Analysing Student Engagement in an Online Course in the Context of Hybrid Learning Environment. An Empirical Study

Michael Wahiu, Fahima Djelil, Laurent Brisson, Jean-Marie Gilliot, Antoine Beugnard

Analysing Student Engagement in an Online Course in the Context of Hybrid Learning Environment. An Empirical Study

Michael Wahiu[1], Fahima Djelil[2][a], Laurent Brisson[2][b], Jean-Marie Gilliot[2][c], and Antoine Beugnard [2][d]

[1]*Computer Science Department, Dedan Kimathi University of Technology, Dedan Kimathi, Kenya*

[2]*IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France*

michael.kamau@dkut.ac.ke, {fahima.djelil,laurent.brisson,jm.gilliot,antoine.beugnard}@imt-atlantique.fr

Keywords: Online Course, Hybrid Learning Environment, Student Engagement, Learning Analytics, Course Design

Abstract: This paper aims to understand student learning engagement in an online course. We describe an empirical study we conducted to investigate learner profiles when interacting with learning content. This study is based on data records about student online navigation and took place in the context of a hybrid environment. The obtained results showed that students mostly select assessment activities and visit the online course content without engaging deeply in the learning activities. This leads us to conclude on the role of assessment to motivate and engage students and on the importance of thinking out the design of the hybrid course. Finally, future work is motivated to study how to provide effective interactions with course content and how this can impact learning engagement and course design.

1 INTRODUCTION

Presently, in higher education, online learning continues to have a rapid expansion, since it removes the space and time restrictions for learning facilitation. As a matter of fact, formalized online learning processes are not time-bound, and supports a blend of both the class-paced and self-paced learning interactions (Kumar et al., 2011; Means et al., 2014). In addition, advances in Learning Analytics have become resourceful for measuring and reporting of data about learner records (Siemens et al., 2011). In particular, Learning Analytics offers many opportunities for online learning analysis (Xiong and Suen, 2018), and go beyond statistical description of learning data (Chatti et al., 2012). Reporting data about students learning activities and describing how students navigate in their courses, help to obtain direct insights from online activities (Lockyer et al., 2013). This helps to acquire objective measurements of learning such as determining whether students progress well and acquire the desired learning outcomes, or whether they are at risk of failure or need adaptive scaffolds.

[a] https://orcid.org/0000-0001-8449-2062
[b] https://orcid.org/0000-0002-5309-2688
[c] https://orcid.org/0000-0003-4569-3552
[d] https://orcid.org/0000-0002-3096-237X

As variations of blended learning evolve, educators need developing better understanding of how effective interaction with course content impacts engagement and learning (Murray et al., 2013). In the context of hybrid learning, understanding how students navigate in an online course when it takes place in a hybrid context implies to examine online learning navigation of a large number of students. This can lead to provide educators synthetic observations of how their students interact with online resources.

In this paper, our objective is to understand students learning engagement in an online course, based on their data records. Our main research questions are the following:

1. What are the indicators of student engagement that can be measured from online learning activities?
2. What are the student profiles that can be obtained from these indicators, and that can lead to some insights on student engagement ?

To this end, based on the literature, we first defined some indicators that are relevant to examine student online navigation, and that may impact learning engagement. We applied Principal Component Analysis (PCA) and an Agglomerative Hierarchical Clustering (AHC) to determine typical learners' behaviours on data learning records.

The present paper is organised as follows. Section 2 gives a description of some online learning engagement indicators. Section 3 describes the methodology employed, including details of the context, as well as data collection and analysis. Results are summarised and discussed in section 4. Finally, conclusions and future research directions are drawn in section 5.

2 LEARNING ENGAGEMENT INDICATORS

Within Learning Management Systems (LMS), such as Moodle (Rice, 2015), online learning is characterized by different kinds of educational interactions (Horton, 2011), including interactions with learning content and interactions between students (Bates et al., 2018).

From various education researches, student learning engagement is related to both a state of the mind and behavioral actions (Mosher and MacGowan, 1985; Skinner et al., 1990; Skinner and Belmont, 1993). Agreeably to a common meaning, authors like (Carini et al., 2006; Robinson and Hullinger, 2008; Kuh, 2009; Lewis et al., 2011; Gunuc and Kuzu, 2015) posit that, student engagement should encompass multidimensional factors, parts of cognitive and behavioral reactions towards educational activities.

