literature review of learning dashboard research', *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 30–41, 2017 (p. 39).

[269] SCIKit-Learn Developers. (2017). Choosing the right estimator, [Online]. Available: http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html (visited on 16th Jun. 2018) (p. 20).

[270] N. Sclater, *Learning analytics explained*. Routledge, 2017 (pp. 32, 35, 47, 51).

[271] N. Sclater, A. Peasgood and J. Mullan, 'Learning analytics in higher education', *London: Jisc. Accessed February*, vol. 8, p. 2017, 2016 (pp. 38, 46).

[272] N. Sclatter and P. Bailey, *Code of Practice for Learning Analytics*, JISC, Aug. 2018. [Online]. Available: http://repository.jisc.ac.uk/6985/1/Code_of_Practice_for_learning_analytics.pdf (p. 51).

[273] SEAtS Software Ltd. (2017). Learning analytics, [Online]. Available: https://www.seatssoftware.com/learning-analytics/ (visited on 12th Jun. 2017) (p. 46).

[274] P. B. Seddon, D. Constantinidis, T. Tamm and H. Dod, 'How does business analytics contribute to business value?', *Information Systems Journal*, vol. 27, no. 3, pp. 237–269, 2017 (p. 16).

[275] A. Seidman, *College Student Retention: Formula for Student Success*, ser. ACE/Praeger series on higher education. Praeger Publishers, 2005, ISBN: 9780275981938. [Online]. Available: https://books.google.co.uk/books?id=ckk5B%5C_ADM%5C_YC (p. 55).

[276] A. Sen and A. P. Sinha, 'A comparison of data warehousing methodologies', *Communications of the ACM*, vol. 48, no. 3, pp. 79–84, 2005 (p. 15).

[277] E. N. Shelton, 'Faculty support and student retention', *Journal of Nursing Education*, vol. 42, no. 2, pp. 68–76, 2003 (p. 55).

[278] G. Shmueli and O. R. Koppius, 'Predictive analytics in information systems research', *Mis Quarterly*, pp. 553–572, 2011 (p. 17).

[279] B. Shneiderman, 'The eyes have it: A task by data type taxonomy for information visualizations', in *Visual Languages, 1996. Proceedings., IEEE Symposium on*, IEEE, 1996, pp. 336–343 (pp. 24, 39, 93).

[280] G. Siemens and R. S. d Baker, 'Learning analytics and educational data mining: Towards communication and collaboration', in *Proceedings of the 2nd international conference on learning analytics and knowledge*, ACM, 2012, pp. 252–254 (p. 39).

[281] G. Siemens, D. Gaevi, C. Haythornthwaite, S. Dawson, S. B. Shum, R. Ferguson, E. Duval, K. Verbert and R. S. J. d. Bake, 'Open learning analytics: An integrated & modularized platform', presented at the Learning Analytics and Knowledge (LAK '11) Conference, Banff, Alberta, Canada, 2011. [Online]. Available: http://solaresearch.org/wp-content/uploads/2011/12/OpenLearningAnalytics.pdf (p. 1).

[282] G. Siemens and P. Long, 'Penetrating the fog: Analytics in learning and education.', *EDUCAUSE review*, vol. 46, no. 5, p. 30, 2011 (p. 33).

[283] S. Slade and P. Prinsloo, 'Learning analytics: Ethical issues and dilemmas', *American Behavioral Scientist*, vol. 57, no. 10, pp. 1510–1529, 2013 (p. 53).

[284] R. E. Slavin, 'Students motivating students to excel: Cooperative incentives, cooperative tasks, and student achievement', *The Elementary School Journal*, vol. 85, no. 1, pp. 53–63, 1984 (p. 118).

[285] J. Stasko and E. Zhang, 'Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations', in *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, IEEE, 2000, pp. 57–65 (p. 42).

[286] S. V. Stehman, 'Selecting and interpreting measures of thematic classification accuracy', *Remote Sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997, ISSN: 0034-4257. DOI: https://doi.org/10.1016/S0034-4257(97)00083-7 (p. 21).

[287] B. Stein and A. Morrison, 'The enterprise data lake: Better integration and deeper analytics', *PwC Technology Forecast: Rethinking integration*, vol. 1, pp. 1–9, 2014 (p. 15).

[288] R. J. Stiles, 'Understanding and managing the risks of analytics in higher education: A guide', *Educause Review*, 2012. [Online]. Available: `http://net.educause.edu/ir/library/pdf/EPUB1201.pdf` (p. 53).

