Visualizing patterns of student engagement and performance in MOOCs

Carleton Coffrin
National ICT Australia
Victoria Research Laboratory
Melbourne, Victoria, Australia
carleton.coffrin@nicta.com.au

Linda Corrin
Centre for the Study of Higher Education
The University of Melbourne
Melbourne, Victoria, Australia
lcorrin@unimelb.edu.au

Paula de Barba
School of Psychological Sciences
The University of Melbourne
Melbourne, Victoria, Australia
paula.de@unimelb.edu.au

Gregor Kennedy
Centre for the Study of Higher Education
The University of Melbourne
Melbourne, Victoria, Australia
gek@unimelb.edu.au

ABSTRACT

In the last five years, the world has seen a remarkable level of interest in Massive Open Online Courses, or MOOCs. A consistent message from universities participating in MOOC delivery is their eagerness to understand students' online learning processes. This paper reports on an exploratory investigation of students' learning processes in two MOOCs which have different curriculum and assessment designs. When viewed through the lens of common MOOC learning analytics, the high level of initial student interest and, ultimately, the high level of attrition, makes these two courses appear very similar to each other, and to MOOCs in general. With the goal of developing a greater understanding of students' patterns of learning behavior in these courses, we investigated alternative learning analytic approaches and visual representations of the output of these analyses. Using these approaches we were able to meaningfully classify student types and visualize patterns of student engagement which were previously unclear. The findings from this research contribute to the educational community's understanding of students' engagement and performance in MOOCs, and also provide the broader learning analytics community with suggestions of new ways to approach learning analytic data analysis and visualization.

Categories and Subject Descriptors

K.3.1 [**Computer Uses in Education**]: Distance learning—*MOOC*; J.1 [**Administrative Data Processing**]: Education

General Terms

Learning Analytics, Visualizations, Online Learning

Keywords

Learning Analytics, Visualization, Completion Rate, Prior Knowledge, Learner Engagement Patterns, MOOC

1. INTRODUCTION

The last five years have seen a rapid rise in the popularity of Massive Open Online Courses or MOOCs. This rise in popularity has been reflected both in the number of students enrolling in these courses and the number of universities now offering courses in this format. While there are many challenges in providing students with high quality educational experiences at such a large scale, MOOCs have created significant opportunities for researchers and educators who are interested in learning analytics. The sheer numbers of students who participate in MOOCs – often in the thousands – means that educators and educational researchers have access to large data sets of students' online learning interactions which, through the use of learning analytics, can be used to develop a greater understanding of students' online learning experiences, processes and outcomes [19]. The large data sets created by students' interactions in MOOCs provide new opportunities to conduct research on online curriculum and assessment structures, students' patterns of engagement with online material and activities, and the more general areas of student retention and success.

The emerging field of learning analytics is constantly developing new ways to analyze data on students' interactions, engagement and performance. Key to the usefulness of learning analytics is the ability to provide data to educators in ways and formats that can help inform their decision making about educational interventions; and curriculum design and redesign. Work is currently being undertaken by many researchers into methods to visualize data in more effective and informative ways [13, 7, 16]. In the context of MOOCs, current access to visualizations of learners' data through the more recent MOOC platforms – e.g. edX, Coursera – is somewhat limited. However, this is an area of development that is being informed by the analysis of data from research into early MOOC offerings [12, 15]. Beyond line and bar charts that display frequency counts and the timing of students' activity, examples of the kinds of visualizations that have been generated using MOOC data include social net-work diagrams [13], Q-Q plots of communication activities in forums [4], and log plots of student access to learning activities [5].

This paper presents the findings from an exploratory investigation of students' interactions within two MOOCs offered by the University of Melbourne. The two main purposes of the research were to develop more refined learning analytic techniques that can be used with MOOC data, and to design visualizations of the output that is meaningful to end users (instructors, researchers, students and administrators).

The research had the broad aim of seeking to use learning analytics to describe students' interactions in and patterns of engagement and success with the two MOOCs being investigated. But more specifically it sought to determine how students' interactions were seen to be similar or different across two courses, which had quite different and distinctive curriculum and assessment designs. Thus, we sought to develop a greater understanding of students' patterns of learning behavior, how we could use learning analytics of interaction and assessment to meaningfully classify student types, and how visual representations could be used to effectively show patterns of student engagement and success across two MOOCs with different curriculum and assessment structures. The findings from this research will not only contribute to the educational community's understanding of students' engagement and performance in MOOCs, but it will also provide the broader learning analytics community with suggestions of ways to approach learning analytic data analysis and the visualization of the results emerging from these analyses.

The research reported in this paper is intentionally descriptive and exploratory. A structured approach to data analysis was adopted, based on a technique we have used in previous research [11]. With this approach the specific variables that are analyzed and the results of these analyses are determined iteratively at increasingly finer levels of granularity. Given this approach, we have structured the paper by combining methods and results sections to assist the reader. Hence, the paper begins with a brief introduction to both courses and some standard analytics, Section 2 and Section 3. It then revisits these analytics with several iterative refinements, Section 4. It then combines these data refinements with some entirely different analytics in Section 5 and concludes with a discussion of the findings in Section 6.

2. COURSE BACKGROUND

This paper examines the interactions of students in the inaugural sessions of two MOOCs developed at the University of Melbourne, Principles of Macroeconomics and Discrete Optimization. These two courses were chosen because, although they were both developed by the same production team and delivered on the same MOOC platform, the course structures and implementations were entirely different. They varied in subject area, prerequisite knowledge, curriculum design, and assessment design, just to name a few. The remainder of this section provides a brief introduction to each of the courses, to provide context for the rest of the paper.

Principles of Macroeconomics is an introductory course with minimal prerequisites. It consists of eight weeks of material broadly presented in a linear structure. Each week students are asked to watch newly released videos which cover key principles of the course, and respond to quizzes that provide both formative and summative assessment of students' knowledge. A discussion forum and social networking sites are provided for students so that they are able to collaborate while they undertake their learning in the course. Of the eight available quizzes, three contribute to students' final grade (in weeks three, five and eight) while the others are for practice (formative assessment). For the remainder of this paper we will use the term *exam* to refer to quizzes that count toward the students' final grade in the course. Principles of Macroeconomics also contained a peer assessment task where students are required to write a 1,500 word essay on a current topic and to review and grade three other students' essays. The inaugural session of this course attracted the interest of 54,217 students, of which 32,598 started the course, and 1,412 students received a certificate of completion for the course (a 4.33% completion rate). This linear course design with quiz and peer-graded assessments is typical of many recent xMOOCs [14].

