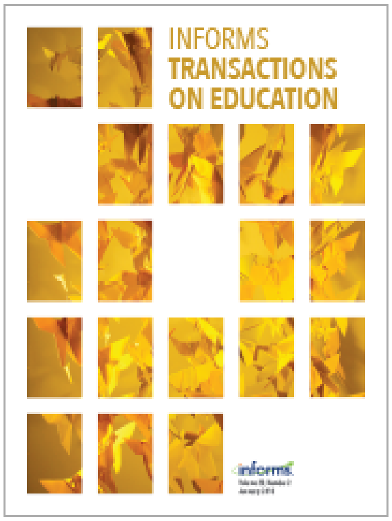


INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Identifying Patterns of Learner Behaviour: What Business Statistics Students Do with Learning Resources

Paula Carroll, Arthur White

informs
http://pubsonline.informs.org/journal/ited/

INFORMS TRANSACTIONS ON EDUCATION
Vol. 18, No. 1, September 2017, pp. 1–13
ISSN 1532-0545 (online)

Identifying Patterns of Learner Behaviour: What Business Statistics Students Do with Learning Resources

Paula Carroll,[a] Arthur White[a]

[a]UCD Quinn School of Business, University College Dublin, Belfield, Dublin 4, Ireland
Contact: paula.carroll@ucd.ie (PC); arthur.white@ucd.ie (AW)

Received: January 8, 2016
Revised: June 17, 2016; September 23, 2016; October 20, 2016
Accepted: October 26, 2016
Published Online in Articles in Advance: June 12, 2017

https://doi.org/10.1287/ited.2016.0169

Abstract. The interactions of early stage business students with learning resources over the duration of an introductory statistics module were analysed using latent class analysis. Four distinct behavioural groups were identified. While differing levels of face-to-face attendance and online interaction existed, all four groups failed to engage with online material in a *timely* manner. The four groups were found to demonstrate significantly different levels of attainment of the module learning outcomes. The patterns of behaviour of the different groups of students give insights as to which analytics education learning resources students use and how their use patterns relate to their level of attainment of the module learning outcomes.

Keywords: analytics education • learning resource development • latent class analysis

1. Introduction

The restructuring of undergraduate business degree programmes at the School of Business in University College Dublin (UCD) Ireland in 2011 identified quantitative and analytical skills as central to the holistic education of business students. Sound decision-making practices based on data analysis were viewed as an essential skill for future business leaders. The review involved consultation with key stakeholders and identified three programme pillars, i.e., business in society, innovation and enterprise, and personal development planning. The Data Analysis for Decision Makers (DADM) module was designed during the review to develop business students' analytical skills by introducing them to standard statistical practices.

The philosophy of the UCD School of Business is to use digital content to underpin knowledge and use face-to-face lectures and tutorials to deepen student learning and facilitate critical skills development. There is a concern that in offering online resources in a blended learning environment, students may substitute online activity for face-to-face activities and may fail to engage with the material at the appropriate level. The availability of online content can cause students not to attend or to become disengaged with lectures (Gordon 2014).

In addition, educators face challenges to engage business students in the development of quantitative skills (McAlevey and Sullivan 2001, Yilmaz 1996). Business students often feel that analysis and statistics are boring and not relevant to business studies. They perceive that analysis is required in business but not that they may become responsible for performing or interpreting that analysis, see for example Cronin and Carroll (2015).

Sensible frameworks to improve teaching in statistical and quantitative subjects are outlined in Moore (1997) and by the American Statistical Association (ASA) (2005, 2016). Moore suggests that the move from instructivist to constructivist learning theories forces statistical educators to incorporate more active learning opportunities for their students (Moore 1997). With this advice in mind, a set of learning objects for the DADM module was developed. Learning objects are defined as any entity, digital or nondigital, that may be used for learning, education or training (IEEE 2002).

This paper examines how learning objects, such as online resources and face-to- face lectures and tutorials, are used by early stage business students. Learning Analytics (LA) are applied with the aim of (1) identifying patterns of learner use of the designed learning objects, (2) quantifying the relationship between patterns of learner use behaviour and levels of attainment of the module learning outcomes, and (3) creating a set of recommendations on the type of learning objects that enable business students' quantitative skills development.

We found that all the learning objects are used but that they are used in different ways by distinct groups of students. We identified four distinct behaviourial groups with different preferences of learning object use. The largest group (34%) gradually substituted face-to-face activities with online activity. Two other groups (55% together) displayed consistent attendance at face-to-face activities but showed differing patterns of use of online resources. The pass rate for these three groups were statistically similar but the average grade was slightly lower for the first group. Lastly, a small cohort of at-risk students were identified who had low engagement with any of the learning objects. The study provides evidence that successful student learning outcomes can be obtained in introductory business statistics using different mixes of learning objects.

1.1. Ecosystem Framework

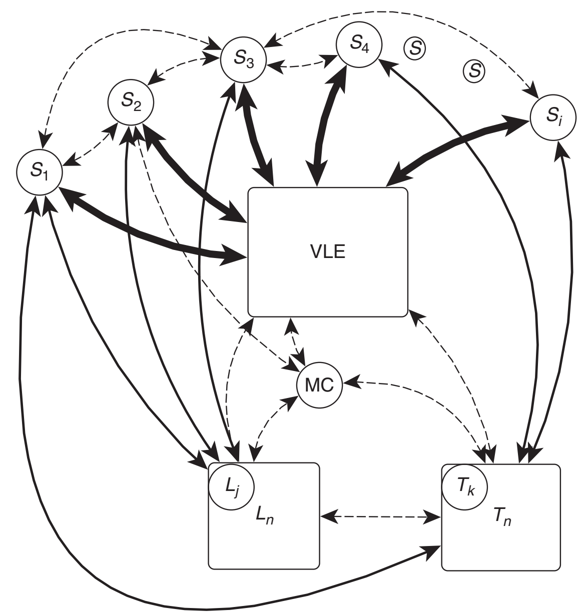
Two philosophies of the root cause of effective teaching could be seen to conflict but in our opinion are complementary. What the student does is the focus of Biggs' constructive alignment theory (Biggs 1999). What the teacher does (with technology) is the focus of the Technological Pedagogical Content Knowledge (TPACK) framework (Koehler et al. 2013). It is more likely that positive interaction between the student and the learning objects can result in effective learning.

There have been some attempts to harness the work in the science domains to view the totality of learner interaction as a learning ecosystem. The characteristics of modern learning settings are comparable with the situation in biotic ecosystems. Holistic ecosystem based models for learning and e-learning are outlined in Gütl and Chang (2008). Figure 1 shows our view of a learning ecosystem with student learners (S_i) interacting with each other, the Virtual Learning Environment (VLE) and the teaching agents (module coordinator (MC), lecturers (L_j) at lectures (L_n), and tutors (T_k) at tutorials (T_n). Interactions between the learners and learning objects, which are the focus of this study, are shown with solid links. Other interaction such as student-student interactions are shown with dashed links (as they are not the focus of this article). Not shown is the external environment in which the learners and learning objects co-exist.

Technology is just one of the ecosystem components that is in a constant state of flux. Successful teaching with technology, as advocated in ASA (2016), requires continually creating, maintaining, and re-establishing a dynamic equilibrium among all three components (Koehler et al. 2013). A range of factors influences how this equilibrium is reached.

