

Using Educational Data Mining to Identify and Analyze Student Learning Strategies in an Online Flipped Classroom

Randall S. Davies, Gove Allen, Conan Albrecht, Nesrin Bakir, & Ball, N.

Instructional Design

Learning Analytics

Education

Online

Learning

Analyzing the learning analytics from a course provides insights that can impact instructional design decisions. This study used educational data mining techniques, specifically a longitudinal k-means cluster analysis, to identify the strategies students used when completing the online portion of an online flipped spreadsheet course. An analysis of these results showed that students did tend to follow a specific learning strategy as they completed this course. However, students also self-regulated to some degree, based on the topic and context of specific lessons. These insights not only improve our understanding about the students taking the course, but they also provide guidance for how the instructional design of the course might be improved. Of note is the practical value of this proof-of-concept study in using educational data mining to improve the instructional design of a course.

1. Introduction

Our ability to create quality online instruction has improved greatly in the past decade. Technology improvements, and our understanding of how to create online learning opportunities, have facilitated these improvements. Moreover, for a variety of reasons, students often prefer to take online, blended, or flipped classes that are facilitated by technology [1]. Still, there is much to understand about how students go about learning in an online setting and how we can use data analytics to improve this type of instruction.

One of the advantages of using online instruction is the ability to track students' activity within the course. Educational data mining capabilities built into many online courses enable learning analytics [2,3]. Basic data analytics capture assessment data and are used to inform teachers and students about an individual's progress towards accomplishing the intended learning objectives of the course. However, carefully crafted data mining efforts can also enhance our

understanding of the course in terms of what students are doing, how we might best provide feedback, and where we might make improvements in the design of our instruction.

This research used the learning analytic capabilities of an online flipped spreadsheet course to identify the strategies students used when completing the online portion of the course. This was then used to better understand where the instruction was functioning well and how the course might be improved. The study is not intended to be a generalizable case for all online courses, as differences in the students and contexts of each course will inevitably lead to different results. However, these insights not only improve our understanding about the students taking this course, but they also provide guidance for how the instructional design of the course might be improved. The study, in this sense, provides a proof-of-concept case study for using educational data mining in order to improve a course. The purpose of this study is to demonstrate, through an authentic case, how information can be data mined in an unobtrusive manner from a learning management system, how these data can be used to better understand the ways in which students go about utilizing course resources, and how this understanding might then be used to inform instructional design improvement when needed.

Previous Work

The research exploring the ways in which students go about learning makes a distinction between a learning approach and a learning strategy. Learning approaches are typically described as either "deep learning" or "surface learning" approaches [4]. Educators and researchers typically praise the virtues of deep learning and devise ways to encourage surface learners to engage more fully in the learning activities provided in order to learn all they can. Unfortunately, students do not always have the academic goals that their instructors might expect. Sometimes, they only intend to attain a sufficient level of learning to earn the grade they want [5]. A criticism of many courses is that they are designed in such a way that deep learning is not rewarded and, in fact, not needed for students to pass a course. Students can often achieve their learning goals with surface learning alone [6]. Through a meta-cognitive process, students devise learning strategies to accomplish their learning goals. These learning strategies may be intended to achieve either surface or deep learning. The strategies students develop are often exposed by the interaction they have with the learning resources made available to them.

A related construct essential to developing learning strategies is that of self-regulation. Self-regulation involves the ability to manage and monitor one's behavior [7,8]. Without the ability to self-regulate, students would not be able to modify their learning strategies. However, the ability to self-regulate one's actions and behaviors does not mean they will; self-regulation is often ancillary to other affective traits and influences [5,7,8].

The learning strategies that students devise are based on personal factors, including a student's academic goals, learning preferences, their self-efficacy and locus of control, as well as their abilities for self-regulation [9,10]. Contextual factors that affect the learning strategies student choose include the difficulty of the task, the student's interest in the topic, as well as the affordances that the instructional design of the course provides to the students [11]. The strategies student use to accomplish instructional activities and learning tasks often reflect a student's desire to learn efficiently, but not always effectively [12]. There are many reasons for this, one reason being that students often have conflicting intentions—they have many courses to study at school and a limited time in which to do them [5]. Often students will modify or completely change their learning strategies as the course progresses. The way a student approaches a learning situation is not inherent, rather, it is developed by the learner and is often dependent on the learning context and situational demands of the course [4]. Understanding the strategies students use to complete courses can help educators and instructional designers improve their courses and may provide actionable information that informs how, and in what ways, an educator might remediate learning gaps and students' misconceptions [13]. In the past, research involving learning strategies has relied primarily on self-report instruments [14]. Self-report, as a data collection method, is notoriously unreliable and constitutes an obtrusive form of data collection. In previous studies, detailed records of the topic focus, media choice, and study times and durations, were difficult to collect. For example, understanding the strategies students use to complete an assignment might require students reporting the time they spent on each problem, where and when students referenced their textbooks, and how students progressed from their initial answers to their submitted answers. These data have been difficult to collect in reliable efficient ways. This is where educational data mining (EDM) comes into play [1].