Discrete Optimization is a graduate level course which assumed that incoming students have a strong background in computer science. It consists of nine weeks of material presented in an open curriculum structure. That is, all of the assignments and lectures are made available in the first week and students design their own study plan to complete at their own pace. The assessments consist of seven programming assignments: one preliminary, five primary, and one extra-credit. Students are permitted to submit unlimited attempts at each assignment and their grade is based on their marks on the final day of the course. Discrete Optimization included a forum for class discussion, but also encouraged student interaction via a gamification element called the *leader board*, where students can see how their solutions compare to those of their peers. The inaugural session of this course attracted the interest of 37,777 students, with 22,731 starting the course, and 795 receiving a certificate of completion for the course (a 3.50% completion rate).

3. UNDERSTANDING A MOOC

As discussed previously, both institutions and instructors are eager to understand how students are engaging with MOOCs. Especially after seeing completion rates between 3% and 5%, instructors and institutions may question the educational success of MOOC offerings. A number of research articles on MOOCs have investigated the issue of student retention [12, 15] and examining just this issue – student retention – was the starting point for the series of iterative, exploratory analyses presented in this paper.

In the recently developed xMOOC platforms, instructors and administrators have access to a set of high-level analytics related to student engagement, performance, and retention. These analytics typically include summaries of data on areas such as unique views of content, assessment item analysis, and distributions of students' assessment outcomes. While the learning analytics capabilities of existing platforms are continually expanding, the range of analytic reports available to course instructors at the time these two courses were conducted were primarily restricted to histograms of student participation (video views and assignment submissions) and assessment performance (i.e. marks). An example of this data is presented in Figure 1, which in-

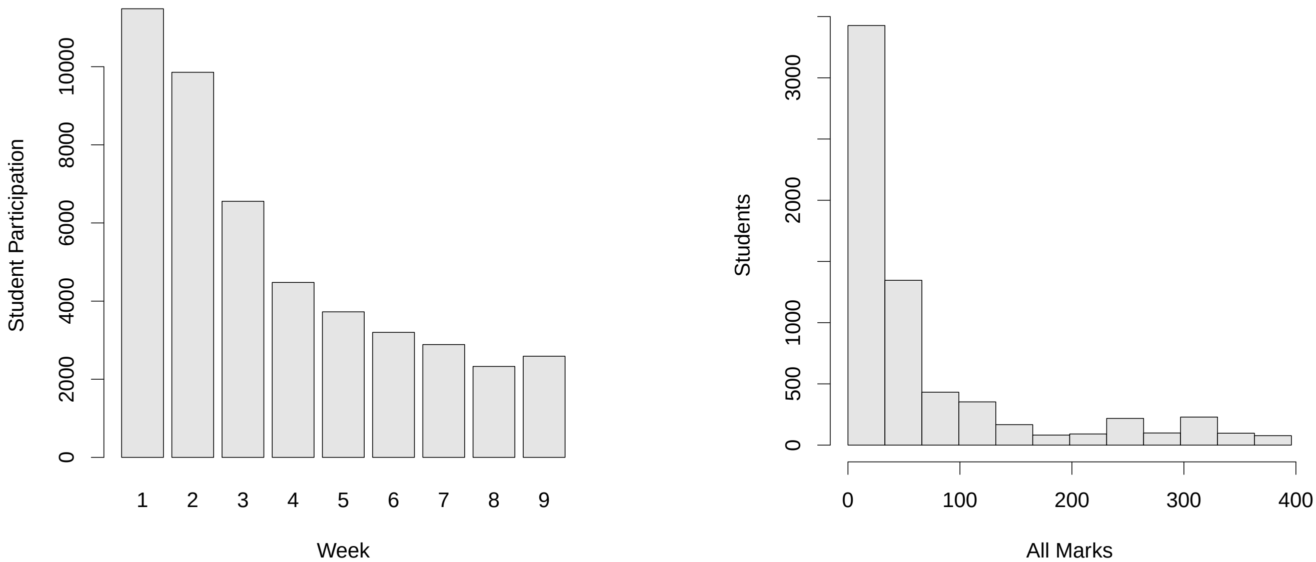


Figure 1: Weekly Student Participation (left) and a Histogram of Student Performance (right) for Discrete Optimization.

cludes the activity of the 22,731 students in Discrete Optimization, broken down by week, and a histogram of students' marks on the assessments.

A preliminary assessment of the analytics presented in Figure 1 indicates that (1) there are many more students viewing the videos than working on the assignments; (2) there is a noticeable and consistent decline in the number of students participating in the course every week; (3) of the subgroup attempting the assignments, the rate of overall success in the course is quite low, with more than half of the students in the first bar of the histogram. These patterns of student behavior shown for Discrete Optimization in Figure 1 are very similar to those for the Principles of Macroeconomics (not shown). One of the most useful insights to be gleaned from these results concerns the number of students who participate in assignments. It can be seen that of the 22,731 students who are active in the course, only 6,633 (i.e. 29.2%) attempted the assignments. Using just the sub-population of students who were active in the course assignments, it is possible to revise the completion rate of Principles of Macroeconomics and Discrete Optimization to 18.1% and 12.0% respectively, which is slightly more appealing.

Conclusions about MOOC completion rates and levels of attrition which are similar to these have been noted by other reports and papers [1, 3, 12, 15]. In many ways, these data visualizations and descriptions mean that Discrete Optimization looks like a typical MOOC – students have a high degree of initial interest and the course shows high attrition. While these types of visualizations are useful in describing the high-level distribution of activity and marks, they are fairly blunt instruments when it comes to making sense of students interactions, engagement and learning. If we were to rely on these types of analytics alone, we would be able to glean very few additional insights about the reasons behind or concomitants to students' success or failure in a particular MOOC. The remainder of this paper outlines a series of alternative analytics that can be harnessed from MOOC data, which provide instructors, researchers and administrators with novel insights into how MOOCs work and how students interact and engage with them.