Biggs' constructive alignment preceded the role of technology and modern social contexts of the teaching and learning environment. It still offers much in

Figure 1. Learning Ecosystem: Student Interaction with Learning Objects



structuring appropriate learning activities and assessment to help students meet well defined learning outcomes. TPACK focuses on the Technology, Pedagogy and (discipline) Content Knowledge. While the Pedagogy knowledge base attempts to understand how learning happens, the TPACK framework misses Biggs' point: It is what the student does that makes learning effective.

LA and Educational Data Mining (EDM) provide options to understand complex phenomenona such as student interaction with technology. These techniques play a role in providing empirical evidence about what students do with the learning objects. These connections are shown as the lines in the learning ecosystem in Figure 1. The borders on the subsystem investigated in this study are necessarily bounded by the data that can be captured. The interactions of the biotic components (student/learners) with biotic (lecturers and tutors) and abiotic components (eLearning content) of the learning ecosystem are also investigated.

1.2. Research Questions

In this article, we focus on the behavioural learning patterns in the ecosystem of undergraduate business students taking the DADM module in the 2013/2014 academic year. The DADM module is offered in a blended learning environment and is described in detail in Section 2. Anecdotal evidence suggested low levels of attendance and engagement in face-to-face activities. However, as the module is delivered to a large cohort of students, it is not immediately apparent whether these claims are valid for all students or just for certain groups of students. Nonengagement is a cause for concern. As noted in Pascarella and Terenzini (2005), academic engagement is a factor in the development of cognitive skills. Since face-to-face attendance and online activity are proxy measures of student academic engagement, they are both measured to establish (a) levels of attendance at contact sessions and (b) levels of online activity. The focus of this paper is to understand what the students are doing with the learning objects and whether that relates to their levels of attainment of learning outcomes as measured by assessment grades.

Two primary research questions are addressed here:

1. What are the identifiable behavioural patterns of learning object use?
2. How do these patterns of learner behaviour relate to levels of attainment in terms of learning outcomes?

Some external factors to the ecosystem such as gender and prior educational attainment are considered. The Leaving Certificate is the Irish end of the secondary education state exam. It is the primary criterion to determine which students are offered places in higher education programmes. In particular, we consider the performance of students in compulsory subjects (English and Mathematics) as well as their overall Leaving Certificate performance.

The structure of the paper is as follows: Section 2 describes the DADM module. We describe the study data set, collection methodology, and analysis approach in Section 3. Results are analysed in Section 4 and some discussion and conclusions are provided in Section 5.

2. Data Analysis for Decision Makers

DADM is a core module taken by all stage one School of Business students in semester 1 at UCD. The learning outcomes say that on successful completion, students should be able to:

1. Prepare spread-sheet models to store, manipulate, and analyse quantitative data using common probability distributions and statistical functions;
2. Calculate, analyse, and present useful statistical measurements from large-scale data sets;
3. Create and interpret inferential statistical statements about population parameters;
4. Interpret the results of data analyses with a view to informing decision making.

Approximately 600 students per annum take the module and use their own device to access digital learning resources on Blackboard, the VLE used in UCD. Online resources are complemented by weekly two hour lectures and one hour tutorials. The students are broken into groups of about 150 for lectures and 50 for tutorials. Faculty staff explain data analysis concepts in a business context during lectures and facilitate active learning tasks. Lectures are followed by tutorials where teaching assistants facilitate a discussion of pre-assigned self-assessment exercises. Additional optional supports were arranged such as guest speakers from industry, drop-in clinics, and hot topic sessions facilitated by the university's Math Support Centre (MSC). A summarised schedule of the DADM activities and continuous assessment (CA) tasks is shown in Table 1.

2.1. Design of DADM Learning Objects

In designing the DADM learning resources we focused on:

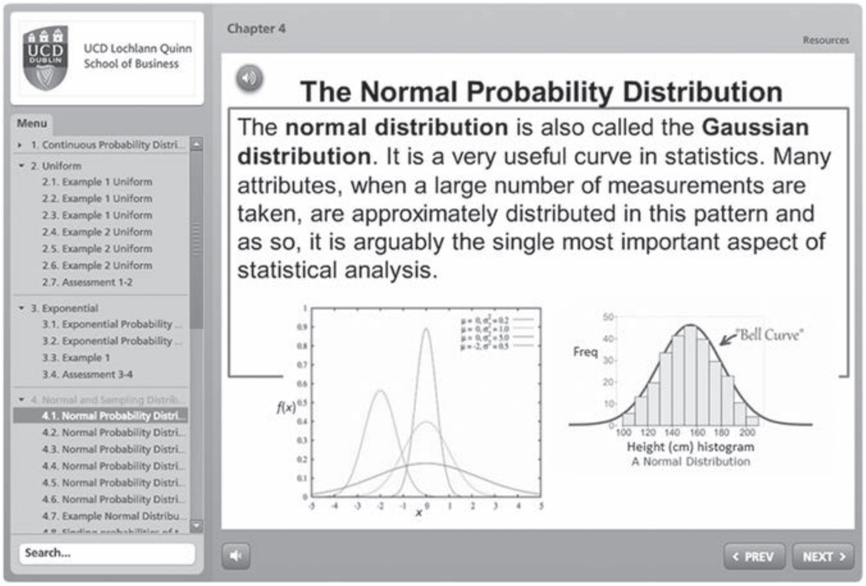
- Using technology to help develop conceptual understanding and for analysing data, and
- Using activities to improve learning and engage students during contact time.

The eLearning courseware was created using a story board approach. The Articulate software Articulate (2013) was used to achieve a consistent layout following

Table 1. DADM Schedule

Week	Lecture content	Tutorial content	CA task	Additional activity
1	Intro to data analysis and descriptive statistics	Intro to excel		Drop-in clinic
2	More descriptive statistics	Descriptive statistics	Excel work	Drop-in clinic
3	Basic probability	Graphs and tables	Excel work	Drop-in clinic & industry speaker
4	Discrete random variable probability distributions	Excel functions	Excel work	Excel training
5	Continuous random variable probability distributions	Excel test	Excel test	Drop-in clinic
6	Normal distributions	Probability exercises		MSC Hot topic & industry speaker
7	Sampling theory	MCQ1	MCQ1	Drop-in clinic
8	Confidence intervals	Probability exercises		MSC Hot topic & industry speaker
9	Confidence intervals	MCQ2	MCQ2	Drop-in clinic
10	Hypothesis testing	Inferential statistics exercises		Drop-in clinic & industry speaker
11	Hypothesis testing	Inferential statistics exercises	Team assignment	
12	Case study	Revision		