EDM is a relatively new term applied to the developing methods educational researchers use to explore the increasingly large-scale data that come from various educational settings, primarily online learning situations. EDM uses a variety of methods to better understand students and the settings in which they learn [15]. Particular to this study is the use of longitudinal k-means cluster analysis to better understand students taking a particular online course. With advances in technology, and increases in technology-enabled instruction, researchers are able to gather considerably more information about the activities students engage in to complete the learning activities required for a course [1,15]. Capturing data within the system allows researchers to analyze the temporal order of the spontaneous individual activities of the students as they complete a course [16]. While this is an imperfect indication of the intents and actions of learners, it can allow researchers to obtain a more accurate description of the students' learning strategies, which can provide the basis for the real-time implementation of adaptive practices.

2. Methods

The setting for this study was an Introduction to Information Systems course. This online course provided the context for the authentic case being studied. The course covered both spreadsheet and database topics, but this study focused on the spreadsheet portion of the course only. The students in this course are typically undergraduate business students and are required to take the class. The class consists of both lecture and asynchronous computer lab sessions following a flipped classroom approach. Students complete assignments (i.e., the computer lab portion of the course) on a website provided by MyEducator, the publisher of the e-text used in the course. The website hosts the textbook and video instruction, as well as the graded assignments. During the lecture portion of the course, instructors review specific tasks and answer questions students may have. However, this particular study focuses on the online portion of the course only. All of the students have basic computing skills (Internet, word processing, and email). Although the course does not require students to have prior experience with Microsoft Excel, the students enter with a variety of spreadsheet skills. Students can move through the labs at their own pace, but the class session and online exams are scheduled.

Most of the instruction for the course is provided online via the MyEducator platform. The website includes a reader that presents the material to be learned, similar to a normal introductory textbook, with chapters and sections, key terms, and a glossary. Students read the textbook on their laptops and mobile devices, or they can listen to the text similar to the way they would listen to a podcast. Learning tools, such as flashcards for key terms, are also available. Each section of the text includes one or more video presentations. The videos are embedded within each web page alongside the text. This was designed to make access to both equally easy and, on the basis of previous evaluations and student comments, this seems to be the case (authors, 2013). The video content complements the text: students can choose to read, to watch video, or to do both. The videos have to be clicked by the student to be played. The data analytics built into the system creates a log of each student's activity as they proceed through the course.

2.1. Participants and Data Collection

This study used educational data mining techniques to analyze extant data gathered from students who completed the MyEducator spreadsheet course. Students taking the course were enrolled in multiple sections of the course at several universities. A total of 997 students were included in this analysis. Only those students who completed all the lessons and assignments required in the course were included in this study. The decision to exclude non-completers was deemed necessary, as the longitudinal aspect of the cluster analysis required a full set of data for each student. Most of the non-completers withdrew from the course prior to completing the second lesson. The remainder withdrew prior to completing the fourth lesson. As these data sets were incomplete, they could not be used for this particular analysis. Data were collected on the student actions taken in the online textbook reader and video player, as well as the actions taken within the Excel workbooks as students completed assignments. All the data used for the study were obtained with student approval, and were only used once the course was completed, following the studies of the Institutional Review Board approved protocols. None of the participants refused to have their data used. The system captured student behavior in four categories: reading, video watching, assignment access, and task guide views (see [Table 1](#)). In addition, the system recorded the order in which students completed various activities and the overlap in which they were completed. The grades that the students achieved on each assignment were also captured.

Table 1. Variables used in the k-means cluster analysis.

Variables	Code	Description
Reading	v	% of instructional text reading completed
Video	r	% of instructional videos watched
Assignments	a	how often students accessed assignments
Task Guide	t	number of times task guide was viewed (task by task)
Task Instructions	T	number of times task instructions were viewed
Order		order in which students completed instruction and assignments
Overlap		degree to which students moved between activities and task guides

Student reading was tracked by client-side scripts that updated the server every 15 s and during page unload. As students use the textbook reader, they scroll the browser window downward through the text. Whenever scrolling pauses long enough, the paragraphs in view are deemed read by the student. Although we cannot determine how carefully a student might have considered the material, the student viewing the text was assumed to be an indication that they read the material to some degree. Embedded videos were split into 5-s blocks and tracked by block. As students played a portion of the video using the inline player, the blocks that played were recorded as watched. Determining the quality of the reading or viewing is always beyond the ability of any research; however, this variable is an indication of quantity, not necessarily quality, on the part of the student.