4. ALTERNATIVE ANALYTICS: AN ITERATIVE APPROACH

There are many possibilities for transitioning from the highly aggregated analytics presented in Section 3 to more detailed analysis of the wealth of student data collected in MOOCs. We have adopted an approach that involved an iterative analysis of analytics data as proposed in [11]. This process involves making a series of repeated passes at the data, with each pass involving further refinement in analysis at a finer level of granularity. On each pass, patterns are observed in the data and those patterns are used to refine the analysis focus. By repeating this process several times, it is possible to cluster either variables into groups and/or student users into subpopulations to gain further insight.

4.1 Revisiting Student Marks

The initial starting point for this type of alternative analysis was completed by revisiting students' marks. In Figure 1 the distribution of student marks was presented as a histogram with 12 bars. A shortcoming of using histograms to present data is that the information they provide is very sensitive to the number of bars used. For example, the mode of a distribution can easily be obfuscated if the number of bars is too small. An alternative way to visualize distribution data is a cumulative distribution plot, a key advantage of which being that it does not require a fixed number of bars.

Figure 2 shows students' marks for Principles of Macroeconomics and Discrete Optimization as cumulative distributions.[1] Marks from all assessments were used in the creation of these distributions, not just those assessments that

[1]Figures 2, 3, and 4 presented throughout this section were produced in **R** [18] using the built-in packages.

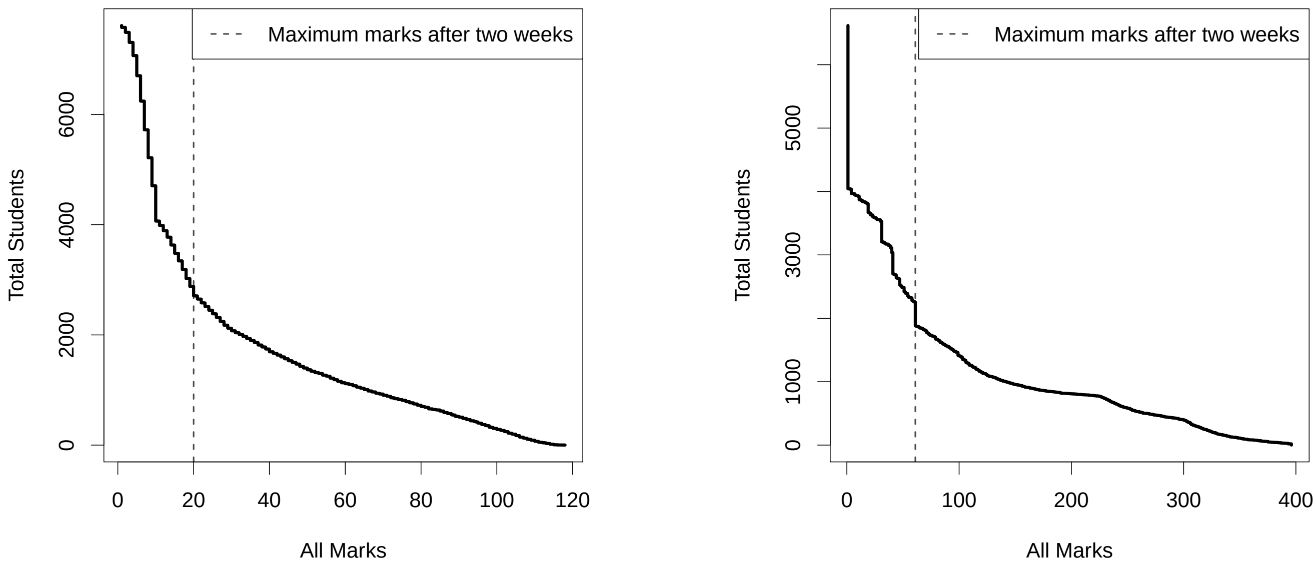


Figure 2: Cumulative Distributions of Student Performance for Principles of Macroeconomics (left) and Discrete Optimization (right).

counted towards the final grade (i.e. the practice quizzes in Principles of Macroeconomics are included). Considering all of the marks allows us to more consistently compare across the courses. With the very granular level of detail provided by the cumulative distribution, we can now see some clear similarities between the courses as evidenced by the similar shapes of the distribution curve. It is also worth noting that for both courses there appears to be an "elbow" in the distribution plot in the region of low student marks. This elbow reflects the point where the significant and sheer drop in the distribution starts to even out into a smoother slope.

Using our knowledge of these courses, we were able to observe that the start of the smooth mark distribution coincides with the number of marks that can be earned in the first two weeks of these courses, as indicated by the dashed line in Figure 2. This suggests that in the first two weeks many students experiment with the assignments but only a dedicated subgroup (less than 50% in these courses) continues to work on the assignments for the entirety of the course. This new insight into these courses inspires a novel hypothesis about students' performance in MOOCs.

HYPOTHESIS 1. *Students' marks in the first two weeks are a good predictor of the final grade in the course.*

In order to explore this hypothesis, we focussed on the marks obtained in the first two weeks in both courses (see Figure 3). To test the hypothesis, we ran a linear regression model for each course considering the association between students' marks at the two week point of the course and their final grade, as summarized in Table 1. These regression analyses were significant for both courses ($p < 0.001$), and there was a higher association for Discrete Optimization ($R^2 = 52.7\%$) than for Principles of Macroeconomics ($R^2 = 20.6\%$). However, we noticed that the Principles of Macroeconomics regression greatly benefits from including the first exam ($R^2 = 51.5\%$), which occurs in the third week of the course. For a consistent comparison between the courses, we elect to focus on the students' marks at the two week point in both courses.

Table 1: Marks Regression Analysis Summary

Course	Marks	R^2	$F(d, de)$	$p <$
Dis. Opt.	2 weeks	52.7%	(1,6612) = 7362.21	0.001
Macro.	2 weeks	20.6%	(1,7614) = 1978.71	0.001
Macro.	3 weeks	51.5%	(1,7614) = 8081.11	0.001

An implication of these findings is that instead of having to wait until the end of the course to run post-hoc analyses of student performance, the group of students most likely to succeed can be identified early in the course. The great advantage of this to instructors is that they can tailor course material to different types of students. They could, for example, choose to give their attention to students who appear to have the required skills to succeed and are actively trying to complete the assignments, leading to customization and interventions while the course is in session.