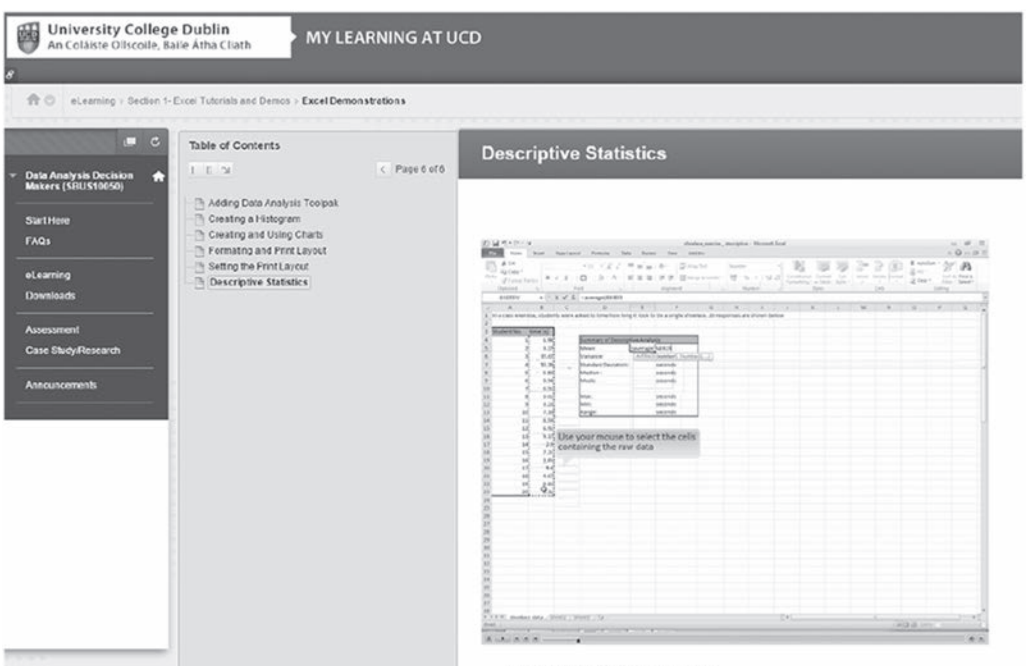
Figure 2. Online Learning Object, Articulate Example



the design principles in Mayer (2001), Mayer and Richard (2003). Content was divided into navigable sections and chapters. Figure 2 shows a sample page. The speaker symbol indicates that this page includes explanatory voiceover. The voiceover of this page says "Note in the figure on the left that there are many normal distributions. Which one we are dealing with is determined by the mean and the variance of the distribution. In the figure on the right we see some data that has been gathered and formed into a histogram. We see that the pattern in the histogram, if we were to throw a blanket over it, almost forms a normal distribution or bell shaped curve."

Links to a curated set of YouTube videos (such as the Khan Academy videos), java applets that, for example, demonstrate sampling distributions, and *How to in Excel* demonstrations were included at relevant locations in the material. Figure 3 shows a screenshot of a sample MS Excel demonstration. These were created using Adobe Captivate. Students can play the demonstration videos and copy the steps to practice implementing data analysis in MS Excel. The Excel

Figure 3. Online Learning Object, MS Excel Captivate Example



learning objects cover an introduction to spreadsheets, show how to enter data, how to use statistical and descriptive statistics functions, and how to draw charts and histograms.

Assessment holders linked to Blackboard self-assessment exercises were interspersed among the content. The Menu panel visible on the left of Figure 2 shows examples of tabs 2.7 and 3.4 which link to self-assessment exercises. Students were advised to attempt these short practical exercises in advance of tutorials. These exercises focused on how to implement the theoretical topic explained in that chapter. Students received automatic feedback showing the calculations steps but did not receive credit for completion of the practice exercises. Each chapter included a conclusions/summary section and finished with a set of review and more detailed *Theory in Practise* (TIP) case study exercises. A sample TIP focuses on the results from the Irish Leaving Certificate. Aggregate data are available from the State Examinations Commission (SEC) (2016). This case study is introduced when descriptive statistics are discussed. The steps to produce histograms in Excel are discussed and students are asked to comment on the shape of the resulting graphs for higher and ordinary level math. Students are then asked to create their own histograms for the exam results in English. Students were asked to estimate measures of central tendency and variation from the histograms and to think about how we might compare results between exam subjects. In general, the TIP topics were selected to be relevant to early stage business students. In this case, the majority of students had recently completed the state exams under discussion.

Materials were also made downloadable. Students could download zipped folders of the full content or a pdf *printable* version of the text content. These are supplemental and more traditional pdf copies of the text of the learning materials. This was to facilitate students who wished to print and bring the hard copy notes to face-to-face sessions.

2.1.1. DADM Assessment Components. The quantitative black-and-white nature of many mathematically-based topics are especially suitable for computer aided assessment (Gordon 2014). However, DADM requires student reflection on interpretation of results and an assessment of how results support business decision making. A mix of CA types are used. Table 2 shows a summary of the CA tasks which count for 40% of the module grade. The remaining 60% is a traditional end of semester exam.

Students submit their MS Excel tutorial work in weeks 2, 3, and 4; these are small tasks designed to encourage engagement and participation. An MS Excel test is given in week 5 where students have to complete a small spreadsheet, calculate some descriptive statistics, and draw a graph. Tutors grade the MS Excel tests.

Table 2. DADM Continuous Assessment Schedule

Week	CA task	Grade (%)
2	MS Excel work	1
3	MS Excel work	1
4	MS Excel work	1
5	MS Excel test	12
7	MCQ 1	7.5
9	MCQ 2	7.5
11	Team project	10

Online open book Multiple Choice Quizzes use closed questions, i.e., there is a single correct answer to a calculation based or definition type question. Students may access their own notes or online resources during the MCQs. The open book nature promotes discovery learning. The MCQs are run in the Blackboard VLE and are automatically corrected; students receive instantaneous feedback. Technology issues can arise since the School operates a bring your own device policy.

A team project asks students to summarise their analysis and interpretation of business data sets in a written report. The topic for analysis and type of data analysis is chosen by the team and thus allows for more open exploration of the DADM concepts. The written report requires an interpretive qualitative deliverable from the students, which also demonstrates their understanding of the quantitative data analysis tasks. This open piece of assessment is corrected by lecturers; the SafeAssign tool is used to deter plagiarism.

2.1.2. DADM Student Guide. With careful thought having gone into the design of the learning objects, a student guide summarises the blended learning teaching philosophy of the DADM module as well as the schedule of lecture, tutorial, and assessment activities. The objective is to manage student expectations and assist students in becoming independent learners. The full guide contains details of each of the CA components as well as information about the university policies on plagiarism, late submission, and the group conduct code of behaviour.

Early stage students face the challenge of transitioning from a teacher-led secondary school learning approach to an independent learning style where students become responsible for their own learning. In a flexible learning environment, the choice of which learning resources best support their learning style is theirs. Flexible access to learning resources with opportunities for formative assessment and feedback are considered essential to support learning in JISC (2009). Since blended learning environments offer students opportunities for independent learning, blended learning should help facilitate student engagement in their academic work. Student engagement in academic work has been identified in Pascarella and Terenzini (2005) as a factor in student knowledge acquisition. The authors argue that academic engagement reduces authoritarianism and increases autonomy and independence.

Taylor argues that early stage students prefer the traditional lecture (Taylor 2000). He further argues that students are likely to resist their new roles as self-directed learners or to express considerable dissatisfaction because their expectations were not met. He suggests that early stage students have not yet developed the ability to adopt flexible learning strategies. This idea is echoed in Hourigan and O'Donoghue (2007) in relation to the situation in Ireland. The authors describe the impact of exam-focused secondary school teaching in Ireland on the preparedness of students to make a successful transition to university. As noted, Leaving Certificate results are the main criterion that determine entrance to higher education. Many Irish students find the transition to becoming a self-directed learner particularly challenging.