In each lesson, a student begins an assignment by downloading an Excel workbook from the MyEducator website. Using Visual Basic for Applications (VBA), the programming language built into Excel, the workbook logs the student's progress as he or she completes the assignment and interacts with the MyEducator servers during submission. The data logs for each student tracks the cell inputs and actions as students work through each problem.

The students are presented with the worksheets necessary to complete the assignment, as well as a set of tools to manage both the completion and the submission of the assignment. Detailed instructions on assignment requirements are included in the workbook and can be opened as a local HTML file (the Instruction Sheet), or presented one step at a time, directly in Excel, within a floating window (the Task Guide). When students have completed their work, they use the Submit tool to have their work graded. While students are working through assignment requirements, the workbook records every change they make to a cell, as well as other activities, such as adding worksheets and creating charts. The workbook also keeps track of when it is opened, each time the instruction sheet is shown, each time the task guide is advanced to show another task, and when the workbook is submitted (see [Figure 1](#) for examples). The data collected by this logging process provides a detailed history of how the student completes the course activities.

Figure 1. Assignment, task guide, and data log examples.

The screenshot shows a window titled "Assignment Tasks" with a close button (X) in the top right. Below the title bar is a green header with the word "Boolean" and navigation arrows. The main content area contains a task instruction:

1 Use the AND function with appropriate arguments in cell H11 to determine if there is a force out at third base. There is a force out at third base if "Runner on 1st" and "Runner on 2nd" are both "Yes". [9 points]

Below the task instruction are two tables. The first table is titled "show task" and has three columns: #, Time, and Task.

#	Time	Task
18	10/7/2017 10:19:18 AM	1.4 Fly ball complete
19	10/7/2017 10:19:22 AM	1.5 Not 2 outs

The second table is titled "range modification" and has seven columns: #, Time, Worksheet, Address, Formula, Value, and Step #.

#	Time	Worksheet	Address	Formula	Value	Step #
20	10/7/2017 10:19:45 AM	Boolean Functions	J11	=NOT(G11=2)	TRUE	1
21	10/7/2017 10:19:56 AM	Boolean Functions	J12:J40	=NOT(G12=2)	TRUE	1

2.2. Data Analysis

For a variety of reasons, technology-enabled online education has increased dramatically in the past decade [17,18]. Among the many benefits of technology-enabled instruction includes the increased amount of data available to educators and researchers. [19] describe the situation as drowning in a digital ocean of data. Certainly, the expectation that educators use data in order to enable educational discussion is not new see [20,21]. However, because of the increased amount of data now available, there is an increased need to make sense of these data and, as a result, the fields of learning analytic knowledge (LAK) and educational data mining (EDM) have gained prominence. We simply do not know what data is valuable, how best to manage it, and what to do with the information derived from these data [22]. EDM, in particular, uses a variety of methods to better understand students and the settings in which they learn. Particular to this study is the use of longitudinal k-means cluster analysis to better understand students taking a particular online course (k = 4 with 10 iterations). Data mining involves sense making [1]. One method for identifying patterns in the data is that of cluster analysis. While a detailed explanation of how cluster analysis works is beyond the scope of this paper, suffice it to say that this study used a longitudinal k-means cluster analysis to identify the optimal groupings that represented the most common strategy patterns used by students to complete each of the ten spreadsheet lessons in this course. The computation was based on data mined from the activity logs. Data were organized, scaled, and normalized to identify which activities were undertaken and how often (i.e., the magnitude and order in which students engaged in specific activities). Table 1 presents the variables used to complete the cluster analysis.

Cluster groupings were analyzed longitudinally by lesson, meaning each student was assigned a strategy group for each lesson. The students' learning strategies for each lesson were compared to identify changes (i.e., variations) made by students in their learning strategies. While the cluster analysis identified three basic groups, some students tended to self-regulate from lesson to lesson. A student's main strategy group was determined based on the strategy group a student followed most often (i.e., 50% of the time or more). Those who followed two strategies equally, or did not follow any strategy consistently, were not assigned a main strategy grouping. Group descriptive statistics were analyzed to help label each group's characteristics.