Based on that, we explore some possible indicators of student learning engagement relaying on the literature:

- Kind of resources: the kind and the diversity of resources the student choose in a given course, that can have a direct impact on his motivation to learning (Coates, 2006; Lei, 2010; Means et al., 2014). Depending on the kind of resource the learning activity may consist of reading, watching, listening, exploring, solving problems, answering questions or discussing with peers.
- Kind of tasks: This may indicate the activeness or the passiveness of the student participation (Kuh, 2009; Ma et al., 2015; Rieber, 2017), that can influence knowledge replication and recall (Lea et al., 2003; Coffrin et al., 2014; Xing et al., 2015). For example, post-tasks and submit-tasks reveal a more active participation than view-tasks. This also includes the collaborative and individual nature of the tasks.
- Interaction intensity: This may indicate how interested and determined is the student by the learning activities (You, 2016; Xiong and Suen, 2018). This is in relation to the number of achieved tasks (Henrie et al., 2015), or even the number of interaction attempts (You, 2016; Xiong and Suen, 2018). For example, the number of clicks to browse a page, play a video or post in a forum.

3 METHOD

Our work aims to examine how students engage in an online learning course. We first consider the student interactions within the same time baseline, then we vectorize the student interactions with the learning resources. To determine typical learners' behaviors, and the set of event indicators underlying similarities and differences between learners, we conducted a Principal Component Analysis (PCA) and an Agglomerative Hierarchical Clustering (AHC). This method has proven to be effective in a similar study (Djelil et al., 2018).

3.1 Experiment context

The study took place in a hybrid course using the Moodle platform (May-August 2019), involving 133 undergraduate students at Dedan Kemathi University of Technology in Kenya. The course was about Computer Information Technology (CIT), which aims students to develop a broad understanding in all areas of Information Technology, including Operating Systems and Database Management Systems. The online course complements in-classroom activities. It is organized as a sequence of learning activities provided in the Moodle platform, comprising 2 forums, 9 quizzes (each quiz is about a classroom lesson), 4 wikis, 1 assignment allowing students to submit work to their teacher. It offers additional resources including e-books, lecture files, HTML pages, URL and workshop resources. The hybrid course also includes other offline tests and a classroom final exam.

3.2 Data Collection

Data records were stored about learning activities characterized by student interactions with online learning resources. Data about a learning activity is characterized by a sequence of events produced by the LMS, each time a student interacts with a learning resource along a period of time. For example, a learning activity for a student can be a participation to a forum at a certain time. Here, the student interacts with a forum resource, and each time he/she views or posts a message, this information is stored as an event into the raw data.

After preliminary data cleaning and filtering, we obtained a data file containing timestamped task-based learning activities records.

3.3 Data Analysis

3.3.1 Time Discretization

To facilitate analysis of student learning activities and allow comprehensive comparisons, we first set an observation period to one day, to get observations with the same time baseline. We split the course timeline into periods of equal lengths, by performing a splitting of data, into sampled records or segmented chunks of daily learning activity observations.

3.3.2 Student Interaction Vectorization

In order to be computed, the data obtained from the previous step of time discretization, is structured using vectors. Each vector stores for each student, during a period of time information about series of events of learning activities including the resource name, its category (group, individual, assessment), the kind of the resulting tasks (active, passive), and the state of the learning activity (began, ongoing, completed).

3.3.3 Data Aggregation

To prepare the data analysis, we built a new data file of aggregated data. For each student, over a period of time, we stored the values of different analysis variables (Table 1). PCA was conducted with this aggregated data after their standardization.

Table 1: Aggregated data (the PCA variables.)

Analysis variable	Signification
Indiv-Act	count of individual activities
Group-Act	count of group activities
Asses-Act	count of assessment activities
Total-Act	count of total activities
Activ-Task	percentage of active tasks
Divers	diversity of resources (Gini Index)

3.3.4 Data Standardization

Since our ultimate goal is to achieve clustering of student profiles on the basis of learning activity indicators having varying units, it is ideally suited to perform a data standardization on aggregated data computed from the student's interaction vectors. Standardization is a scaling technique where the data values are centered around the mean with a unit standard deviation. The result is that the mean of the attributes becomes zero and the resultant distribution has a unit standard deviation.