Having established an association between performance in the first two weeks and overall success in the course, we can now identify a particular group of students who had performed well in the first two assessments. We call this group the *qualified* students, and define them as the students who obtained marks above the 60^{th} percentile in the first two weeks of the course, indicated by a dashed line in Figure 3. This group of students are regarded as *qualified* because their relatively high scores on the first two assessments indicates they have substantial prior knowledge [6, 9, 17] in the discipline area, and moreover, they have invested the time required to complete the assignments. For a consistent comparison across both courses we chose the 60^{th} percentile somewhat arbitrarily to define this group. If applied to other courses, this value could be determined on a case-by-case basis after inspecting the two-week mark-distribution plot (i.e. Figure 3).

The qualified subgroup provides another lens for understanding course completion rates. If we consider the qualified students as reflecting a subgroup who had the necessary prerequisites for completing the course – much like prerequisites for taking more traditional university courses – then

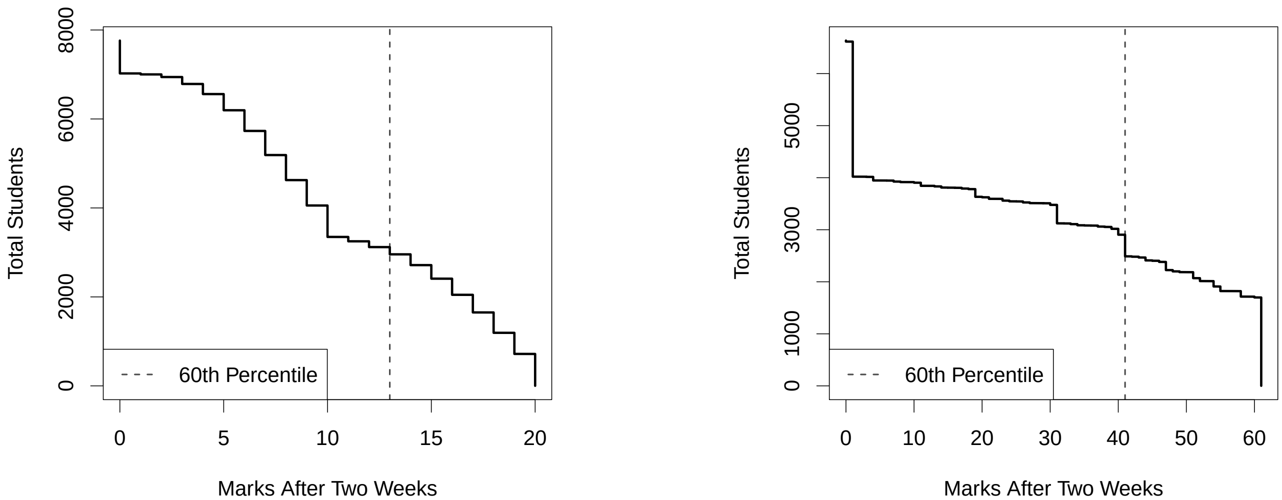


Figure 3: Cumulative Distributions of Student Performance for Principles of Macroeconomics (left) and Discrete Optimization (right).

we can revise the completion rate of Principles of Macroeconomics and Discrete Optimization to 42.1% and 27.4% respectively, which is **much** more appealing than the 5% to 3% we started with. Table 2 summarizes the completion rate calculations discussed thus far. Along the lines of [11], the identification of this *qualified* subgroup enables us to revisit **all** of the course data in a more detailed manner. The remainder of this paper focuses on how this subgroup can be used to enhance our understanding of learning analytics data with a view to helping instructors better understand the engagement behavior of students in MOOCs.

Table 2: Completion Rate Calculations

Course	All	Active	Qualified
Macroeconomics	4.33%	18.1%	42.1%
Discrete Optimization	3.50%	12.0%	27.4%

4.2 Revisiting Weekly Participation

In this section we revisit the weekly participation plot (Figure 1) using the insights from Section 4.1, in terms of both assignment activity and qualified students. Initially total weekly participation was divided into three mutually exclusive subgroups:

1. *Auditors*, students who watched videos in a particular week, but did not participate in any assessments.

2. *Active*, students who participated in an assessment in a particular week.

3. *Qualified*, students who watched a video or participated in an assessment and met the two assessment qualification criteria from Section 4.1.

This analysis is inspired by [12], but the qualified student group provides a new level of detail in the analysis. Figure 4 presents the students' weekly participation in both courses and the percentage in each bar indicates the relative percentage for that week. Now that the data have been divided into groups, a more detailed story of student engagement is revealed in both courses.

For example, some of the novel insights revealed by Figure 4 include: (1) The relative proportions of auditor, active, and qualified students are remarkably similar across both courses, despite the substantial differences in the designs of the courses and their target students; (2) Of the students who are still engaged in the course, the relative proportion of qualified students is broadly maintained over time, unlike the population of active students, which decreases steadily. This is consistent with the "transformative shift" findings of [20]; (3) The reduced number of active students in weeks five and seven of Macroeconomics suggests that some students only complete the exams and skip the quizzes; and (4) The relative proportion of students who are discontinuing across both courses came from the Active group.

In summary, despite the significant differences in the course structures and assessment designs, these courses exhibit remarkable similarities in relation to student activity, in both performance (total marks) and relative weekly participation. Given these similarities, an obvious question that emerges is: Do students take advantage of the open course structure of Discrete Optimization, or do they naturally conform to the linear delivery structure of Principles of Macroeconomics?

5. UNDERSTANDING TEMPORAL ENGAGEMENT

In the previous sections we have shown how the qualified subgroup can be used to inform completion rate calculations (Section 4.1) and weekly participation analytics (Section 4.2). In both cases there were striking similarities between Principles of Macroeconomics and Discrete Optimization. In this section we investigate students' temporal engagement with the course material. Specifically, we consider the degree to which students in each course conform to