2.3. Learning Strategy Patterns

While the cluster analysis provides the cluster grouping based on the variables provided, researchers must still make sense of the grouping mathematically obtained. In order to do this, a student activity pattern was created for each student using a string of activity action codes. This was done in order to provide a human-friendly view into student strategies and allow researchers to visually inspect student strategies. Each letter represents the completion of about 10% of the different learning activities, although students may have completed activities such as accessing the assignments repetitively. The degree to which students accessed the assignments is indicated in the assignment variable, while the pattern only indicates when students completed at least 10% of the assignment. [Table 2](#) presents an example of one student's activity pattern. Note that each assignment provides a task guide for each part of the assignment, as well as the option to view all the task instructions at once. Students could view the task guides individually (represented by the letter t) or, optionally, they could view all the task instructions at once (represented by the capital letter T). The pattern is slightly different from the task guide and instruction variables used in the cluster analysis in that those variables represent how often the task guide and instructions were accessed, whereas the activity pattern depicts how much of the instructions were viewed. In the example presented in [Table 2](#), the student viewed all the task guides (not always the case) but did not use the task instructions option (which shows the entire task guide at once). This student may have viewed the task guide more than once, which is captured in the cluster analysis variables. The pattern codes were created for the researchers to better understand and interpret the student's learning strategy.

Table 2. Activity code example for an individual student: rrrrvvrrtttaataatata.

Code	Description
rrrr	started by reading 40% of the lesson text
vvv	watched 30% of the video blocks in that lesson
rr	read another 20% of the chapter text
ttt	viewed 30% of the task guides in the lesson assignment
aa	worked on 20%+ of the primary assignment
tataatata	continued by alternating between task guide and assignment

3. Results

The optimized cluster analysis results identified three learning strategy groupings. [Table 3](#) presents a description of these groups (including pattern examples) and the proportion of students who follow each of the strategies a majority of the time. Most students (58%) taking this course followed what we have labeled a Knowledgeable Confident strategy. These students completed less than 50% of the reading and viewed little of the video instruction (less than 4%). Primarily, they worked on the assignments quickly and, on average, achieved high scores. A second group (21%) we labeled Novice Careful. These students completed much of the reading (63%, more so in the earlier lessons), viewed a moderate amount of the video (29%, especially in later lessons), and tended to access the assignments more, with a higher number showing task use and task assignment overlap. The students in this group were either unfamiliar with the topic or, perhaps, were simply being careful or diligent. The last group we labeled Confident Traditional. Only 14% of students followed this strategy a majority of the time. These students completed a moderate amount of the reading (52%), only watched about 10% of the video content, and tended to complete assignments with little task guide use or overlap. The last group (7% of students) did not follow any one strategy to any great extent or switched equally between two strategy groups. Achievement by group and lesson is presented in [Table 4](#). Overall, there was little difference in

achievement between groups or lessons. While the difference in means were statistically significant [$F(3,9147) = 32.26, p < 0.001, \eta^2 = 0.007$], the group averages were less than two points apart, and the practical significance was negligible (less than 1% of the variance was explained by the students' main learning strategy).

Table 3. Learning strategy main grouping patterns.

Approach Strategy Description	Overall Proportion	Sample Activity Patterns
Novice Careful Higher reading (63%), more so in the earlier lessons. Moderate video use (29%), especially in later lessons. Higher showing task use and task assignment overlap.	21%	Rrrrrrvttatartaatataata rrrrrrTtaataataataa
Confident Traditional Moderate reading (52%). Low video use (10%). Low task description use. Low task assignment overlap.	14%	Rrrtataaraataaaa rrrrTataaataaaa
Knowledgeable Confident Lower reading (41%). Low video use (4%). Higher show task use and task assignment overlap.	58%	Rrtataaataataaa rtataaaaaaa
No Main Group Followed a particular strategy less than 50% of the time.	7%	

Table 4. Strategy group average scores by lesson.

Learning Strategy Group	Lesson										Average
	1	2	3	4	5	6	7	8	9	10	
Novice Careful	96	97	94	97	99	92	89	98	97	97	96
Confident Traditional	95	95	94	97	98	90	90	97	96	97	95
Knowledgeable Confident	96	97	98	99	99	92	93	98	98	98	97
No Main Group	96	97	96	98	98	89	90	99	98	99	96

3.1. Self-Regulated Patterns

There are several reasons that might explain why students switch learning strategies. The degree to which students tended to switch between strategies is presented in [Table 5](#). The overall strategy students tended to use by lesson is presented in [Table 6](#). Of note is that students tended to switch from their main strategy in Lessons 1 and 10, but stick to their main strategy in Lessons 2 through 9. In Lesson 1 (Excel Basics), most students used a Confident Traditional strategy (64%) before moving to a more stable main strategy. In Lesson 10 (optimization using Excel's Solver feature, arguably a more difficult lesson), 79% of students chose a Confident Traditional strategy, with most moving from the Knowledgeable Confident strategy. Other than in Lesson 1 and Lesson 5, those students in the Novice Careful group seemed to stick to their main strategy most consistently. Lesson 5 (Charts and Graphs) was arguably the easiest lesson. In this lesson, most students (78%) chose to follow the Knowledgeable Confident strategy, with many from the Novice Careful group changing strategies for this lesson.