3.3.5 Principal Component Analysis

The key objective with PCA is to reduce the dimensionality of a dataset with a large number of interrelated variables, while retaining as much as possible the variation present in the dataset (Jolliffe, 2003). This reduction is achieved by converting the initial variables into a new set of uncorrelated variables, called principal components. Principal components are ordered so that the first few retain most of the variation present in the dataset. The principal components are also called PCA axes or factors. The PCA returns the principal components with their corresponding eigenvalues, reflecting the variability of the reduced initial data. Ideally, a small number of factors with high eigenvalues are retained to ensure good visual representations of data. Correlation refers to the degree of dependence between two variables. In our case, it is measured according to Pearson's correlation coefficient, giving a value between -1 and +1 inclusive. The closer the coefficient is to -1 or +1, the greater is the correlation between the variables.

3.3.6 Agglomerative Hierarchical Clustering

In order to achieve our ultimate goal, which is to identify groups of learners, we conducted an AHC on the new observations' coordinates in the subspace containing the chosen PCA factors. AHC works from the dissimilarities between the observations to be grouped together (Day and Edelsbrunner, 1984). It is an iterative classification method whose process is based on calculating the dissimilarity between observations. Two observations are clustered together when that minimizes a given agglomeration criterion, thus creating a class comprising these two objects. Then the dissimilarity between this class and other observations is calculated using the agglomeration criterion. This process continues until all the observations have been clustered.

4 RESULTS AND DISCUSSION

4.1 Characterizing Student Interactions

The PCA returned 6 factors estimated from the aggregated data listed beforehand, after undergoing standardization (Table 2). The first three eigenvalues represent 72.3% of the initial variability of the data (cumulative variability). We retained the first three factors and ignored the last ones that have low eigenvalues.

Table 2: Principal components (factors) returned by the PCA, with their corresponding eigenvalue, eigenvalue variability and cumulative variability.

Factors	Eigenvalue	Variability (%)	Cumulative (%)
F1	1.97	32.8	32.8
F2	1.37	22.8	55.6
F3	1	16.7	72.3
F4	0.97	16.1	88.4
F5	0.68	11.6	100
F6	0	0	100

Table 3 shows the set of the initial variables defining the first three retained factors. Variables that scored high on the factor F1 are Asses-Act (count of assessment activities) and Total-Act (count of total activities), with respective correlation values of 0.64 and 0.66. We can observe that these two correlation values are very close. This can be interpreted by the fact that students who contributed to defining F1, complete a large number of their activities by accessing assessment resources such as quizzes.

F2 is positively correlated with the variables Indiv-Act (count of individual activities), and Divers (diversity of resources) respectively with the values 0.66 and 0.44. This can be explained by the fact that the course incorporates various individual resources as offered by the platform Moodle, and so students choose varied individual activities.

F3 is significantly correlated with the variable Group-Act (count of group activities) and negatively correlated with the variable Divers. Correlation values between these variables and F3 are respectively 0.80 and -0.49. These results reflect a lack of diversity of group activities used in the course. Consequently, students choosing collaborative activities are those who access to resources that are not diverse.

Based on these statistical analysis, we can conclude that it is possible to distinguish students by: 1) a large number of assessment activities compared to other activities; 2) the number of individual activities with a relatively significant diversity of resources they access to; and contrarily 3) the number of group activities with a relatively non-significant diversity of resources they access to.

4.2 Characterizing Student Profiles

The AHC was conducted on the new observations' coordinates in the 3-dimensional space with respect to the chosen factors F1, F2, and F3. The algorithm re-

Table 3: Correlations between the PCA variables and the first three factors. Refer to table 1 for variable names explanation.

PCA Variable	F1	F2	F3
Indiv-Act	0.01	**0.66**	-0.25
Group-Act	0.16	0.40	**0.80**
Asses-Act	**0.64**	-0.12	-0.22
Total-Act	**0.66**	0.24	0.05
Activ-Task	0.36	-0.36	-0.06
Divers	-0.03	**0.44**	**-0.49**

turned four homogeneous clusters showing, for each student within a time period, the cluster he/she belongs to (Figure 1). Table 4 complements the visual result by giving the size of each cluster as well as some descriptive statistics related to the analysis variables that define significantly the retained factors F1, F2 and F3.