[289] B. E. Stoliker and K. D. Lafreniere, 'The influence of perceived stress, loneliness, and learning burnout on university students' educational experience', *College Student Journal*, vol. 49, no. 1, pp. 146–160, 2015 (p. 74).

[290] K. D. Strang, 'Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes?', *Education and information technologies*, vol. 22, no. 3, pp. 917–937, 2017 (p. 56).

[291] K. A. Strike, 'The legal and moral responsibility of teachers', in *The Moral Dimensions of Teaching*, J. I. Goodlad, R. Soder and K. A. Sirotnik, Eds., Jossey-Bass Publishers, 1990 (p. 27).

[292] 'Student acheivement and university classes: Effects of attendance, size, peers and teachers', *IZA Discussion Papers*, no. 2490, [Online]. Available: `https://ssrn.com/abstract=955298` (p. 55).

[293] S. R. Sukumar and R. K. Ferrell, 'big datacollaboration: Exploring, recording and sharing enterprise knowledge', *Information Services & Use*, vol. 33, no. 3-4, pp. 257–270, 2013 (p. 15).

[294] J. A. Suykens, T. Van Gestel and J. De Brabanter, *Least squares support vector machines*. World Scientific, 2002, ISBN: 9812381511 (p. 19).

[295] D. T. Tempelaar, B. Rienties and Q. Nguyen, 'Towards actionable learning analytics using dispositions', *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 6–16, 2017 (p. 49).

[296] E. Terhart, 'Formalised codes of ethics for teachers: Between professional autonomy and administrative control', *European Journal of Education*, vol. 33, no. 4, pp. 433–444, 1998 (p. 26).

[297] I. G. Terrizzano, P. M. Schwarz, M. Roth and J. E. Colino, 'Data wrangling: The challenging journey from the wild to the lake.', in *CIDR*, 2015 (p. 15).

[298] N. Thai-Nghe, T. Horváth and L. Schmidt-Thieme, 'Factorization models for forecasting student performance.', in *EDM*, 2011, pp. 11–20 (p. 36).

[299] R. H. Thaler, *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press New Haven & London, 2008 (p. 117).

[300] J. J. Thomas and K. A. Cook, 'A visual analytics agenda', *IEEE computer graphics and applications*, vol. 26, no. 1, pp. 10–13, 2006 (p. 16).

[301] L. C. Thomas, 'A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers', *International journal of forecasting*, vol. 16, no. 2, pp. 149–172, 2000 (p. 18).

[302] J. A. Tice, S. R. Cummings, R. Smith-Bindman, L. Ichikawa, W. E. Barlow and K. Kerlikowske, 'Using clinical factors and mammographic breast density to estimate breast cancer risk: Development and validation of a new predictive model', *Annals of internal medicine*, vol. 148, no. 5, pp. 337–347, 2008 (p. 17).

[303] L. Tickle, 'Opening windows, closing doors: Ethical dilemmas in educational action research', *Journal of Philosophy of Education*, vol. 35, no. 3, pp. 345–359, 2001 (p. 52).

[304] V. Tinto, 'From theory to action: Exploring the institutional conditions for student retention', in *Higher education: Handbook of theory and research*, Springer, 2010, pp. 51–89 (p. 55).

[305] K. Tirri, 'Teachers' perceptions of moral dilemmas at school', *Journal of Moral Education*, vol. 28, no. 1, pp. 31–47, 1999 (p. 28).

[306] M. A. Titus, ‘An examination of the influence of institutional context on student persistence at 4-year colleges and universities: A multilevel approach’, *Research in higher education*, vol. 45, no. 7, pp. 673–699, 2004 (p. 75).

[307] M. Tory and T. Möller, ‘A model-based visualization taxonomy’, School of Computing Science, Simon Fraser University, 2002 (p. 39).

[308] ——, ‘Rethinking visualization: A high-level taxonomy’, in *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, IEEE, 2004, pp. 151–158 (p. 39).

[309] Tribal Group plc. (2015). Student insight brouchure, [Online]. Available: `http://www.tribalgroup.com/media/72452/po-tribal-student-insight-v2-mar15.pdf` (visited on 12th Jun. 2017) (p. 46).

[310] A. M. Turing, ‘I.computing machinery and intelligence’, *Mind*, vol. LIX, no. 236, pp. 433–460, 1950. DOI: `10.1093/mind/LIX.236.433` (p. 18).

[311] R. J. Vallerand, L. G. Pelletier, M. R. Blais, N. M. Briere, C. Senecal and E. F. Vallieres, ‘The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education’, *Educational and psychological measurement*, vol. 52, no. 4, pp. 1003–1017, 1992 (p. 28).