Table 5. Percent of those with a main strategy who changed from main strategy group by lesson.

	Lesson										Avg. Overall % Who Changed
	1	2	3	4	5	6	7	8	9	10	
% who changed	56	17	19	12	24	25	17	17	15	71	27%

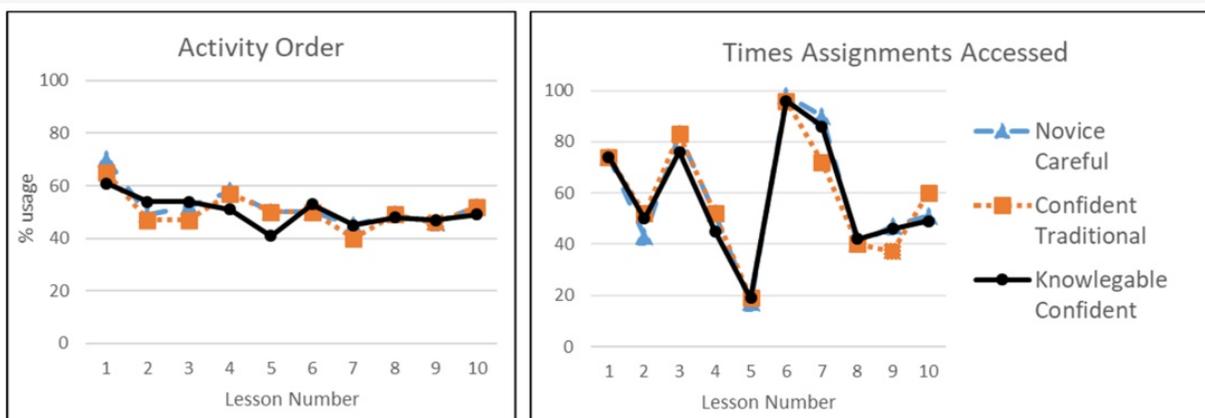
Table 6. Percent following a specific strategy by lesson.

	Lesson										% in Main Group
	1	2	3	4	5	6	7	8	9	10	
Novice Careful	8.6	24.0	27.4	26.4	5.6	35.7	29.4	23.6	20.7	20.7	21
Confident Traditional	64.0	17.5	13.4	15.1	16.4	20.4	17.0	16.9	15.9	78.7	14
Knowledgeable Confident	27.4	58.5	59.2	58.5	78.0	43.8	53.6	59.5	63.4	0.5	58

3.2. Common Strategies and Patterns

In some ways, the students in this study all followed a common pattern of learning (see [Figure 2](#). Regardless of the main strategy group students were inclined to follow, they tended to do the reading and view the video first (if they did these activities at all), prior to attempting the assignment. Very few went back to the reading and videos once they started the assignments. Likewise, how often students accessed the assignments was more a function of the lesson than the strategy students tended to follow. Lessons 6 and 7 (Beginning and Advanced Modeling) were the lessons where students accessed the assignment most often. Overall, students tended to get lower scores on these two lessons. Lesson 5 (Charts and Graphs) was a lesson where students tended to complete the assignment quickly without needing to go back to the assignment multiple times. Students tended to get near perfect scores on this lesson.

Figure 2. Activity order and times assignment was accessed by each strategy group.