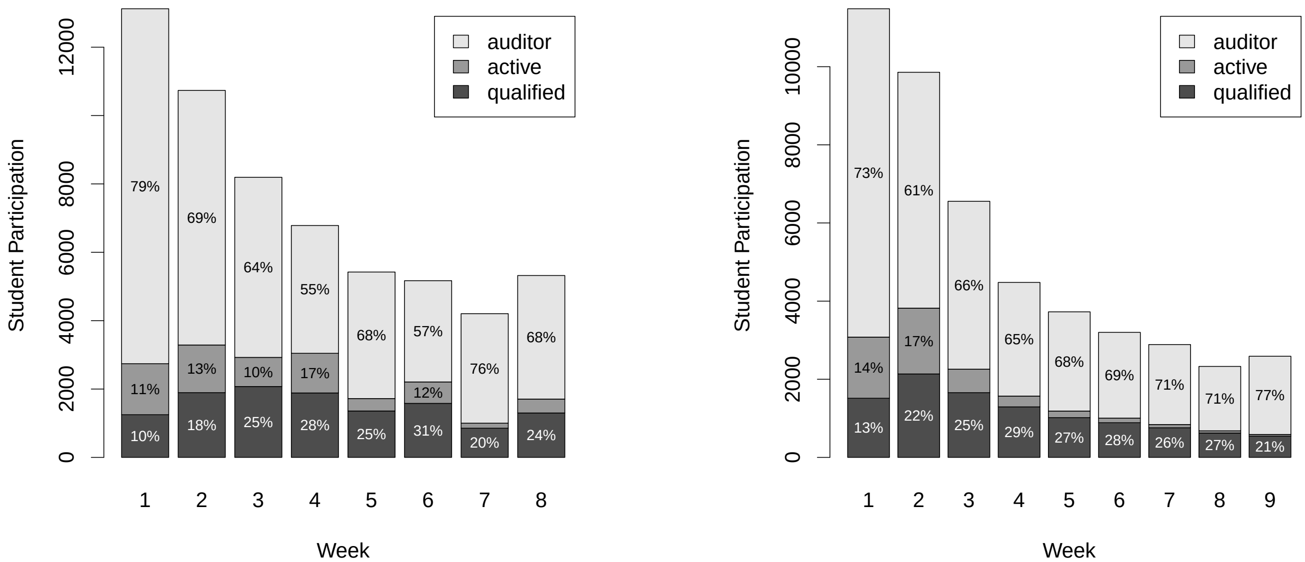


Figure 4: Weekly Student Participation by Student Subgroups for Principles of Macroeconomics (left) and Discrete Optimization (right).

a more traditional linear delivery structure and whether students in Discrete Optimization particularly take advantage of the open course structure. To investigate this question, we introduce another visualization, called a state transition diagram. We will show how the state transition diagram combined with an analysis of student subgroups can clearly illustrate the differences in students' temporal engagement in the two courses.

5.1 State Transition Diagrams

Contemporary MOOC platforms collect detailed data about students' online interactions. It is possible to study students' interactions and engagement at both a macro-level (e.g. how many students completed an assignment) and a micro-level (e.g. a click-stream, tracking every interaction a student has with the learning platform). The detail provided by data contained at the micro-level can be seductive, but it can also be incredibly difficult to make sense of such data. In this section, we introduce state transition diagrams as a way of taking detailed time-stamped interaction data and reducing it to a simple visualization which retains useful temporal information.

Originating as finite state automata in the computer science discipline, the purpose of state transition diagrams is to represent how a system moves from one state to another state over a sequence of events. These diagrams are typically visualized as a graph where nodes represent the states and the lines connecting nodes reflect probabilistic or weighted transitions between states. Previous educational researchers have used state transition diagrams with learning analytics. For example, [10] used state transition diagrams, among other techniques, to analyze logs of students' interactions with an online drag-and-drop learning activity. In the context of MOOCs, [12] used a form of state transition diagram to show student movement between categories of engagement over assignment periods.

Figure 5 provides a legend to understand the state transition diagrams presented in this paper. The legend shows

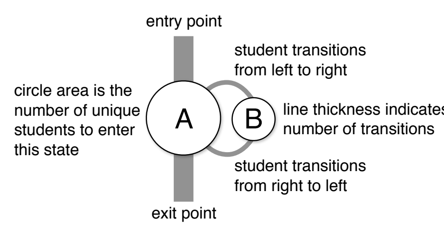


Figure 5: State Transition Diagram Legend.

two possible states, A and B. The figure shows that students enter the State A from the top, make transitions between the states, and then exit the State A from the bottom. The arcs above the states indicate transitions from left to right (i.e. from A to B) while the arcs below the states indicate transitions from right to left (i.e. from B to A). Additionally, line thickness represents the relative number of transition made, while the size of the circle for any given state indicates the number of unique students that entered the state. Each state's name and size is presented as text inside the circle to aid in the explanation of the visualization in this paper, but this information is not strictly necessary in practice. For consistency across both courses, we present the states linearly, based on the recommended path through the course material. Hence, students following a linear curriculum structure will always transition from left to right. Transitions from right to left (i.e. below the states) indicate students exploring the curriculum in a different way. The state transition diagrams presented throughout this section (i.e. Figures 6 and 7) were produced in HTML5 and D3 [2] using customized data processing and layout instructions.

This temporal visualization enables us to determine the impact of an open course structure, such as that used in Discrete Optimization, on the degree to which students switch or "jump around" the course. Furthermore, the visualization is fairly general, so it can be applied to various forms of temporal data. In the next two sections we will consider the student state transitions with regards to both video views and assignment submissions.

5.2 Video Views Transitions

It is typical in MOOCs to break lecture topics into short 5 to 15 minute segments. If each of these micro-lectures were to represent a state in a state transition diagram, too much detail would be provided, making it very difficult to interpret. Instead, if videos are able to be grouped by conceptual theme or topic we can observe the major state transitions more easily.[2] In Principles of Macroeconomics the video lectures are grouped into six topic areas, while in Discrete Optimization they are grouped into four topic areas.

Figure 6 presents the video view state transitions diagrams from both Principles of Macroeconomics and Discrete Optimization segmented by the qualified and non-qualified student groups. For clarity, states and transitions that were used by less than 1% of students are omitted from the diagrams. Some of the key observations are: (1) Comparing across the qualified and non-qualified groups, we can see that the quantity of non-qualified students is greater than the qualified group. Hence, we have a very different insight into video viewing behavior when the qualified group is analyzed in isolation; (2) The transition behavior of the non-qualified group is remarkably similar across the two courses; and (3) When the qualified groups are compared across the two courses, it appears that these students switch between video topics in both courses more than non-qualified students, however, the switching behavior is more pronounced in Discrete Optimization.

In the next section we repeat this same analysis using students' assignment submission as the key variable.