[312] M. L. Van Blerkom, ‘Class attendance in undergraduate courses’, *The Journal of psychology*, vol. 126, no. 5, pp. 487–494, 1992 (p. 118).

[313] V. N. Vapnik, ‘An overview of statistical learning theory’, *IEEE transactions on neural networks*, vol. 10, no. 5, pp. 988–999, 1999 (p. 18).

[314] C. P. Veenstra, ‘A strategy for improving freshman college retention’, *The Journal for Quality and Participation*, vol. 31, no. 4, p. 19, 2009 (p. 3).

[315] K. Verbert, E. Duval, J. Klerkx, S. Govaerts and J. L. Santos, ‘Learning analytics dashboard applications’, *American Behavioral Scientist*, vol. 57, no. 10, pp. 1500–1509, 2013 (p. 39).

[316] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss and M. McKeon, ‘Manyeyes: A site for visualization at internet scale’, *IEEE transactions on visualization and computer graphics*, vol. 13, no. 6, 2007 (p. 40).

[317] J. Wakeford, ‘It’s time for universities to put student mental health first’, *The Guardian*, Sep. 2017. [Online]. Available: `https://www.theguardian.com/higher-education-network/2017/sep/07/its-time-for-universities-to-put-student-mental-health-first` (visited on 1st Sep. 2018) (p. 29).

[318] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev and A. T. Campbell, ‘Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones’, in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, ACM, 2014, pp. 3–14 (p. 29).

[319] C. Ware, *Information visualization: perception for design*. Elsevier, 2012, ISBN: 0123814642 (pp. 24, 25).

[320] S. Weale, ‘Uk universities call for joined-up mental health care for students’, *The Times*, May 2018, ISSN: 0261-3077. [Online]. Available: `https://www.theguardian.com/society/2018/may/11/uk-universities-call-for-joined-up-mental-health-care-for-students` (p. 29).

[321] Y. Weinstein, M. Sumeracki and O. Caviglioli, *Understanding how we learn: a visual guide*. Routledge, 2018 (p. 28).

[322] D. West, H. Huijser, D. Heath, A. Lizzio, D. Toohey, C. Miles, B. Searle and J. Bronnimann, ‘Higher education teachers experiences with learning analytics in relation to student retention’, *Australasian Journal of Educational Technology*, vol. 32, no. 5, pp. 48–60, 2016 (p. 55).

[323] D. West, A. Luzeckyj, D. Toohey, J. Vanderlelie and B. Searle, ‘Do academics and university administrators really know better? the ethics of positioning student perspectives in learning analytics’, *Australasian Journal of Educational Technology*, 2019 (p. 130).

[324] S. C. Whiston, 'Accountability through action research: Research methods for practitioners', *Journal of Counseling & Development*, vol. 74, no. 6, pp. 616–623, 1996 (p. 32).

[325] M. Whitehead, R. Jones, J. Pykett and M. Welsh, 'Geography, libertarian paternalism and neuro-politics in the uk', *The Geographical Journal*, vol. 178, no. 4, pp. 302–307, 2012 (p. 117).

[326] A. W. Whitney, 'A direct method of nonparametric measurement selection', *IEEE Transactions on Computers*, vol. C-20, no. 9, pp. 1100–1103, Sep. 1971, ISSN: 0018-9340. DOI: `10.1109/T-C.1971.223410` (p. 20).

[327] L. Wilkinson and M. Friendly, 'The history of the cluster heat map', *The American Statistician*, vol. 63, no. 2, pp. 179–184, 2009 (p. 42).

[328] A. Williams, E. Birch and P. Hancock, 'The impact of online lecture recordings on student performance', *Australasian Journal of Educational Technology*, vol. 28, no. 2, 2012 (p. 56).

[329] J. Williams, 'Constructing consumption: What media representations reveal about todays students', *The marketisation of higher education and the student as consumer*, pp. 170–182, 2011 (p. 131).

[330] I. H. Witten, E. Frank, L. E. Trigg, M. A. Hall, G. Holmes and S. J. Cunningham, 'Weka: Practical machine learning tools and techniques with java implementations', 1999 (p. 19).

[331] Workday Inc., *Workday student*, 2018. [Online]. Available: `https://www.workday.com/en-us/applications/student.html` (visited on 26th Oct. 2018) (p. 37).