With these challenges in mind, the study guide and a set of responses to frequently asked questions are made available at the "StartHere" and "FAQ" tabs as shown on the left of Figure 3.

3. Methodology

The problems of data gathering by survey are known in Chen et al. (2014). Students are asked to reflect on what they were thinking and doing sometime in the past; their recollections may have become blurred over time or coloured by assessment grades. Socially Desirable Responding (SDR) is a cause of survey bias where respondents, perhaps unconsciously, tend to give answers that make them look good.

Additional problems are associated with the use of anonymous online surveys. UCD uses anonymous online evaluations at the end of each module. There was a 24% response rate to the 2013/2014 DADM evaluation. Ninety-three percent of respondents said they had attended more than two thirds of face-to-face sessions while 42% said they attended 100%. The established attendance is detailed in Section 4, which shows that the survey is biased toward regular attenders or that the results suffer from SDR or inaccurate recall. In either case, it makes interpretation of such survey feedback challenging.

3.1. Data to be Analysed

An alternative approach to gain insight to student experience is use of LA on recorded face-to-face attendance and online activity. LA is among the innovations identified in Sharples et al. (2013) as having a potentially profound influence on education. As the DADM online resources are managed in the Blackboard VLE, learner interaction can be captured for subsequent analysis. Student-VLE interaction is reflected by the solid heavy line in Figure 1.

Following receipt of ethical approval from the university research ethics committee, we informed students at the start of the semester about the research study and its objectives. Students' online activity was extracted from the VLE. Students swiped their student cards at a barcode scanner observed by the lecturer or tutor during contact sessions. UCD uses a system of smart cards; by using barcode scanners, only IDs and no other student information was recorded. Learner attendance is reflected by the solid light line in Figure 1. As an incentive to participate in the study, we used the attendance data to form teams for the CA team project. A common complaint in team work is the formation of teams and the non-participation of some members.

We were initially concerned that this attendance data gathering process would take too much time and might result in data integrity issues with students swiping cards for absent friends. However, it rarely took more than a couple of minutes and gave rise to some interesting discussions on queueing theory. We are not aware of students swiping multiple cards. Students would occasionally forget their card, so their ID was manually added to the file. We also recorded the semester week, session type (lecture or tutorial), and lecture/tutorial group.

Student age, gender, and Leaving Cert or access route information were provided by the university registry office. The full data sets were collated and anonymised for research purposes.

3.2. Detecting Learner Behaviour Patterns

In terms of Baker's EDL/LA taxonomy (Baker et al. 2014), this study is concerned with *Structure Discovery*, i.e., determining what structure emerges naturally from the data rather than using the data to predict outcomes. We use latent class analysis (LCA) to analyse the learning behaviours of 524 students taking the DADM module to determine whether distinct clusters of student behaviour exist, and to identify which students belong to them. LCA (Goodman 1974) is a model-based clustering method (Fraley and Raftery 2007) for multivariate categorical or binary data. An implementation is available in BayesLCA (White and Murphy 2014) on the open source statistical software platform **R** (R Core Team 2014).

We analyse student behaviour with respect to five learning objects over the 12 teaching weeks of the module delivery to try to understand how students are using the following resources:

1. lecture attendance;
2. tutorial attendance;

3. online scheduled: access to the scheduled eLearning material (i.e., accessing online resources in the week in which it was covered in face-to-face sessions);
4. print: access to printable versions of scheduled materials (i.e., accessing printable materials in the week in which it was covered in face-to-face sessions);
5. online full: access to any of the eLearning material, regardless of whether it was scheduled for that week.

We analyse 59 variables in total for each student: 12 weeks of binary (i.e., yes/no as to whether a contact session was attended or material was accessed) data for lectures, tutorials, unscheduled online material and printable material, and 11 weeks for scheduled online material since no new learning material was introduced during the last week of the module.

Two sets of parameters are estimated, i.e., the relative size of each cluster, and the proportion of each cluster that interacts with a given learning object during a given week. These can be used to characterise the data at a global level. Additionally, we estimated the cluster membership of each student. We use these to assess cluster attributes in a post-hoc analysis. To select the number of clusters, we used the Bayes information criterion (BIC).

4. Results and Analysis

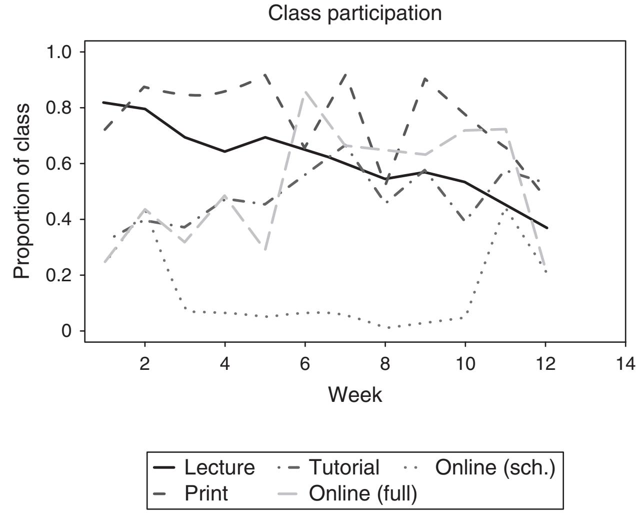
The majority of DADM students are early stage students, with 87% entering directly from the Irish secondary school system in the 2013/2014 academic year. Most had not been exposed to a blended learning environment before taking the DADM module. Students were on average 19.1 years of age at the start of semester. A breakdown of gender and Leaving Certificate mathematical level is shown in Table 3. Broadly speaking, students who took the higher level exam would be expected to have better mathematics ability.

Figure 4 shows the overall proportions of students who attended lectures, tutorials, and accessed various online materials on a weekly basis. There is a noticeable decline in lecture attendance over the duration of the module. Tutorial attendance peaks in Weeks 5, 7, and 9, which coincide with continuous assessments. Perhaps most notable is the lack of engagement with the scheduled online materials. That is, the proportion of students accessing the online materials during the weeks when that content is covered in contact sessions is low; between weeks 3 and 10 of the module, the percentage of students accessing the material ranges from 7% to a low of only 1%.

Table 3. Student Breakdown by Gender and Math Level

	Higher level	Ordinary level	Subtotals
Female	0.28	0.13	0.41
Male	0.40	0.19	0.59
Subtotals	0.68	0.32	1.0

Figure 4. Proportion of Students Attending Lectures, Tutorials or Engaging with Online Materials by Week



LCA was applied to the anonymised data, over a range of numbers of groups: $G = 1, \ldots, 10$. Using BIC, the optimal number of groups was found to be $G = 4$; see Table 4. The largest group accounted for about 34% of students, with the smallest containing about 11%. See Table 5.

We obtained qualitatively similar results when we considered clustering solutions obtained using different numbers of groups. Broadly speaking, for the five-cluster solution, the majority of students clustered together into Group 1 in the four-group solution were split into two groups. In a three-cluster solution, many of the students in Groups 2 and 3, whose behaviours are similar in several ways, were merged into a single group. We interpret the four- cluster solution in more detail below.