5.3 Assignment Submissions Transitions

Comparing the assignment transitions between these courses is particularly interesting because the rolling deadlines in Principles of Macroeconomics strongly constrain students to the structure of a linear curriculum, while the open curriculum structure of Discrete Optimization allows students to follow their own learning pathways, in their own time. Figure 7 presents the assignment submission state transitions diagrams from both Principles of Macroeconomics and Discrete Optimization, again segmented by the qualified and non-qualified student groups. As before, the diagram excludes states and transitions below 1% of the students. Some of the key observations are: (1) When qualified and non-qualified groups are compared, again we can see that the number of non-qualified students is far greater than qualified students. Once again, this allows for a very different insight into assignment submission behavior of students when the qualified group is analyzed in isolation; (2) The behavior of the non-qualified group of students is similar in the two courses. However, it is clear that a subgroup of non-qualified students in Macroeconomics prefer to skip the practice quizzes and only take the exams. This is evident in the state transition diagram 7(a) where a population of approximately 100 students jump from Quiz 1, to Exam 1, then to Exam 2, and finally to Exam 3. In contrast, no students from the non-qualified group complete the assignments in Discrete Optimization; (3) When comparing the two qualified groups across the courses, it is clear that students took great advantage of the open course structure in Discrete Optimization as evident by the number of forward and backward transitions between nodes. It seems that many students repeatedly revised their assessments over the 9 week course period.

Overall, both of these state transition analyses suggest that students took advantage of the open course structure provided by Discrete Optimization. The differences between the state transition diagrams of non-qualified and qualified groups of students, across video viewing and assessment submission, for both courses, suggest the valuable insights that can be gained from isolating specific subgroups of students for analyses and visualizations using learning analytic data.

6. DISCUSSION

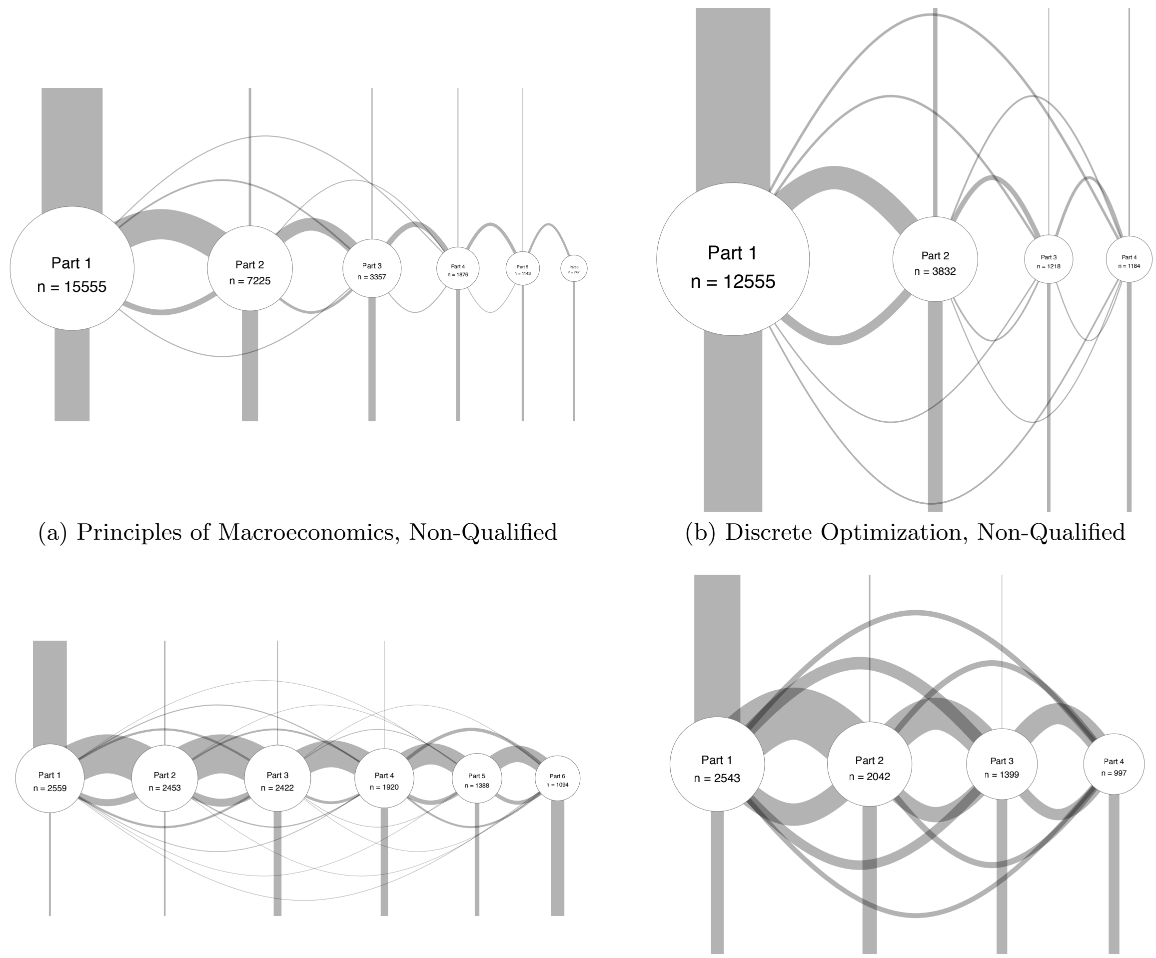
The aim of this exploratory investigation was to develop learning analytics techniques that could be applied to large data sets that emerge from students' participation in MOOCs. Additionally, we sought to develop a greater understanding of patterns of students' learning behavior across two distinct MOOCs through the use of different visual representation techniques. The starting point for this investigation was an acknowledgement that while mainstream MOOC providers are continually developing their learning analytics capacity, the current learning analytics output from these platforms is limited.

The findings from our investigation show the power of different visualizations of learning analytics output. When the two courses were compared in terms of students' interaction patterns it was somewhat surprising to see such similar patterns given the differences in each course's curriculum and assessment design (see Figures 2 and 4). For example, the cumulative distribution for students' performance for Principles of Macroeconomics and Discrete Optimization were very similar. Moreover, when visualizations of patterns of engagement are differentiated by types of user (auditor, active, qualified), very similar patterns of activity across the two courses emerge, despite one having a linear curriculum design and the other employing an open curriculum design. While it may be assumed that different curriculum structures engender different patterns of engagement, clearly for some variables and at particular levels of analysis, this is not the case.