[332] Xyleme, Inc. (2017). Product | analyze, [Online]. Available: `http://www.xyleme.com/product/analyze` (visited on 12th Jun. 2017) (p. 46).

[333] H. Yavuzer, E. men-Gazolu, A. Yildiz,. Demr, A. Kiliçaslan, Ç. Serteln *et al.*, 'The teacher altruism scale: Development, validity and reliability.', *Educational Sciences: Theory & Practice*, vol. 6, no. 3, 2006 (p. 28).

[334] C. Ye and G. Biswas, 'Early prediction of student dropout and performance in moocs using higher granularity temporal information', *Journal of Learning Analytics*, vol. 1, no. 3, pp. 169–172, 2014 (p. 38).

[335] J. S. Yi, Y.-a. Kang, J. T. Stasko and J. A. Jacko, 'Understanding and characterizing insights: How do people gain insights using information visualization?', in *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, ACM, 2008, p. 4 (p. 39).

[336] J. S. Yi, Y. ah Kang, J. T. Stasko, J. A. Jacko *et al.*, 'Toward a deeper understanding of the role of interaction in information visualization', *IEEE Transactions on Visualization & Computer Graphics*, no. 6, 2007 (p. 47).

[337] M. Yorke, P. Bridges and H. Woolf, 'Mark distributions and marking practices in uk higher education: Some challenging issues', *Active learning in higher education*, vol. 1, no. 1, pp. 7–27, 2000 (p. 100).

[338] A. Zafra, C. Romero and S. Ventura, 'Multiple instance learning for classifying students in learning management systems', *Expert Systems with Applications*, vol. 38, no. 12, pp. 15 020–15 031, 2011 (p. 38).

[339] T. Z. Zarsky, 'Incompatible: The gdpr in the age of big data', *Seton Hall L. Rev.*, vol. 47, p. 995, 2016 (p. 51).

[340] X. Zeng, D. F. Wong and L. S. Chao, 'Constructing better classifier ensemble based on weighted accuracy and diversity measure', *The Scientific World Journal*, vol. 2014, 2014 (p. 137).

[341] J. Zilvinskis, J. Willis and V. M. H. Borden, 'An overview of learning analytics', *New Directions for Higher Education*, vol. 2017, no. 179, pp. 9–17, DOI: `10.1002/he.20239` (pp. 32, 33).

[342] K. Zivin, D. Eisenberg, S. E. Gollust and E. Golberstein, 'Persistence of mental health problems and needs in a college student population', *Journal of affective disorders*, vol. 117, no. 3, pp. 180–185, 2009 (p. 29).

Appendix A

Candidate Metrics for Predictive Model

A.1 Engagement Metric Options

- Extra Curricular
  - Athletic Union Team Membership
  - Campus Life[1] Involvement
  - Careers/Employability Service Involvement, inc. Bangor Employability Award Participation
  - Event/Sports Fixture Attendance
  - Gym / Sports Centre Usage
  - Peer Guide Scheme[2] Involvement
  - Student Society Membership
  - Student Society Management
  - Students' Union Election Participation
  - Students' Union Sabbatical Involvement
- Educational - Attendance
  - Classes, Laboratories, Lectures, Seminars.
  - Examinations / Tests.
  - Optional Academic Events
  - Optional School Events
  - Personal Tutor Sessions
  - Project Supervisor Sessions
- Educational - IT Services
  - Helpdesk Cases Raised
  - User Account Logins
  - VLE Content Views
  - VLE Logins
  - VLE Submissions

[1] Campus Life is a scheme at Bangor University designed to motivate students residing in University accommodation to interact and make the most of the facilities available to them.
[2] Peer Guides are returning students that volunteer to help newer students settle into life at Bangor University.

- Educational - Library
  - Electronic Resource Accesses
  - Inter-library Loans Requested
  - Items Borrowed
  - Items Returned
  - Items Returned Late
  - Searches Conducted
  - Visits
- Educational - Registration
  - Home School/Department
  - Programme
  - Programme Type
  - Repeating Year/Level?
  - Total Years Studied
  - Year of Programme
- Educational - Undergraduate Research Involvement
  - Publication/Research Co-authorship
  - Research Conference Attendance
  - Research Conference Presentation
- Student Representation / Feedback
  - Course Representative Involvement
  - Module Evaluation Completion
  - National Student Survey Completion
  - Staff-Student Liaison Committee Attendance
  - Staff-Student Liaison Committee Membership
  - Student-Led Teaching Awards Nominating
  - Student-Led Teaching Awards Voting
- Student Support
  - Support Service Assessment
  - Support Service Registration
  - Support Service Tutor/Advisor Visits
- Unofficial / Social Media
  - Posts in Groups for School, Cohort, Programme, or Project
  - ‘Help Wanted’ Posts
  - Responses to ‘Help Wanted’ Posts