We can observe that the largest number of observations (979) consists of students (by time period) whose assessment activities account for almost activities totaled in the online platform. In fact, the mean value of the variable Asses-Act is the highest one compared to the other clusters. Moreover, in this cluster, the variables Asses-Act and Total-Act have very close values (Table 4). We can also observe from Figure 1, that one part of this cluster points (triangular red symbols) are plotted along the F1 axis with relatively high coordinates values (these points contributed significantly to the definition of F1 axis - we already observed that F1 axis is mostly defined by the two variables Asses-Act and Total-Act). This cluster allows qualifying students by time period with assessment-oriented learning profiles.

A second cluster defines group-oriented learning profiles (74 observations). In fact, we can observe that this cluster is mainly characterized by the largest mean number of group activities and the largest mean number of total activities. Students belonging to this cluster register also a significant average number of assessment activities. In fact, this cluster points are plotted along the axis F1 (blue rectangular symbols in Figure 1), with relatively high coordinates values.

A third cluster allows defining individual-oriented learning profiles (92 observations). The main characteristic of this cluster is a large average number of individual activities, as well as the most significant diversity of resources accessed by students. This may be explained by the diversity of individual resources inherent to the Moodle Plateform. We can observe in Figure 1 that this cluster points are mostly plotted along F2 axis with relatively high coordinates values (magenta diamond symbols). We already observed that F2 axis is defined by the number of individual

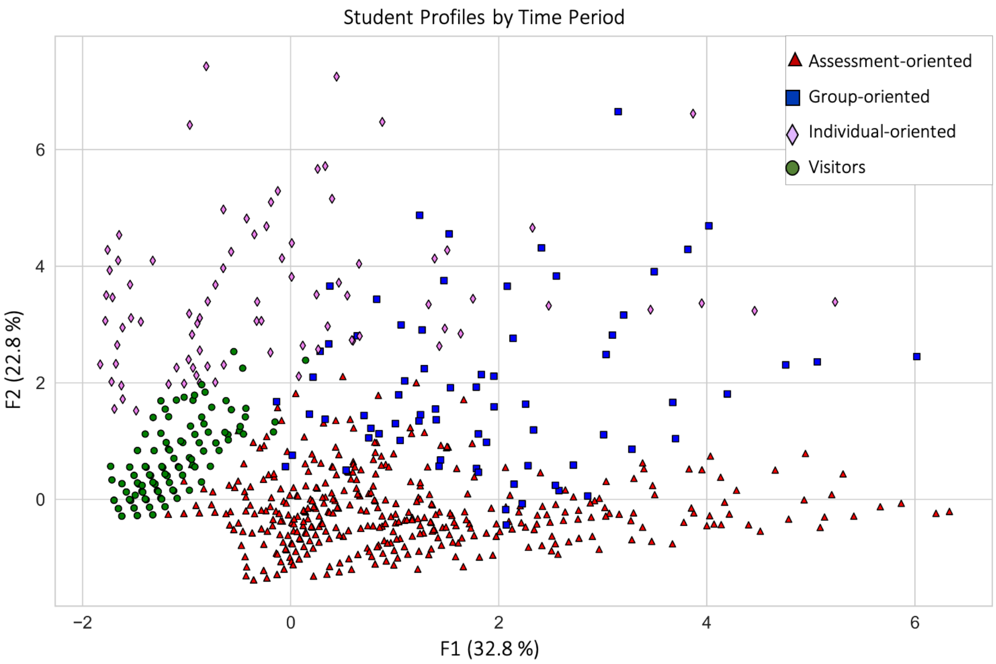

Figure 1: Map showing the partition of students by time period into clusters defining four different profiles.

activities the student performs and the diversity of resources.

The last cluster is the second-largest one (559 observations). It is characterized by low average values on all the variables compared to other clusters. We can observe in Figure 1 that these cluster points are plotted in the positive 2D-plane F1-F2 around the origin, having coordinate values close to zero (green circle symbols). This allows defining students that access to all the learning activities and resources without really achieving tasks and being active. They rather access to the different learning content to only visit and view the course resources. They can be qualified by visitors or content viewers.