4.1. Interpreting the Clusters

Before directly discussing our research questions, we first interpret each of the groups based on their behaviour. The distinct behavioural patterns for the

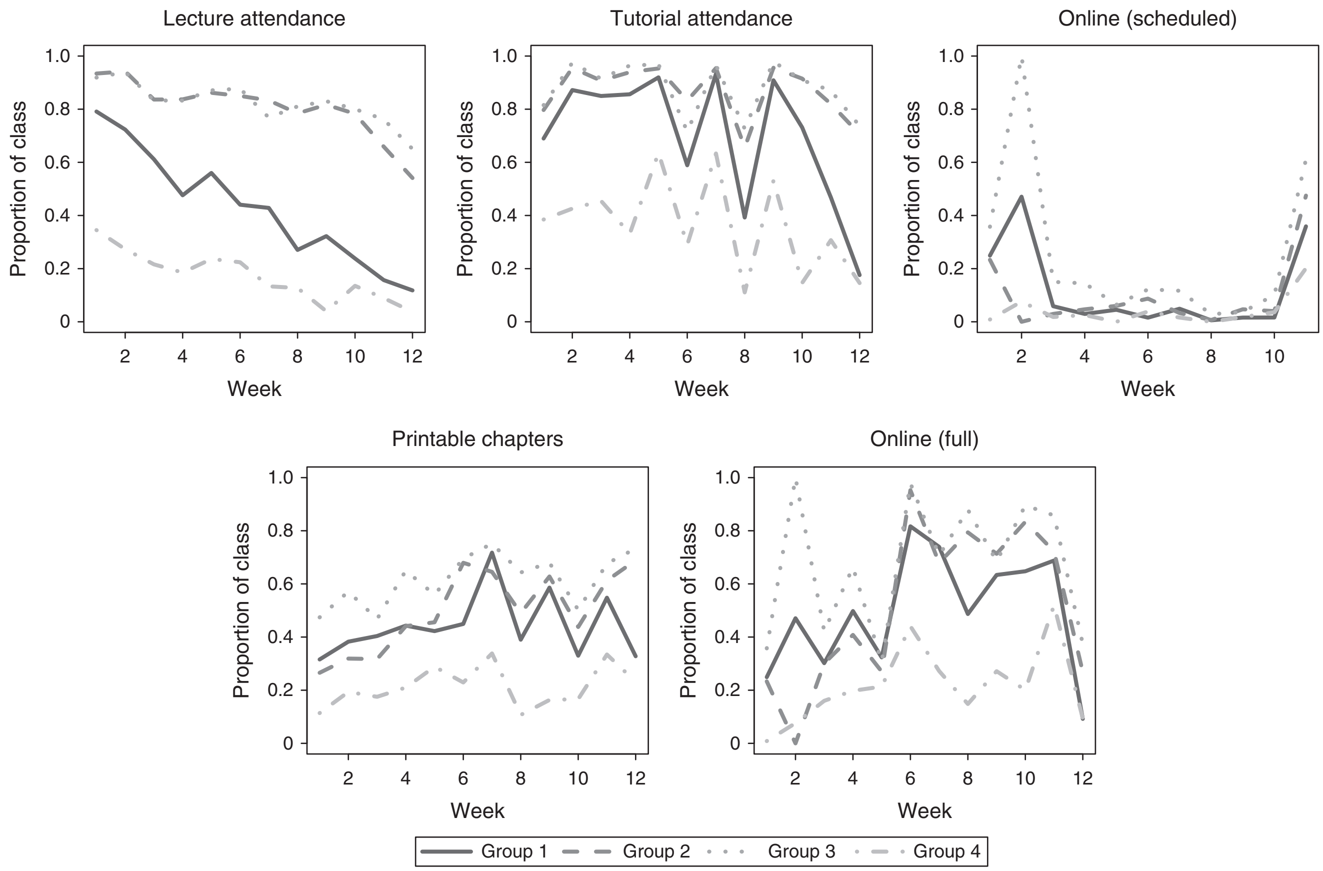
Table 4. Model Choice: Identifying the Optimal Number of Groups

	1	2	3	4	5	6
BIC	−34,241	−32,264	−31,966	**−31,812**	−31,822	−31,872

Table 5. Estimated Group Probability, Four Groups Solution

	Group 1	Group 2	Group 3	Group 4
Group proportions	0.34	0.28	0.27	0.11

Figure 5. Proportion of Clusters Attending Lectures and Tutorials or Engaging with Online Materials by Week



four groups are shown in Figure 5. We see that the groups use the learning resources in different ways.

Good intentions. Group 1 is the largest group of students. Lecture and tutorial attendance decrease sharply as the module progresses. The sharp peaks in tutorial attendance for CA tasks are most pronounced for this group. For example, compare the mean attendance of 92% for these tutorials with an attendance rate of only 11% for the final week's lecture. A similar trend is also present for printable chapters, although the rate at which the group accessed full online material improved for most of the second half of the course. Perhaps a realistic interpretation for this group is that while they may initially have good intentions for behavioural engagement, they struggle to perform unless directly incentivised.

Conscientious attenders, late online adopters. Group 2 have distinctly higher attendance for tutorials and lectures, with mean attendance of 86% and 80%, respectively, compared to Groups 1 and 4, although a slight downward trend appears to be present. This suggests that students were more conscientious in their attendance and less affected by direct incentives. This group was slower to engage with online materials, so that a majority of students in this cluster only routinely accessed online materials after week 6, i.e., halfway through the module.

Conscientious attenders, early online adopters. The behaviour of Group 3 is quite similar to that of Group 2 in that their engagement with lectures and tutorials is consistently high throughout the module delivery. Of the four groups, students in this cluster consistently displayed the highest rates of engagement with online materials. This distinction was especially clear in the initial weeks of the course, which suggests that students in this cluster adopted the online material earlier than their peers.

Poorly engaged. The smallest group, Group 4, display the lowest levels of engagement with respect to all learning resources. For example, the highest attendance rates of this cluster at tutorials occur in weeks 5 and 7, when assessments were held (63% in both cases). This is still lower than the tutorial attendance of Groups 2 and 3 for any week. This group's lack of engagement is particularly pronounced during the first six weeks of the course.

We considered the timing of student access to learning resources compared with their total participation. We compared LCA models of weekly scheduled student engagement to a cluster analysis of the aggregated attendance and online participation over the duration of the course. A k-means algorithm was applied to this data. Timely consideration of the data allows us to obtain additional information. For example, for students clustered into Group 1 under the LCA model, the decline in lecture attendance over time is clear, whereas their online use generally increases over the duration of the course. When clustering with the aggregated data, it is unclear if the student's behaviour is changing over time or whether they are simply engaging with the material less sporadically than the least engaged group of students. Thus, it is more difficult to distinguish between k-means clusters. Other trends are also identifiable by the LCA model: For example, the belief that many student's attendance levels are affected by incentivisation could only be conjectured when examining the aggregated data, but is clearly demonstrated under the LCA model, particularly for Groups 1 and 4.

4.2. Answering the Research Questions

We now discuss the research questions posed in Section 1. First, attendance at lecture and tutorial activities is established. If we examine the top row of figures in Figure 5, we can see that of the four groups, Groups 2 and 3 have attendance regularly above 80% throughout, while the attendance of students in Group 1 substantially declines to levels below 30% by Week 12 for lectures and tutorials. Students in Group 4 have poor attendance throughout the module. It appears that Group 1 is incentivised by continuous assessment, which is not unusual with early stage students.