The advantage of visualizing learning analytics output can also be seen in the use of state transition diagrams. This analysis and their visualization clearly showed how different types of users (qualified, non-qualified) show different patterns of transition between key MOOC resources (videos) and assessment activities. Our analyses provide new insights into how students engage with MOOCs and, moreover, suggest how different patterns of engagement impact on student performance. These types of analyses further our understanding of how different patterns of student engagement in MOOCs – and potentially other online learning environments – may lead to success. This has clear implications for student retention and attrition in MOOCs, which is covered in more detail below.

The use of state transition diagrams also sheds light on how differences in course curriculum and assessment design may impact on patterns of student engagement. This is most obviously seen by comparing the patterns of engagement

[2] We considered both a week-based grouping and a topic-based grouping. The latter was more informative as key conceptual themes were covered over multiple weeks and any one week could cover at least two conceptual themes.



(a) Principles of Macroeconomics, Non-Qualified

(b) Discrete Optimization, Non-Qualified

(c) Principles of Macroeconomics, Qualified

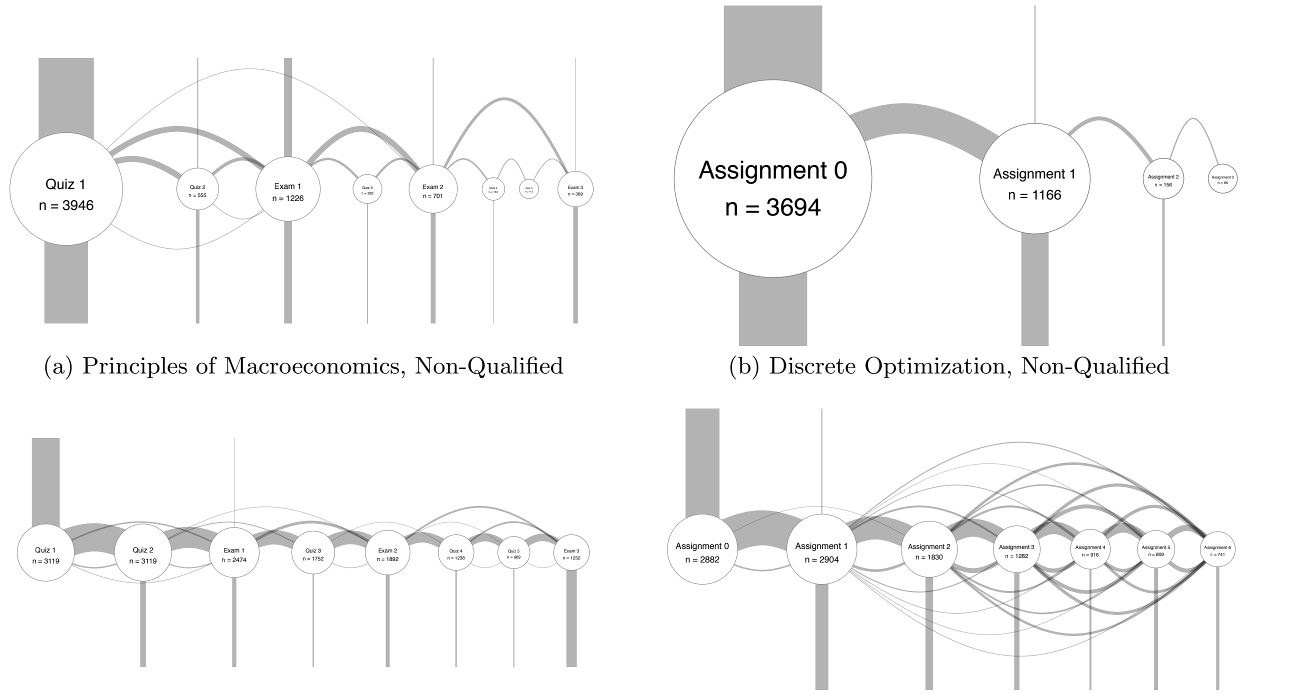
(d) Discrete Optimization, Qualified

Figure 6: Student Video Viewing Transitions Broken Down by Subgroups, Non-Qualified (top) and Qualified (bottom) for Principles of Macroeconomics (left) and Discrete Optimization (right).

of qualified students in Principles of Macroeconomics and Discrete Optimization. Students in Discrete Optimization clearly switched between assignments to a greater extent than students in Macroeconomics and also switched between videos provided in the course. Importantly, it seems that students in Discrete Optimization were more likely to go back to material they had already covered or viewed. While this is not surprising given the curriculum structure, it does indicate that by designing different curricula, and learning and assessment tasks in different ways, instructors have a strong influence on students' learning activities and behaviors [8]. These types of analyses and visualizations can also be very useful in informing curriculum redesign for future offerings of the course. If students are repeatedly viewing certain videos this could indicate that the concept or principle being explained in the video may require further clarification or supporting resources. The visualization of students' switching between concepts that are separated within a curriculum structure may also indicate where changes to the flow of concepts within the curriculum are needed.

A clear finding from this investigation was that the student activity and success in the first couple of weeks of the course was significantly associated with students' outcomes at the end of the course. The important role of prior knowledge in student learning has been thoroughly investigated by educational researchers and previous research has demonstrated that prior knowledge is linked to student success [6, 9]. Although we did not explicitly measure prior knowledge in this investigation, students' performance on their first two (Discrete Optimization) or three (Macroeconomics) assessments was able to be used as a way to determine whether students had the necessary prior knowledge to accommodate the new, related information they covered in the course into their existing knowledge framework [17].

In practical terms, these findings highlight the potential value of providing students with informal diagnostic tests at the start of a MOOC. Such tests could be embedded within curriculum and could be used as early signals to both instructors, and students themselves, about how well students are suited to the course material. This type of diagnostic test could be used to support automated adaptive learning techniques as well as more traditional forms of remediation and feedback. For example, if instructors and teaching assistants were able to view the patterns of engagement and the early



(a) Principles of Macroeconomics, Non-Qualified

(b) Discrete Optimization, Non-Qualified

(c) Principles of Macroeconomics, Qualified

(d) Discrete Optimization, Qualified

Figure 7: Student Assignment Submissions Transitions Broken Down by Subgroups, Non-Qualified (top) and Qualified (bottom) for Principles of Macroeconomics (left) and Discrete Optimization (right).

student performance shown in Figure 4, this could be used as the basis for providing information and support to students. Strategies for real-time remediation in MOOCs could include highlighting common areas of misunderstanding in weekly student communications, launching new discussion threads to assist particular groups of students or targeting problem areas, and providing additional resources on topics requiring further explanation.