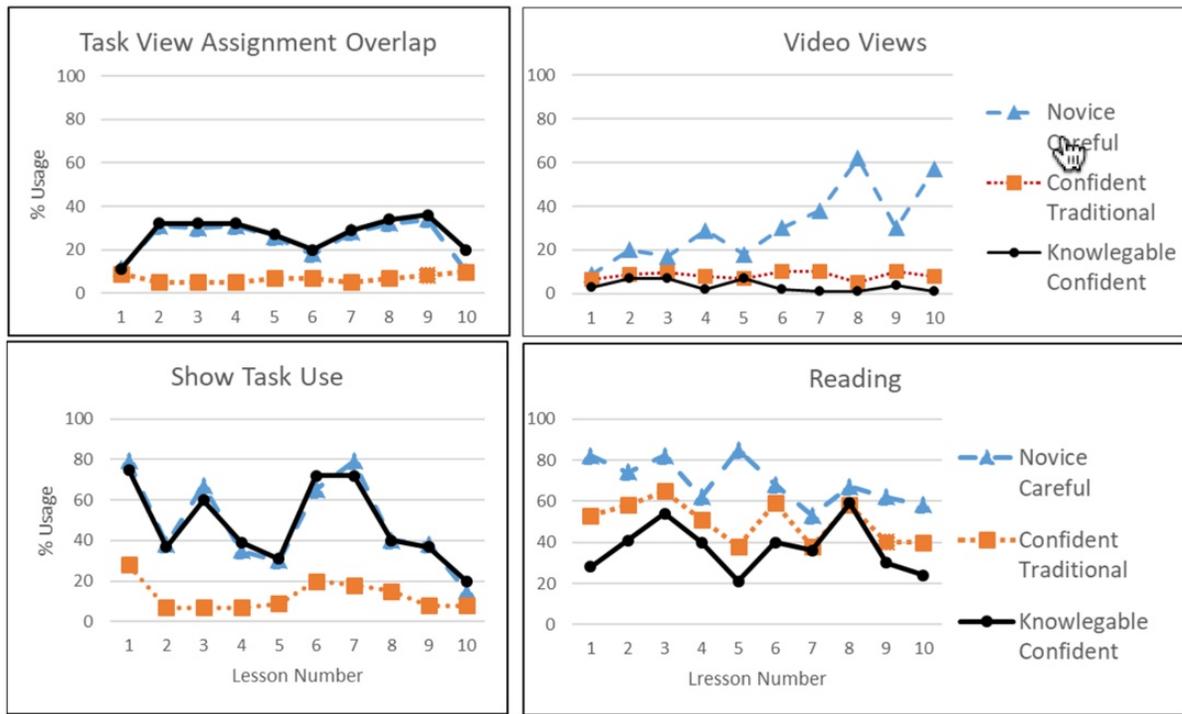


3.3. Unique Strategies and Patterns

In several ways, students in this study followed a unique pattern of learning (see [Figure 3](#). While the Novice Careful and Knowledgeably Confident groups used the task view and overlapped the task view and assignments often, the Confident Traditional strategy groups tended to view the task instructions and use the task guide less often with less overlap. With

regard to reading, the Confident Traditional and Knowledgeably Confident strategy groups tended to read about half, or less than half, of the readings, depending on the lesson (with the exception of Lessons 3 and 8, where these students tended to do more of the reading on average). Those following a Novice Careful strategy tended to do most of the reading in the first part of the course, and then less of the reading in later lessons. Their video views, however, increased in later lessons. This was especially true for Lessons 8 and 10, arguably two of the more difficult lessons, based on instructor comments. Those following the Confident Traditional and Knowledgeably Confident strategies rarely viewed any of the video.

Figure 3. Unique activity patterns based on cluster grouping for each lesson.



4. Discussion and Conclusions

The purpose for conducting this study was to demonstrate the potential use of EDM techniques to better understand student behavior and identify the ways in which the instructional design of a particular course might be improved [15]. On the basis of the average student achievement results alone, one might conclude that the course needs no improvement. This particular course was a fairly easy introductory course. The average scores obtained by the students were quite high; still, some students struggled. Our analysis identified a group of students who may have struggled in specific ways. This suggests that improving the instructional design for some lessons might benefit this group of less-than-adept students, even though, on average, the performance of these students was adequate overall. In addition to this, we were also better able to understand student behavior in general [15] on the basis of the learning strategies they incorporated while taking this course.

On the basis of a longitudinal analysis of student behavior, and somewhat as expected [2,7,8,9], students did tend to follow specific learning strategies as they completed this course. In this course, the majority of students (58%) followed what we called a Knowledgeable Confident strategy. They watched very few videos and read less than half of the instructional text provided in the course. They tended to get right to the assignments and any extra effort, in terms of accessing the assignment and task guides, seemed to be a function of the lesson difficulty. Another common strategy, followed by 21% of students in this course, was the Novice Careful strategy. These students read considerably more of the text, and viewed much more of the videos, especially those provided in the later lessons where they spent less time reading and more time watching. However, students do seem to self-regulate.

About 23% of the time, students switched strategies for a specific lesson. For this course, students tended to switch the most at the beginning and end of the course. In Lesson 1, about 56% of the students deviated from their main strategy.