4.2.1. What Are the Identifiable Behavioural Patterns of Learning Object Use? Four distinct groups were identified by LCA. Recall that DADM is taken by all stage one business students in semester one, the bulk of whom have come directly from the secondary school system. The patterns of behaviours show that some groups are adapting well and transitioning to self-directed learning over the course of the semester.

In Figure 4, the dotted line shows the percentage of students that accessed the online materials at the same time as those materials were being covered in lectures. The value is consistently low, and, with the exception of weeks 2 and 11, below 20%. These values remain low when we consider them by cluster also (see the left figure in the middle row of Figure 5). However, if we consider the proportion of students who access online material regardless of whether the content matches the scheduled lecture content, the rate of engagement steadily increases for the most part, by contrast to the lecture and tutorial levels. The same holds true for students accessing the printable material. This is consistent with a lag between material being covered at lectures and that being accessed by students at a later time. This is possibly for revision purposes for those who attended lectures but possibly a first view for students who did not attend and are subsequently prompted by a piece of CA.

Inspecting this behaviour by cluster, we see that Group 3 regularly accessed the online material, while as before, Group 4 had the lowest levels of activity of the four groups. In terms of online activity, Groups 1 and 2 behaved similarly, with a gradually increasing engagement with online material. The behaviour of Group 1 is slightly more sporadic. This again may reflect the timing of assessment, with students in this group more likely to access online material during the weeks of an assessment.

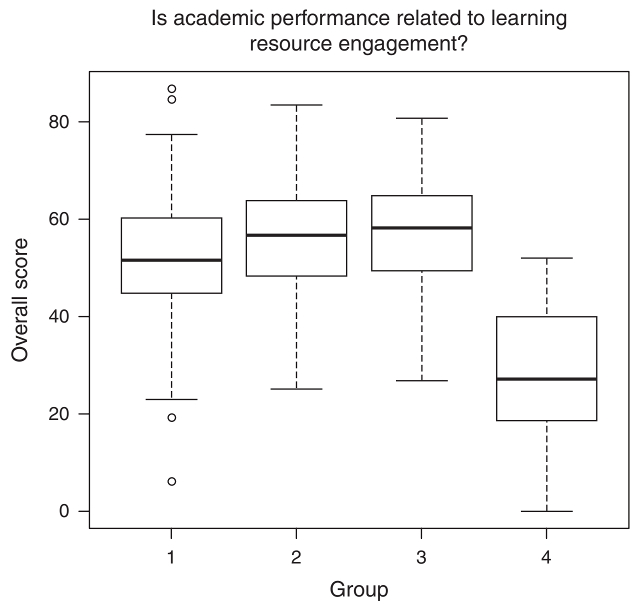
Biggs (1999) contends that some students have an innate knowledge of how to learn. Students such as those in Group 3 seem to have the flexibility to quickly adapt to and take advantage of a blended learning environment. Group 2 adapted more slowly. Groups 1 and 4 have not yet become self-directed learners by the end of the DADM stage one semester one module and may need further support.

Examining the four groups together, there do not seem to be clusters of students with strong fixed preferences between different types of engagement with the learning objects. However, we note that Groups 1 and 2 substantially increase their engagement with online material as the module progresses. In particular, Group 1 seems to do this at the expense of lecture attendance. This class predominantly consists of first year students new to university learning, which suggests that students may still be discovering which learning resources they prefer. Perhaps many will more strongly prefer using online material to attending lectures in future.

It is also apparent from Figure 5 that students in all groups access the printable version of the materials in a more *timely* fashion than the Online (scheduled) materials. Technology Enhanced Learning (TEL) offers students opportunities for personalised learning. Students find their own pathway through learning material at a pace and place that suits their learning style. Students who are developing self-directed learning skills may find the transition to a blended learning environment challenging. This may be an indication of the point raised in Tinto (1993): Early stage students are reluctant to abandon their traditional learning behaviours. The facility to access printable versions of the content may act as a comfort blanket to early stage students.

However, the increased use of Online (Full) activity by all groups in the second half of the semester suggests that most students do find their way to the online resources. We did note a small number of students with an excessively high number of hits on the same online resources in a short time frame. Perhaps this indicates that some students experienced difficulty navigating the online resources or have insufficiently developed IT literacy skills.

Figure 6. Final Score for Students by Cluster



4.2.2. How Does the Pattern of Learner Behaviour Relate to Levels of Attainment of Learning Outcomes?

Grade bands are used at UCD to show the level of attainment of learning outcomes. A score of at least 40 is a pass grade. A student's final grade is a weighted average across the assessment components described in Section 2.1.1. Performance by cluster is shown in the boxplot shown in Figure 6.

There is a clear performance discrepancy between Group 4 and the other groups. However, Groups 2 and 3 also appear to perform slightly better overall than Group 1. Performing an analysis of variance (ANOVA) test of performance based on group membership, we obtain a highly significant result ($F_{3,519} = 84.70$, p-value < 0.001.).

We can also perform pairwise comparisons of the groups. These comparisons are shown in Table 6. From this table, we see that, unsurprisingly, the performance of Group 4 is markedly worse than the performance of the other groups, with a mean difference in overall score lower than 20% in all cases.

Groups 3 and 1 are also significantly different, although by a smaller margin of about 4.8%. A 95% confidence interval for this score difference is (1.29, 8.27).

Table 6. Group Score Differences—Upper and Lower Bounds at 95% Confidence

Group mean comparison	Difference	Lower	Upper	Adj. p-value
Group 2 – Group 1	3.31	−0.12	6.73	0.06
Group 3 – Group 1	4.78	1.29	8.27	0.00
Group 4 – Group 1	−24.31	−29.11	−19.51	0.00
Group 3 – Group 2	1.47	−2.18	5.13	0.73
Group 4 – Group 2	−27.62	−32.54	−22.70	0.00
Group 4 – Group 3	−29.09	−34.05	−24.13	0.00

Table 7. Breakdown of the Number of Students Who Passed and Failed by Group

	Group 1	Group 2	Group 3	Group 4
Fail	23	15	13	41
Pass	159	133	126	13

The differences in overall score between Groups 1 and 2 and Groups 2 and 3 are not significantly different at a 5% level, however, there is evidence of a difference between Groups 1 and 2 at the 10% level. In summary, Groups 2 and 3, who are the more regular attenders, do marginally better in demonstrating attainment of learning outcomes. Group 3, who also adopts early use of the online resource, has the least variable academic performance.

Examining the pass/fail rates of students, we see that, unsurprisingly, the majority of students in Group 4 failed the module. See Table 7 for a breakdown by group of the number of students who passed and failed; Figure 7 presents this information in bar chart form. Removing Group 4, a χ^2 test suggests that there is no discernible difference between the failure rates of Groups 1, 2, and 3. Despite their different behaviours, there is no significant impact on their fail rate. It could be argued that Group 1 are the more strategic learners rather than simply being students with "good intentions." Their pass rate is similar to Groups 2 and 3 even though their average grade is slightly lower. Note that the DADM module is taken by stage one students and that the grade does not count toward their final degree calculation. Group 1 in particular demonstrate an assessment driven interaction with the learning objects. While these students may be playing a strategic game and balancing competing demands from their academic and personal lives, their lower attainment may reflect a shallower approach to learning in this module.