A.2 Demographic Detail Options

- Standard Demographics
  - Age
  - Birth Country
  - Biological Sex
  - Home Country/Region
  - Marital Status
  - Nationality
  - Primary Language
  - Race
  - Religion
- Biographical
  - Criminal Record
  - Dependants
  - Disabilities
  - Driving License?
  - Employment History
  - Home Schooled?
  - Parental Education Level
  - Parental Relationship Status
  - Prior Education Level
  - Siblings
  - Technology/Device Ownership
- Financial
  - Credit Rating
  - Familial Socio-economic Group
  - Personal Socio-economic Group
  - Scholarships/Bursaries
  - Student Finance Arrangements
- Medical / Care
  - Gender Dysphoria?
  - Gender Identity
  - Mental Health Contact?
  - Medical History
  - Medications
  - Referrals?
  - Sexual Orientation
  - Social Services Contact?
  - Special Educational Needs Contact?
  - Specific Learning Difficulty (SpLD)
  - Substance Abuse/use?
- Social
  - Social Network Memberships
  - Society/Club/Group Memberships
  - Travel History

Appendix B

Data Silo Diagram

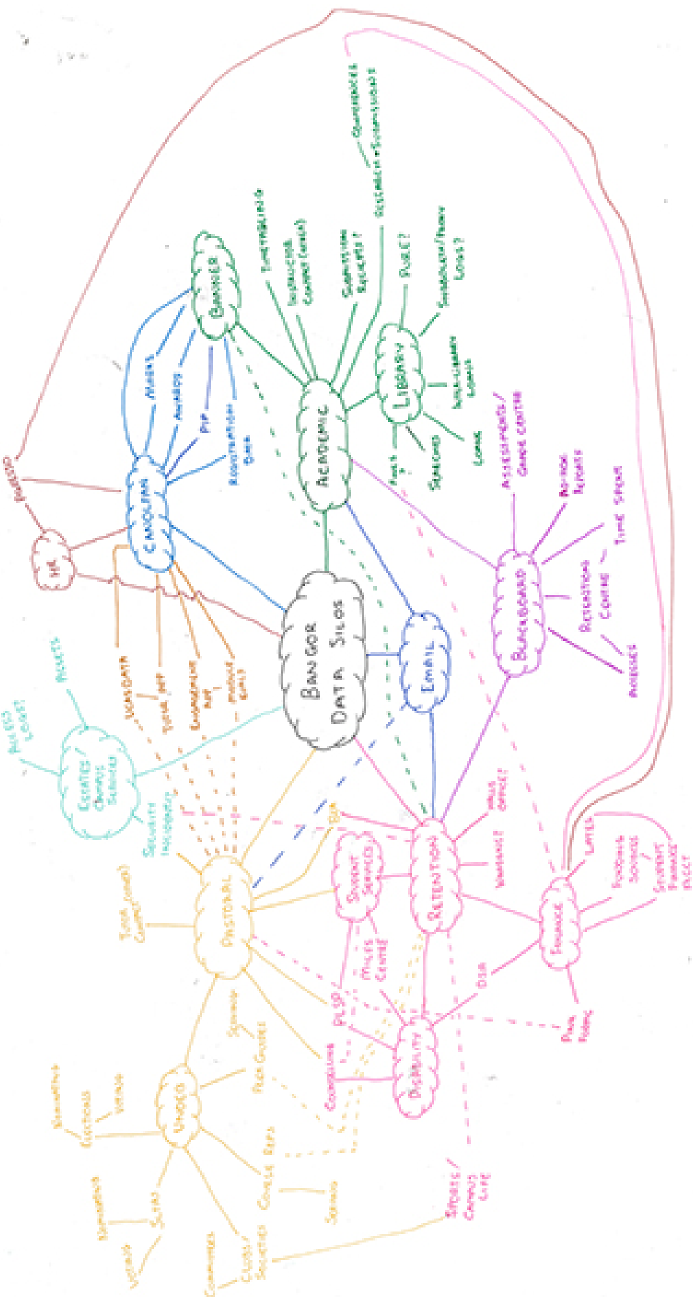


Figure B.1: Schematic overview of data storage locations/systems at Bangor University. This is based on a user perspective and may not match exactly with physical storage.