4.3 Understanding Student Online Engagement

From these observations we can conclude that students mostly select assessment resources, compared to collaborative or individual ones. We also observed that most of the time, students visit the course content without engaging deep interactions with the provided tasks.

This may be explained by the context of the study, which is a hybrid environment that is also offering social presence and face-to-face dialogues between the teacher and the students. In fact, students may consider that the classroom learning situation is good enough for acquiring knowledge, and decide to keep asynchronous online interactions producing assignments and complete online quizzes.

Students are used to looking for a great documentation from the net and may discard the ones recommended by teachers. They may visit the online course content out of curiosity and decide that the content received in presence would be enough for their learning.

Students also meet in a physical classroom and communicate directly with each other without resorting to a communication tool provided inside the online course. They may decide to perform group activities in presence and use other means of online communication. This may explain the low intensity of online group activities that was measured.

5 CONCLUSION AND FUTURE DIRECTIONS

In this work we have described our methodology and the results of a study that we conducted to investigate learners' profiles when interacting with online course content. We processed a PCA and an AHC in the anal-

Table 4: Cluster size and descriptive statistics related to the variables defining the first three factors. Refer to table 1 for the PCA variable names explanation.

		Assessment	Group	Individual	Visitors
Cluster size		979	74	92	559
Asses-Act	Mean	**11.83**	8.65	4.66	0.83
	std	8.04	8.86	7.75	0.94
Total-Act	Mean	**11.88**	**24.3**	13.67	1.87
	std	8.17	9.13	10.1	2.53
Indiv-Act	Mean	0.4	1.65	**8.62**	0.81
	std	1.11	2.78	5.29	1.45
Group-Act	Mean	0.61	**14.5**	1.48	1.24
	std	1.46	5.96	2.59	2.05
Divers	Mean	0.22	0.57	**10.25**	0.04
	std	1.8	2.74	11.49	0.43

ysis of data, leading to the characterization of different profiles of students.

Regarding the first research question "what are the indicators of student engagement that can be measured from online learning activities?" Insights from the literature helped us to define aggregated data that served as inputs to the PCA. Five variables among six allowed to characterize students interactions with online content. This led to indicators of engagement that are dependent on the experiment context. In fact, these indicators may be different in another context. Regarding the second research question, "what are the student profiles that can be obtained from these indicators, and that can lead to some insights on student engagement ? " The AHC performed on the PCA results allowed for the description of four distinct profiles. These profiles are relative to a hybrid context but may be similar to other contexts such as classroom environments.

In fact, the obtained results showed that students are very selective in their interaction with learning content. They mostly select assessment activities, and visit the learning content without accessing deeply into it. Accessing to individual and group activities is of minimal interest compared to assessment activities and visiting content.

These results show that assessment activities are of a high interest for students, and this is much more likely to engage them. Indeed, the literature argues on the important role of assessment in learning. For example, (Earl, 2012) describes assessment as a motivator for learning. Assessment can motivate students by stimulating their intrinsic interest, and reinforcing the idea that they have control over learning. This can build confidence in students who need to take risks. Assessment can be appealing to student imagination, and provide the scaffolding the students need to succeed. While these observations relate to the classroom environment, our study shows that it is also similar to an online course in a hybrid context. Moreover, this study showed that the course needs to be effectively designed to provide students with high engaging and diverse activities. Improvements need to be done to allow richer interactions with the learning content provided to students.

From this study we can conclude that assessment is a kind of activity which is much more likely to engage students. More importantly, simply making diverse resources available to students is not enough to engage students in learning. The process of designing effective online, face-to-face or blended learning environments needs to be more carefully thought out. More particularly, as the use of online courses increases and blended learning opportunities grow, student interaction with course content will play an important role in the teaching and learning process. Designing effective online courses and blended learning environments requires high-quality instructional content, but also a good combination of face-to-face courses with asynchronous online content, leading to a coherent continuum between the two modalities.

We plan to define new experiments and explore other methods to study the impact of student interactions with online content on learning engagement in hybrid or in face-to-face contexts. The implication on course design should also be considered in future research.

ACKNOWLEDGEMENTS

We would like to thank Campus France and The French Embassy in Kenya for the PhD scholarship program which allowed financing thesis stays in France during which this work was carried out. We also thank Dedan Kimathi University of Technology in Kenya for its support to this PhD work.