At this stage of the course, students may be making decisions about how much effort they will need to exert in order to satisfactorily complete the course and achieve their learning goals. They may also be assessing the degree to which instructional resources will help them accomplish their learning goals. After the first lesson, students seem to settle into a specific learning strategy. The lesson topic also seems to be a factor where students self-regulate. For example, in Lesson 5 (Charts and Graphs), students tended to move to a Knowledgeable Confident strategy, likely due to how easy the lesson was or, perhaps, based on the possibility that many students had previous experience with this topic. However, in Lesson 10 of this course, a large majority of students from the Knowledgeable Confident group abandoned their main strategy for completing lessons. One explanation for this might be that Lesson 10 (using Excel's solver) was something these students were unfamiliar with and they needed more assistance in completing the task. Analysis of these data help instructional designers focus their efforts. It is true that a mixed methods approach may be needed in order to fully understand how to improve a course. However, using learning analytics not only helped identify how students went about completing their learning, but also helped us identify where, and in what ways, the instruction might be improved. For example, anecdotal self-report evidence, based on student comments, suggests that the videos were a well-used and well-received element of this course. The students gave positive ratings with regard to the convenience of watching the videos on demand, and to the fact that they could pause and rewind the videos, and even watch them at an increased speed. This perception did not mirror the empirical usage patterns we observed. There may be several ways to interpret this information. One conclusion might be that the materials need to be changed. Certainly, on the basis of an analysis of the learning analytics for this course, the video portion of the course likely needs to be evaluated. Given the high video usage of some students, the video components for more challenging lessons may need to be revised or improved. Still, many students do not seem to utilize video resources (an important finding on its own). However, improving these components may lead to greater use and more efficient learning, especially for those lessons found to be more challenging [1,15]. For example, many students seem to need, or could benefit from, enhancing the video resources in the later lessons. Given that many students struggle with specific lessons, these could be the focus of instructional design efforts. More study is needed for this aspect of the course.

Author Contributions

Conceptualization, R.D., G.A., C.A., N.B. (Nesrin Bakir) and N.B. (Nick Ball); Data curation, G.A. and N.B. (Nesrin Bakir); Formal analysis, R.D., G.A. and C.A.; Investigation, R.D., G.A., C.A., N.B. (Nesrin Bakir) and N.B. (Nick Ball); Methodology, G.A., C.A. and N.B. (Nick Ball); Project administration, R.D.; Resources, N.B. (Nick Ball); Writing—original draft, R.D.; Writing—review & editing, G.A. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

The study was conducted according to the guidelines of the Institutional Review Board (IRB) at Brigham Young University which concluded that the study did not require formal IRB approval as it used existing data with all identifying information removed. No additional data was collected from human subjects.

Informed Consent Statement

Even though this study did not require an IRB research protocol approval, informed consent was obtained from all subjects involved in the study. As student begin this online course, each was asked permission to have their data analytics used for research purposes. All data utilized in this study was used with the permission of each individual student.