Appendix C

Classifier Benchmark Results

Table C.1: Full Results of Classifier Evaluation. The delta Δ metrics are between the two protocols, not other combinations. This table is sorted by the Δ Accuracy metric, then by the $\Delta\Sigma$ F-Fail (F-Measure for all classes but PA) metric.

Classifier	Discrim.	Protocol	Acc. %	Per-Class F-Measures					Σ F-Fail*	$\Delta\Sigma$ F-Fail*	Δ F-PA	Δ Acc.
				PA	FC	RY	FN	RS				
Random Tree	Program	Resub.	96.01	0.978	0.817	0.881	0.87	0	2.568	2.342	0.09	17.12
Random Tree	Program	LOO CV	78.89	0.888	0.134	0.041	0.051	0	0.226	2.342	0.09	17.12
RandomTree Regres.	Program	Resub.	96.02	0.978	0.817	0.881	0.87	0	2.568	2.346	0.079	16.99
RandomTree Regres.	Program	LOO CV	79.03	0.899	0.13	0.016	0.076	0	0.222	2.346	0.079	16.99
Random Forest	Program	Resub.	96.01	0.965	0.697	0.746	0.75	0	2.193	1.934	0.064	14.76
Random Forest	Program	LOO CV	81.25	0.901	0.146	0.047	0.066	0	0.259	1.934	0.064	14.76
RandomTree Regres.	School	Resub.	93.6	0.965	0.693	0.736	0.739	0	2.168	1.863	0.059	13.38
RandomTree Regres.	School	LOO CV	80.22	0.906	0.117	0.067	0.121	0	0.305	1.863	0.059	13.38
Random Tree	School	Resub.	93.56	0.965	0.688	0.736	0.739	0	2.163	1.897	0.064	12.74
Random Tree	School	LOO CV	80.82	0.901	0.124	0.029	0.113	0	0.266	1.897	0.064	12.74
Random Forest	School	Resub.	93.6	0.965	0.697	0.746	0.75	0	2.193	1.878	0.057	11.37
Random Forest	School	LOO CV	82.23	0.908	0.127	0.016	0.172	0	0.315	1.878	0.057	11.37
RotationForest	Program	Resub.	88.31	0.936	0.256	0.323	0.437	0	1.016	0.736	0.012	2.74
RotationForest	Program	10-Fold CV†	85.57	0.924	0.093	0.024	0.163	0	0.28	0.736	0.012	2.74
RotationForest	School	Resub.	88.07	0.935	0.246	0.283	0.386	0	0.915	0.693	0.011	2.48
RotationForest	School	LOO CV	85.59	0.924	0.043	0	0.179	0	0.222	0.693	0.011	2.48

Table C.1 continued from previous page

Classifier	Discrim.	Protocol	Acc. %	Per-Class F-Measures					Σ F-Fail*	$\Delta\Sigma$ F-Fail*	Δ F-PA	Δ Acc.
				PA	FC	RY	FN	RS				
Multi-layer Perceptron	School	Resub.	86.28	0.929	0.102	0.049	0.341	0	0.492	0.25	0.004	0.75
Multi-layer Perceptron	School	10-Fold CV†	85.53	0.925	0.08	0	0.162	0	0.242	0.25	0.004	0.75
Self-Organising Map	Program	Resub.	86.03	0.925	0.033	0.074	0.048	0	0.155	0.155	0.003	0.64
Self-Organising Map	Program	LOO CV	85.39	0.922	0	0	0	0	0	0.155	0.003	0.64
Naïve Bayes	Program	Resub.	85.51	0.925	0.048	0	0.264	0	0.312	0.098	0.004	0.78
Naïve Bayes	Program	LOO CV	84.73	0.921	0.024	0	0.19	0	0.214	0.098	0.004	0.78
Multi-layer Perceptron	Program	Resub.	85.75	0.923	0	0	0	0	0	0.096	0.002	0.01
Multi-layer Perceptron	Program	10-Fold CV†	85.74	0.925	0.029	0	0.067	0	0.096	0.096	0.002	0.01
DecisionTable	Program	Resub.	86	0.925	0	0	0.136	0	0.136	0.023	0.001	0.16
DecisionTable	Program	LOO CV	85.84	0.924	0	0	0.113	0	0.113	0.023	0.001	0.16
DecisionTable	School	Resub.	86	0.925	0	0	0.136	0	0.136	0.023	0.001	0.16
DecisionTable	School	LOO CV	85.84	0.924	0	0	0.113	0	0.113	0.023	0.001	0.16
Naïve Bayes	School	Resub.	84.95	0.922	0.02	0	0.208	0	0.228	0.017	0.001	0.14
Naïve Bayes	School	LOO CV	84.81	0.921	0.012	0	0.199	0	0.211	0.017	0.001	0.14
Self-Organising Map	School	Resub.	85.75	0.923	0	0	0	0	0	0	0	0.02
Self-Organising Map	School	LOO CV	85.73	0.923	0	0	0	0	0	0	0	0.02
C4.5 Tree	Program	Resub.	85.75	0.923	0	0	0	0	0	0	0	0