Identifying different student types based on their patterns of engagement also opens up the possibility of instructors tailoring specific feedback to particular sub-populations of students. For example, in the context of the current study, instructors may decide to tailor their feedback to those in the student audience who are active and qualified, rather than those who are auditors. If instructors know that a significant cohort of students are active in the course – they are watching videos and attempting assessments – but they are also likely to disengage from the course before its conclusion, they could design strategies to support these particular students.

There has been considerable commentary in the educational community about the rate of student attrition in MOOCs. Much of the consternation comes from comparing students' participation in traditional, often campus-based courses for which they often pay considerable fees, with students' participation in free MOOCs that are delivered online. In many respects these are not fair comparisons as it is likely that students come to each type of course with different levels of motivation, expectations and degrees of commitment. This notwithstanding, the findings from this paper provide an alternative way to calculate metrics of student retention in MOOCs. Rather than calculating these proportional completion metrics on the basis of the number of students who expressed an interest in the course, or the number of students who logged on, it may be more informative to calculate retention metrics on the basis of those students who are deemed to have sufficient aptitude or prior knowledge to complete the course, as based on early diagnostic assessment. In the case of the latter, completion rates would be based on the number of students who have shown a baseline competency in the course material. While this may not be appropriate for some MOOCs that are of general interest, it might be an alternative and useful metric for more technical or professional courses. It is important to note that we are not advocating any type of restriction on open participation in MOOCs, rather that metrics of retention and attrition also be calculated on the basis of those students who are deemed to have sufficient prior knowledge to complete the course.

7. CONCLUSIONS

The analyses and visualizations of learning analytics data presented in this paper go some way to providing greater insights into student activity in MOOCs. However, there is still a great deal of research that can be done in this area. For example, the segmentation of students into auditors, active and qualified groups could be investigated further, across a range of MOOC contexts and disciplines. Also a more detailed analysis could be undertaken to understand the transitions students make from being active participants to auditors. The analysis of the state transition diagrams in terms of performance in the course could also add to our knowledge of the patterns of behaviors in MOOCs that can contribute to success. Further research could also examine the concept of student motivation and the relationship this has to patterns of engagement and performance in MOOCs.

As noted at the start of this paper, in the last few years there has been considerable interest and hype about MOOCs generally, and the potential of collecting "big data" on learners using these platforms. However, current MOOC platforms are limited in their ability to provide data feeds and visualizations that can be easily used to assist instructors, administrators, designers and students' in their decision making. In this paper we have demonstrated that with a relatively small amount of extra thought and analysis we are able to generate sophisticated visualizations and diagnostics associated with learning analytics data. While the creation of these visualizations is currently a post-hoc process applied to data extracted from one particular MOOC platform, the approach may be easily generalized to a live process on many other platforms. In fact, the true value of learning analytics to MOOCs – and to online learning more generally – will only be realized when MOOC and other platforms for virtual learning embed the kinds of analyses and visualizations presented in this paper for end users to routinely and easily access and use.

8. ACKNOWLEDGMENTS

The authors would like to thank Victor Pillac and Colin Tatham for thier insight in the preparation of this paper. This work has been conducted as part of a collaboration supported through the University of Melbourne's Learning Analytics Research Group, which fosters interdisciplinary research in the area of learning analytics. This work was conducted in part at NICTA and is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

9. REFERENCES

[1] G. Balakrishnan and D. Coetzee. Predicting student retention in massive open online courses using hidden markov models. http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-109.html, 2013.

[2] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.

[3] L. B. Breslow, D. E. Pritchard, J. De Boer, G. S. Stump, A. D. Ho, and D. Seaton. Studying learning in the worldwide classroom: Research into edx's first mooc. *Research & Practice in Assessment*, 8:13–25, 2013.

[4] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong. Learning about social learning in moocs: From statistical analysis to generative model. *CoRR*, abs/1312.2159, 2013.

[5] D. Clow. Moocs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 185–189, New York, NY, USA, 2013. ACM.

[6] F. Dochy, M. Segers, and M. M. Buehl. The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of educational research*, 69(2):145–186, 1999.

[7] E. Duval. Attention please!: learning analytics for visualization and recommendation. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, LAK '11, pages 9–17, New York, NY, USA, 2011. ACM.

[8] R. Ellis and P. Goodyear. *Students' experiences of e-learning in higher education: the ecology of sustainable innovation*. Routledge, 2009.

[9] G. D. Haertel, H. J. Walberg, and T. Weinstein. Psychological models of educational performance: A theoretical synthesis of constructs. *Review of Educational Research*, 53(1):75–91, 1983.

[10] T. S. Judd and G. E. Kennedy. More sense from audit trails: Exploratory sequential data analysis. In *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference*, pages 476–484, 2004.

[11] G. E. Kennedy and T. S. Judd. Making sense of audit trail data. *Australian Journal of Educational Technology*, 20(1):18–32, 2004.

[12] R. F. Kizilcec, C. Piech, and E. Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 170–179, New York, NY, USA, 2013. ACM.

[13] R. Kop, H. Fournier, and J. Mak. A pedagogy of abundance or a pedagogy to support human beings? participant support on massive open online courses. *The International Review of Research in Open and Distance Learning*, 12(7), 2011.

[14] A. Littlejohn. Understanding massive open online courses. http://cemca.org.in/ckfinder/userfiles/files/EdTech%20Notes%202_Littlejohn_final_1June2013.pdf, 2013.

[15] MOOCs@Edinburgh Group. Moocs @ edinburgh 2013: Report #1. http://hdl.handle.net/1842/6683, May 2013.

[16] M. Olmos and L. Corrin. Learning analytics: A case study of the process of design of visualizations. *Journal of Asynchronous Learning Networks*, 16(3):39–49, June 2012.

[17] P. R. Pintrich, R. W. Marx, and R. A. Boyle. Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational research*, 63(2):167–199, 1993.

[18] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[19] G. Siemens and P. Long. Penetrating the fog: Analytics in learning and education. *Educause Review*, 46(5):30–32, 2011.

[20] M. Waite, J. Mackness, G. Roberts, and E. Lovegrove. Liminal participants and skilled orienteers: Learner participation in a mooc for new lecturers. *MERLOT Journal of Online Learning and Teaching*, 9(2):200–215, 2013.