Data Availability Statement

Data supporting reported results can be obtained from the principle investigators of this study by request.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Davies, R.; Nyland, R.; Bodily, R.; Chapman, J.; Jones, B.; Young, J. Designing technology-enabled instruction to utilize learning analytics. *TechTrends* **2017**, *61*, 155–161. [[Google Scholar](#)] [[CrossRef](#)]
2. Bannert, M.; Reimann, P.; Sonnenberg, C. Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognit. Learn.* **2014**, *9*, 161–185. [[Google Scholar](#)] [[CrossRef](#)]
3. Davies, R.; Ball, N.; Dean, D. Flipping the classroom and instructional technology integration in a college level information systems spreadsheet course. *Educ. Technol. Res. Dev.* **2013**, *61*, 563–580. [[Google Scholar](#)] [[CrossRef](#)]
4. Entwistle, N.; Ramsden, P. *Understanding Student Learning (Routledge Revivals)*; Routledge: New York, NY, USA, 2015. [[Google Scholar](#)]
5. Davies, R. *Exploring the Meaning and Function of Learner Intent for Students Taking Online University Courses*; VDM Publishing House Ltd.: Saarbrücken, Germany, 2009. [[Google Scholar](#)]
6. Dolmans, D.H.; Loyens, S.M.; Marcq, H.; Gijbels, D. Deep and surface learning in problem-based learning: A review of the literature. *Adv. Health Sci. Educ.* **2016**, *21*, 1087–1112. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)] [[Green Version](#)]
7. Boekaerts, M.; Zeidner, M.; Pintrich, P.R. (Eds.) *Handbook of Self-Regulation*; Elsevier: Amsterdam, The Netherlands, 1999. [[Google Scholar](#)]
8. Carver, C.S.; Scheier, M.F. *On the Self-Regulation of Behavior*; Cambridge University Press: Cambridge, UK, 2001. [[Google Scholar](#)]
9. Bonk, C.J.; Lee, M.M.; Kou, X.; Xu, S.; Sheu, F.R. Understanding the self-directed online learning preferences, goals, achievements, and challenges of MIT OpenCourseWare subscribers. *J. Educ. Technol. Soc.* **2015**, *18*, 349. [[Google Scholar](#)]
10. Oxford, R. *Language Learning Strategies: What Every Teacher Should Know*; Heinle and Heinle Publishers: Boston, MA, USA, 1990. [[Google Scholar](#)]
11. Ben-Eliyahu, A.; Bernacki, M.L. Addressing complexities in self-regulated learning: A focus on contextual factors, contingencies, and dynamic relations. *Metacognition Learn.* **2015**, *10*, 1–13. [[Google Scholar](#)] [[CrossRef](#)] [[Green Version](#)]
12. Nisbet, J.; Shucksmith, J. *Learning Strategies*; Routledge: New York, NY, USA, 2017. [[Google Scholar](#)]
13. Midgley, C. *Goals, Goal Structures, and Patterns of Adaptive Learning*; Routledge: New York, NY, USA, 2014. [[Google Scholar](#)]
14. Kember, D.; Biggs, J.; Leung, D.Y. Examining the multidimensionality of approaches to learning through the development of a revised version of the learning process questionnaire. *Br. J. Educ. Psychol.* **2014**, *74*, 261–279. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
15. Baker, R.S. Challenges for the Future of Educational Data Mining: The Baker Learning Analytics Prizes. *JEDM J. Educ. Data Min.* **2019**, *11*, 1–17. [[Google Scholar](#)] [[CrossRef](#)]
16. Stice, J.; Stice, E.K.; Albrecht, C. Study Choices by Introductory Accounting Students: Those Who Choose to Study by Reading Text Outperform Those Who Choose to Study by Watching Video Lectures. In *Proceedings of the AAA Conference: Western Region, Seattle, WA, USA, 2 June 2016*. [[Google Scholar](#)]
17. The U.S. Higher Education System. The National Science Foundation. 2016. Available online: <https://edtechbooks.org/-YiJV> (accessed on 19 October 2021).
18. Li, C.S.; Irby, B. An overview of online education: Attractiveness, benefits, challenges, concerns and recommendations. *Coll. Stud. J.* **2008**, *42*, 449–458. [[Google Scholar](#)]
19. DiCerbo, K.; Behrens, J. Implications of the digital ocean on current and future assessment. In *Computers and Their Impact on State Assessment: Recent History and Predictions for the Future*; Lissitz, R., Jiao, H., Eds.; Information Age: Charlotte, NC, USA, 2012; pp. 273–306. [[Google Scholar](#)]
20. Skinner, B.F. *The Technology of Teaching*; Appleton-Century Crofts: New York, NY, USA, 1968. [[Google Scholar](#)]
21. Tyler, R.W. *Basic Principles of Curriculum and Instruction*; The University of Chicago Press: Chicago, IL, USA, 1949. [[Google Scholar](#)]
22. Watters, A. *Learning Analytics: Lots of Education Data... Now What?* Paper presented at LAK12. 2012. Available online: <https://edtechbooks.org/-UAS> (accessed on 19 October 2021).

Previous Citation(s)

Davies, R.; Allen, G.; Albrecht, C.; Bakir, N.; Ball, N. (2021). Using Educational Data Mining to Identify and Analyze Student Learning Strategies in an Online Flipped Classroom. *Educ. Sci.* 2021, 11, 668. <https://doi.org/10.3390/educsci11110668>



Randall S. Davies

Brigham Young University

Dr. Randall S. Davies is an associate professor of Instructional Psychology & Technology at Brigham Young University (BYU), where he teaches courses in educational inquiry, measurement, and evaluation. He previously worked at Indiana University South Bend as an Assistant Professor of Educational Research and at IU Bloomington in the Center for Evaluation and Education Policy. He received his Ph.D. in Instructional Psychology & Technology from BYU as well as a B.Ed. in secondary education and a B.Sc. in mathematics & computing science from the University of Alberta.



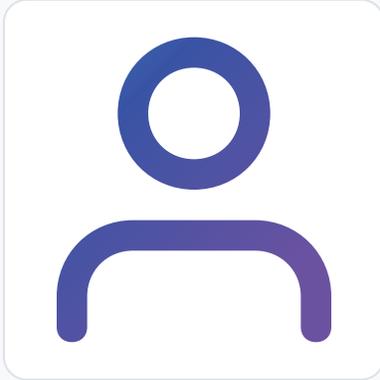
Gove Allen

Brigham Young University



Conan Albrecht

Illinois State University



Nesrin Bakir

Illinois State University



Ball, N.

Utah Valley University



This content is provided to you freely by BYU Open Learning Network.

Access it online or download it at

https://open.byu.edu/light_learning_2022/Data_Mining_in_an_Online_Flipped_Classroom.