Table C.1 continued from previous page

Classifier	Discrim.	Protocol	Acc. %	Per-Class F-Measures					Σ F-Fail*	$\Delta\Sigma$ F-Fail*	Δ F-PA	Δ Acc.
				PA	FC	RY	FN	RS				
C4.5 Tree	Program	LOO CV	85.75	0.923	0	0	0	0	0	0	0	0
C4.5 Tree	School	Resub.	85.75	0.923	0	0	0	0	0	0	0	0
C4.5 Tree	School	LOO CV	85.75	0.923	0	0	0	0	0	0	0	0

Key: Acc. - Accuracy, Discrim. - Discriminator, Regres. - Regression, Resub. - Resubstituion.

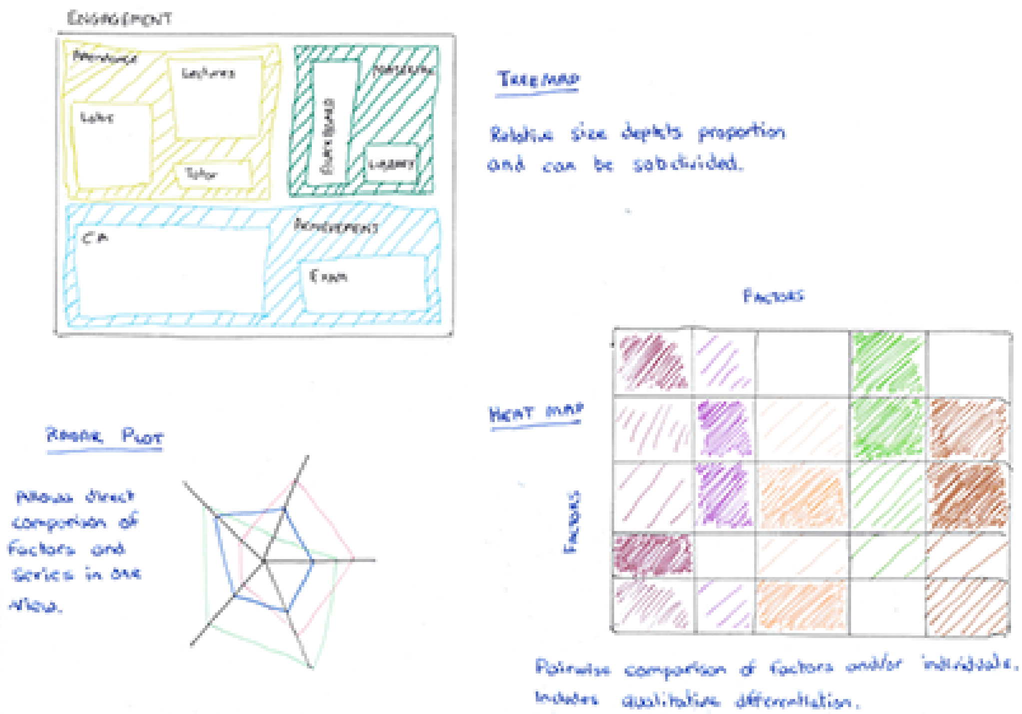
* F-Fail represents the F-Measure scores for all classes except PA (Pass).

†10-Fold CV Experiments were substituted for Leave One Out CV due to time constraints.