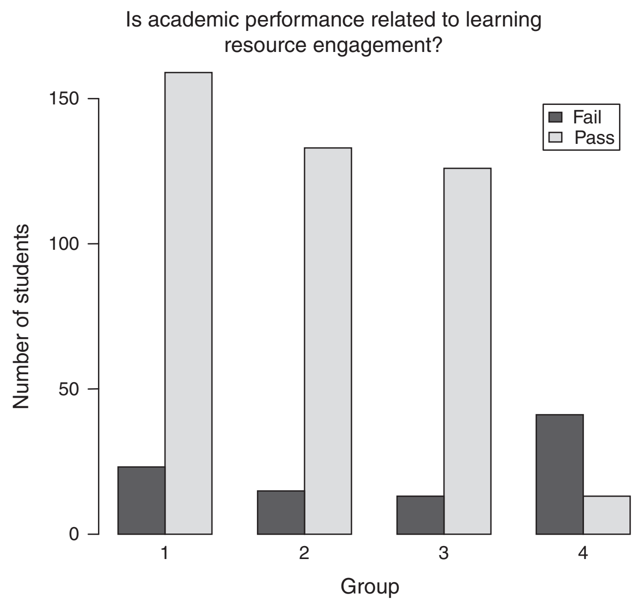
4.3. Additional Analysis

As an additional analysis, we considered whether a significant relationship existed between students' cluster membership and their sex and previous performance in the Leaving Certificate, in terms of the points they achieved in English, Mathematics, and overall. Points are awarded depending on the grade achieved and level of difficulty (higher, ordinary or foundation), with a larger number of points achieved reflecting a better performance.

We found that:

- Gender did not have a significant relationship with cluster membership at a 5% significance level. ($\chi^2_3 = 7.06$; $p = 0.07$.)
- Performance in English did not have a significant relationship with cluster membership. ($F_{3,347}$; $p = 0.49$.)

Figure 7. Number of Students Who Passed or Failed This Module by Cluster



• Differences in mathematics performance existed between the clusters. ($F_{3,392} = 7.58$; $p < 0.001$.) See Figure 8. In particular, students in Group 4 had worse performance at math, and were more likely to have taken ordinary level math. ($\chi^2_3 = 11.4983$; $p = 0.009$.) The majority (approx. 75%) of students in Group 4 performed worse than the median scores of the other groups, whose performance is almost identical.

• Overall performance in the Leaving Cert. was related to cluster membership. ($F_{3,442} = 6.664$; $p < 0.001$.) See Figure 9.

• Using a multiway ANOVA, overall Leaving Cert points, math result and cluster membership were all found to have significant bearing ($p < 0.001$, in all cases) on overall score in the DAM.

5. Discussion

The main contribution of our paper is the application of LCA to discover the structure of distinct early stage business student learner behaviours in a core data analysis module. In particular, we have identified groups of students exhibiting diverse learner behaviours based on the *timeliness* of their interaction with learning objects. In summary:

• We applied LCA to discover the structure of early stage student learner behaviours. The suite of learning objects was used by different groups of students in different ways. Students exhibited distinct behavioural patterns of attendance and engagement with online resources.

• Timeliness was an important distinguishing feature of these patterns. Students with low engagement across learning objects during the first half of the course were most at risk of failing.

• Perhaps unsurprisingly, we found an association between poor mathematics skills and low student engagement.

Figure 8. Boxplot of Leaving Cert. Mathematics Results (in Terms of Points Awarded) by Cluster

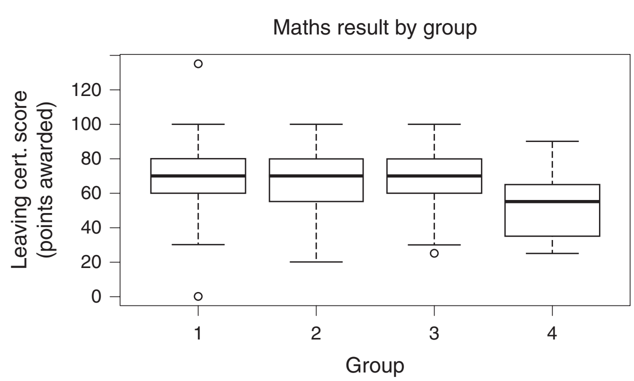
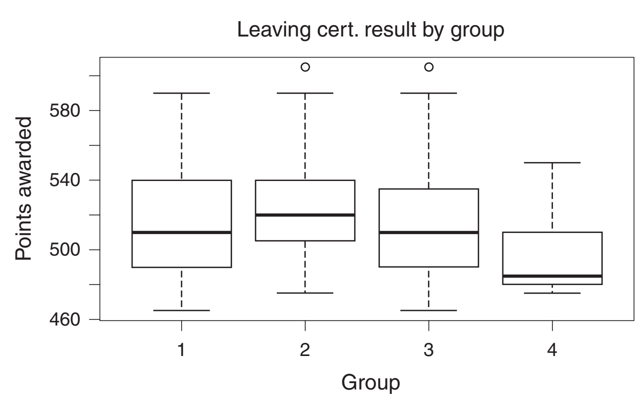


Figure 9. Boxplot of Leaving Cert. Results by Cluster



Note that the majority (approx. 75%) of students in Group 4 performed worse than the median scores achieved by Groups 1, 2, and 3.

- A substantial number of students (34%) appeared to require assessment prompted engagement. Despite inconsistent attendance, these students still demonstrated acceptable (but quite variable) academic performance.
- There was evidence to suggest that some early-stage students were developing a preference for online material instead of lectures and tutorials.
- There was no evidence to suggest that early-stage students consistently accessed online material in an appropriately scheduled timeframe.

Returning to our ecosystem in Figure 1, the weights of the connections between the learning objects and the different groups of students vary considerably. Students in Groups 2 and 3 are better connected to the biotic components, i.e., lecturers and tutors, than the other groups. These well maintained connections may play a role in providing structure and managing workload for students. Group 3 have the strongest connections to all the learning objects and the least variable academic performance. This group seem to confirm Taylor's assertion that early stage students prefer the traditional lecture format; yet their early adoption of online resources conflicts with his idea that they are incapable of adopting flexible learning strategies (Taylor 2000).

The insights about student learning behaviour gleaned from our analysis can inform future development of learning objects. In the main, the designed learning objects have had the desired impact. It may be disheartening for the lecturer who faces a half-empty lecture hall but our analysis shows that certain types of students are consistently choosing to attend. Our study shows the diversity in the learning behaviour among the student body and indicates that students tailor their personal learning experience in their use of the learning resources with the majority adapting well to the university blended learning environment.

In particular, Groups 2 and 3 accounted for 55% and these students seem to be successfully making the transition to self-directed learners at university in their use of the learning objects. Because Group 3 may have an innate ability to learn, as described by Biggs, they might play a role in peer mentoring future students. It is clear from the pass rates that all but Group 4 have or are developing successful university learning strategies. Group 4, with evidence of generally weaker initial mathematical ability, may benefit from further directed support in general university study skills as well as from additional tailored resources in the DADM module.