REFERENCES

Bates, A. T. et al. (2018). *Teaching in a digital age: Guidelines for designing teaching and learning*.

Carini, R. M., Kuh, G. D., and Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in higher education*, 47(1):1–32.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):318–331.

Coates, H. (2006). *Student engagement in campus-based and online education: University connections*. Routledge.

Coffrin, C., Corrin, L., de Barba, P., and Kennedy, G. (2014). Visualizing patterns of student engagement and performance in moocs. In *Proceedings of the fourth international conference on learning analytics and knowledge*, pages 83–92. ACM.

Day, W. H. and Edelsbrunner, H. (1984). Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of classification*, 1(1):7–24.

Djelil, F., Muller, P.-A., and Sanchez, E. (2018). Investigating learners' behaviours when interacting with a programming microworld. In *Open Conference on Computers in Education*, pages 67–76. Springer.

Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Gunuc, S. and Kuzu, A. (2015). Student engagement scale: development, reliability and validity. *Assessment & Evaluation in Higher Education*, 40(4):587–610.

Henrie, C. R., Bodily, R., Manwaring, K. C., and Graham, C. R. (2015). Exploring intensive longitudinal measures of student engagement in blended learning. *The International Review of Research in Open and Distributed Learning*, 16(3).

Horton, W. (2011). *E-learning by design*. John Wiley & Sons.

Jolliffe, I. (2003). Principal component analysis. *Technometrics*, 45(3):276.

Kuh, G. D. (2009). The national survey of student engagement: Conceptual and empirical foundations. *New directions for institutional research*, 141(1):5–20.

Kumar, S., Gankotiya, A. K., and Dutta, K. (2011). A comparative study of moodle with other e-learning systems. In *2011 3rd International Conference on Electronics Computer Technology*, volume 5, pages 414–418. IEEE.

Lea, S. J., Stephenson, D., and Troy, J. (2003). Higher education students' attitudes to student-centred learning: beyond'educational bulimia'? *Studies in higher education*, 28(3):321–334.

Lei, S. A. (2010). Intrinsic and extrinsic motivation: Evaluating benefits and drawbacks from college instructors' perspectives. *Journal of Instructional psychology*, 37(2):153–161.

Lewis, A. D., Huebner, E. S., Malone, P. S., and Valois, R. F. (2011). Life satisfaction and student engagement in adolescents. *Journal of Youth and Adolescence*, 40(3):249–262.

Lockyer, L., Heathcote, E., and Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10):1439–1459.

Ma, J., Han, X., Yang, J., and Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, 24:26–34.

Means, B., Bakia, M., and Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*. Routledge.

Mosher, R. and MacGowan, B. (1985). Assessing student engagement in secondary schools: Alternative conceptions, strategies of assessing, and instruments.

Murray, M., Pérez, J., Geist, D., and Hedrick, A. (2013). Student interaction with content in online and hybrid courses: Leading horses to the proverbial water. In *Proceedings of the Informing Science and Information Technology Education Conference*, pages 99–115. Informing Science Institute.

Rice, W. (2015). *Moodle e-learning course development*. Packt Publishing Ltd.

Rieber, L. P. (2017). Participation patterns in a massive open online course (mooc) about statistics. *British Journal of Educational Technology*, 48(6):1295–1304.

Robinson, C. C. and Hullinger, H. (2008). New benchmarks in higher education: Student engagement in online learning. *Journal of Education for Business*, 84(2):101–109.

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S. P., Shum, S., Ferguson, R., Duval, E., Verbert, K., Baker, R., et al. (2011). Open learning analytics: an integrated & modularized platform.

Skinner, E. A. and Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology*, 85(4):571.

Skinner, E. A., Wellborn, J. G., and Connell, J. P. (1990). What it takes to do well in school and whether i've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of educational psychology*, 82(1):22.

Xing, W., Guo, R., Petakovic, E., and Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, 47:168–181.

Xiong, Y. and Suen, H. K. (2018). Assessment approaches in massive open online courses: Possibilities, challenges and future directions. *International Review of Education*, 64(2):241–263.

You, J. W. (2016). Identifying significant indicators using lms data to predict course achievement in online learning. *The Internet and Higher Education*, 29:23–30.