Appendix D

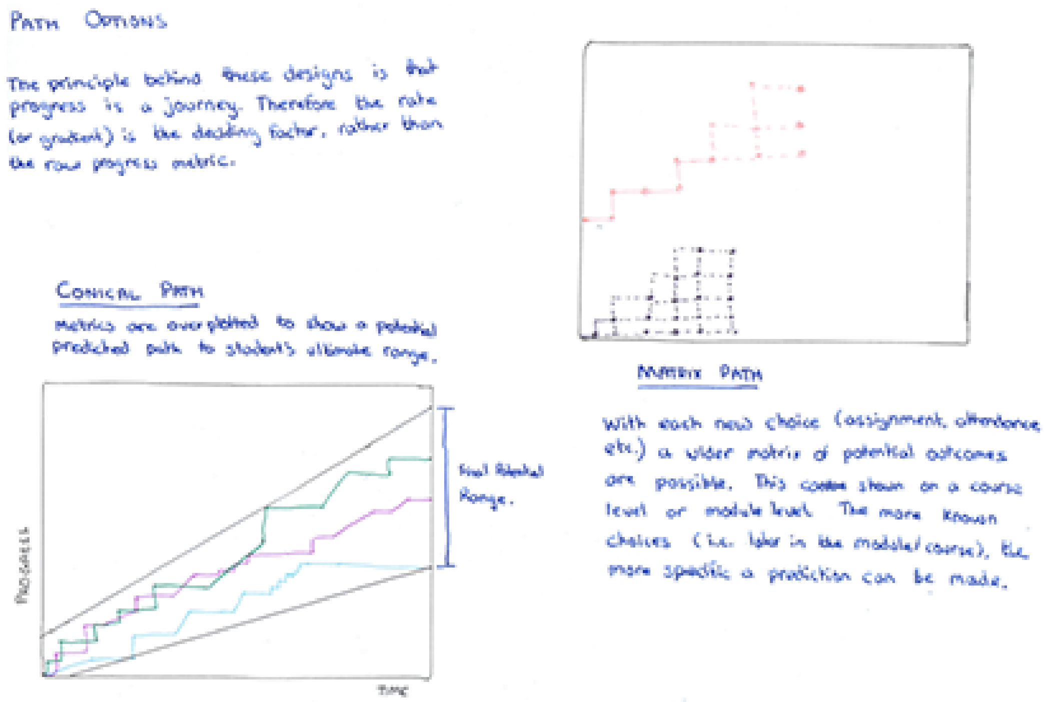
Initial Designs for the Student Journey Tool

(a) Potential methods to show multivariate data in one chart.

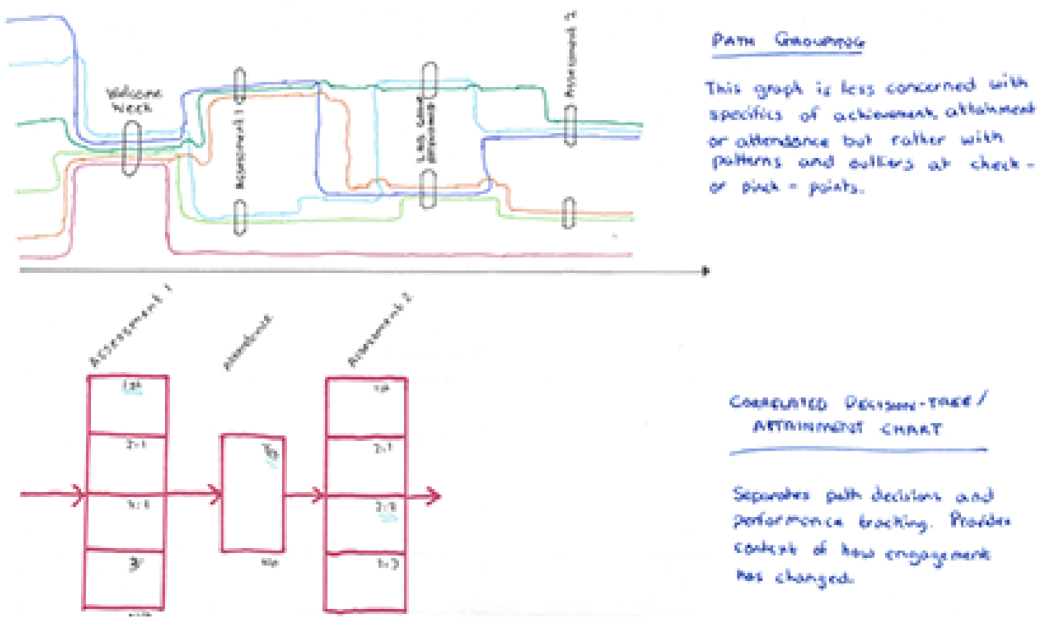


(b) Potential methods to show data as an explicit hierarchy in one chart.

Figure D.1: Initial Design Ideas for the Student Journey Visualisation



(c) Potential methods to show journeys as a path.



(d) Potential methods using path grouping to show multiple progressions (from different areas of student life) on one chart.

Figure D.1: Initial Design Ideas for the Student Journey Visualisation (cont.).