Group 1 can be considered "successful" if their objective is to pass the module. The variability of their results suggests they may have the potential to improve their performance. Our study shows evidence of a *Just-in-Time* approach to learning by many early stage students. Assessment prompted engagement is particularly evident for Group 1 type students, who accounted for 34% of the students in this study. Group 1 type students warrant further analysis and research. We need a better understanding of what motivates them, which new TEL innovations they might prefer, and how best to support the development of their learning habits. Group 1 students may be strategic game players. This forces us to ask if other forms of assessment, such as business simulation games, might be more appealing to them, and might lead to deeper learning. However, such games might not appeal to the other groups and would require additional resources to develop. A careful balance of continuous assessment that: (1) does not disenfranchise Groups 2 and 3 type students, and (2) weans Group 1 type students from an assessment dependency culture, is required in blended learning early stage modules. Some Group 1 students may also overestimate their own capabilities and underestimate the work involved in meeting the learning outcomes without attending lectures.

The more consistent use of the printable resources over the module delivery (compared to access to scheduled online resources) suggests that multiple forms of content availability should be considered. There are resource implications in facilitating this type of choice, whereby additional flexibility comes at an increased staff cost in terms of preparation (Gordon 2014).

Questions of the ethics of using student VLE data also arise. Policy decisions at the institute level need to be made before LA can be applied at the module and individual student level. Early intervention styled on the *Signals* system described in Chen et al. (2014) for Group 4 type students could help with retention, however issues of privacy and resource availability arise.

The apparent correlation between attendance and academic success has led some institutions and staff to adopt more rigorous attendance requirements (Gordon 2014). Some may take the view that students are making poor choices in their use of learning resources so the decision should be taken out of students' hands and attendance enforced. Issues of institutional and administrative support for such an approach arise and effectively move the higher education environment back toward a secondary school environment with academics taking on the role of attendance enforcers. This shifts the responsibility for learning from the student back to the *teacher* and, in our opinion, would disenfranchise students such as those in Groups 2 and 3.

Furthermore, this approach risks conflating the observational study with the experiment. The apparent benefits in attendance may be due to hidden characteristics of regular attenders. The benefits of forcing students to attend are unclear. Such action may lead to poor engagement and an unconstructive learning atmosphere for regular attenders. An authoritarian approach may further disenfranchise nonattenders such as Group 4 type students and does not, in our opinion, allow Group 1 type students to discover for themselves the benefits of lecture attendance. The risk of over dependence on continuous assessment and monitored attendance and the lack of an opportunity to develop an independent learning style are apparent. Our study provides some evidence that students in Group 1 (the "Good Intentions" cluster) are in the process of developing their own individual independent learning styles. Awarding some marks for timely attempts at online assessment exercises, which are linked to lecture activities, could prompt better engagement from Group 1 type students. This is being given careful consideration for future iterations of the module in seeking a balance of encouraging students to take responsibility for their learning and fostering positive learning behaviours.

In summary, all the learning objects designed for the DADM module are used by the different groups of students but in different ways. Despite the resource implications of maintaining and developing the array of resources, we recommend this approach as suitable in analytics education, particularly for students not majoring in quantitative disciplines. We believe that by offering a blend of online learning resources and face-to-face contact sessions, the majority of business students can be supported to achieve the learning outcomes of analytics modules such as DADM.

A number of interesting research avenues arise from this study. In future work, we explore students' performance in the different assessment components. Because some students are particularly assessment driven, we will analyse the assessment components in an assurance of learning framework.

Acknowledgments

The authors thank the Business eLearning team at UCD for their advice and support in developing the online learning resources. The authors also thank the students of the DADM module for their participation in our research project.

References

Articulate (2013) Articulate storyline. Accessed November 25, 2016, http://www.articulate.com/.

ASA, American Statistical Association (2005) Guidelines for assessment and instruction in statistics education. Technical report, American Statistical Association, Alexandria, VA.

ASA, American Statistical Association (2016) Guidelines for assessment and instruction in statistics education GAISE draft college report. Technical report, American Statistical Association, Alexandria, VA.

Baker R, Ryan S Inventado RS (2014) Educational data mining and learning analytics. Larusson JA, White B, eds. *Learning Analytics* (Springer, New York), 61–75.

Biggs J (1999) What the student does: Teaching for enhanced learning. *Higher Ed. Res. Development* 18(1):57–75.

Chen X, Vorvoreanu M, Madhavan K (2014) Mining social media data for understanding students' learning experiences. *Learn. Tech., IEEE Trans.* 7(3):246–259. doi:10.1109/TLT.2013.2296520.

Cronin A, Carroll P (2015) Engaging business students in quantitative skills development. *eJournal Bus. Ed. Scholarship of Teaching* 9(1):119–131.

Fraley C, Raftery A (2007) Model-based methods of classification: Using the mclust software in chemometrics. *J. Statist. Software* 18(6):1–13.

Goodman LA (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61(2): 215–231.

Gordon N (2014) Flexible pedagogies: Technology-enhanced learning. Technical report, Higher Education Academy, York, UK.

Gütl C, Chang V (2008) Ecosystem-based theoretical models for learning in environments of the 21st century. *Internat. J. Emerging Tech. Learn.* 3(2008):50–60.

Hourigan M, O'Donoghue J (2007) Mathematical under-preparedness: The influence of the pre-tertiary mathematics experience on students' ability to make a successful transition to tertiary level mathematics courses in ireland. *Internat. J. Math. Ed. Sci. Tech.* 38(4):461–476.

IEEE (2002) Draft standard for learning object metadata. ieee standard 1484.12.1. Technical report, Learning Technology Standards Committee, Institute of Electrical and Electronics Engineers, New York.

JISC (2009) Effective practice in a digital age: A guide to technology-enhanced learning and teaching. Technical report, Joint Information Systems Committee. Accessed November 25, 2016, http://www.jisc.ac.uk/media/documents/publications/effectivepracticedigitalage.pdf.

Koehler MJ, Mishra P, Cain W (2013) What is technological pedagogical content knowledge (TPACK)? *J. Ed.* 193(3):13–19.

Mayer R, Richard RM (2003) Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist* 38(1):43–52.

Mayer RE (2001) *Multimedia Learning* (Cambridge University Press, New York).

McAlevey LG, Sullivan JC (2001) Making statistics more effective for business? *Internat. J. Math. Ed. Sci. Tech.* 32(3):425–438.

Moore DS (1997) New pedagogy and new content: The case of statistics. *Internat. Statist. Rev./Revue Internationale de Statistique* 65(2):123–137.

Pascarella E, Terenzini P (2005) *How College Affects Students: A Third Decade of Research* (Jossey-Bass, San Francisco).

R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria). Accessed November 23, 2016, http://www.R-project.org/.

SEC (2016) State examinations committee.

Sharples M, McAndrew P, Weller M, Ferguson R, FitzGerald E, Hirst T, Gaved M (2013) Innovating pedagogy 2013. Technical report, Open University, Milton Keynes, UK.

Taylor PG (2000) Changing expectations: Preparing students for flexible learning. *Internat. J. Acad. Development* 5(2):107–115.

Tinto V (1993) *Leaving College: Rethinking the Causes and Cures of Student Attrition* (University of Chicago Press, Chicago).

White A, Murphy TB (2014) BayesLCA: An R package for Bayesian latent class analysis. *J. Statist. Software* 61(13):1–28. Accessed November 23, 2016, http://www.jstatsoft.org/v61/i13/.

Yilmaz MR (1996) The challenge of teaching statistics to non-specialists. *J. Statist. Ed.* 4(1), http://ww2.amstat.org/publications/jse/v4n1/yilmaz.html.