




Article

Insights from Learning Analytics for Hands-On Cloud Computing Labs in AWS

Germán Moltó *, Diana M. Naranjo * and J. Damian Segrelles *

Instituto de Instrumentación para Imagen Molecular (I3M), Centro mixto CSIC—Universitat Politècnica de València, 46022 Camino de Vera s/n, Spain

* Correspondence: gmolto@dsic.upv.es (G.M.); dnaranjo@i3m.upv.es (D.M.N.); dquilis@dsic.upv.es (J.D.S.)

Received: 10 November 2020; Accepted: 16 December 2020; Published: 21 December 2020



Abstract: Cloud computing instruction requires hands-on experience with a myriad of distributed computing services from a public cloud provider. Tracking the progress of the students, especially for online courses, requires one to automatically gather evidence and produce learning analytics in order to further determine the behavior and performance of students. With this aim, this paper describes the experience from an online course in cloud computing with Amazon Web Services on the creation of an open-source data processing tool to systematically obtain learning analytics related to the hands-on activities carried out throughout the course. These data, combined with the data obtained from the learning management system, have allowed the better characterization of the behavior of students in the course. Insights from a population of more than 420 online students through three academic years have been assessed, the dataset has been released for increased reproducibility. The results corroborate that course length has an impact on online students dropout. In addition, a gender analysis pointed out that there are no statistically significant differences in the final marks between genders, but women show an increased degree of commitment with the activities planned in the course.

Keywords: learning analytics; cloud computing

1. Introduction

The beginning of the digital era has exponentially increased the amount of data generated. The new data analytic techniques have received special attention in the research from the industrial and academic sectors [1]. The data sources that have emerged in the past years have allowed better characterization of student behavior.

The use of learning management systems (LMSs) has increased in recent years, especially in universities that offer online courses where students can immerse themselves in an individual and collaborative learning experience. In higher education, the use of analytics is an active area of research. Indeed, there are many different concepts when it comes to analytics, and finding a definition that fits all profiles can be complicated [2]. In the work by Barneveld et al. [3], several definitions are collected according to different terms, and they proposed the following: “an overarching concept described as data-driven decision making”.

The growing interest in improving the students’ learning methods has led to the creation of different institutions specialized in exploring the role and impact of analytics in teaching and learning in the education sector. In 2011, the Society for Learning Analytics Research (SoLAR: <http://www.solaresearch.org>) was founded as a non-profit interdisciplinary network of international researchers to explore the impact of big data and learning analytics in the education sector. Learning analytics in academia focuses on gathering the data generated by students during courses to manage student success, including early warning processes where the need for intervention by a teacher can be justified.

With the growth of interest in research on learning analytics, the large amount of data generated in the education sector, and the rapid development of applications and software for data collection, it is important that researchers and educators recognize the characteristics of educational data mining (EDM) and learning analytics and knowledge (LAK). These data-intensive education approaches are increasingly prominent in sectors such as government, health care, and industry [4].

Indeed, as described in [5], the analysis of the data generated by students from the use of online technologies can provide information on the students' progress and the quality of the teaching and curriculum implemented. The analysis of student follow-up data has led to the emergence of new technologies around adaptive learning and recommendation systems for adaptive practice.

In [6], Blinkstein defines multimodal learning analytics (MMLA) as "a set of techniques that can be used to collect multiple sources of data in high frequency (video, logs, audio, gestures, biosensors), synchronize and code the data, and examine learning in realistic, ecologically valid, social, mixed-media learning environments." In this line, the work by Worsley [7] indicates that MMLA "utilizes and triangulates among non-traditional as well as traditional forms of data in order to characterize or model student learning in complex learning environments".

We believe that cloud computing instruction can benefit from the automated compilation of learning analytics, which can then be coupled to additional analytics obtained from LMSs to provide further insights. With this aim, this paper focuses on capturing and analyzing learning analytics gathered from a fee-paying online course in cloud computing with Amazon Web Services (AWS) in order to reveal data-driven insights concerning the students. These data include: (i) activity logs from their interaction with multiple cloud services to determine the students' percentage of progress for each hands-on session and (ii) results of the self-assessment tests that students can optionally take after each module. To gather data from the hands-on lab activities carried out by the students, we introduced a distributed serverless platform to collect and process the logs from AWS CloudTrail, a service that registers the accesses to the AWS services performed by the students. We then translated these data into meaningful activity traces in order to determine the progress percentage for each student in each education activity performed in the cloud.

Several studies are available in the literature concerning how students behave in freely available online courses, such as MOOCs (Massive Open Online Courses) [8,9]. However, fewer studies focus on online fee-paying courses. In these courses, the expectation of dedication of participants has been previously measured to be significantly different compared to freely available MOOCs, as indicated in the work by Cross et al. [10], which found out a statistically significant difference between the anticipated time commitments of fee-paying students and no-fee learners. In addition, the completion rate among fee-paying students for online courses is typically higher compared to students that do not pay a fee, as reported in the work by Koller et al. [11]. This contribution analyzes the data gathered by means of a platform designed from a fee-paying online course with a significant population ($N = 427$). We postulate that the analysis of these data for a cohort of students can aid in the identification of improvements in the course design in order to reduce the dropout rate. We also aim to better characterize the students' behavior when carrying out the hands-on activities in the cloud.

After the introduction, the remainder of the paper is structured as follows. First, Section 2 introduces the related work in the area. Next, Section 3 describes the educational data analytics that are gathered from the students, together with the cloud-based tools designed. Later, Section 4 analyzes the aforementioned data to obtain further insights from the students taking the course, including a gender analysis. Finally, Section 5 summarizes the main achievements of the paper and describes future work.

2. Related Work

Learning analytics can greatly help researchers understand and optimize learning by collecting relevant traces of users and using them for personal improvement [12]. However, for the large amount of data that can be generated, it is important to consider the way in which they are presented for both

students and teachers. A study by Charleer et al. [13] exposed that the “zoom + context” technique together with the use of technologies can be applied for data analysis to help teachers and students deal with large amounts of data.

Previous studies used different sources to produce data, such as click-stream, eye tracking, electroencephalography (EEG), or gesture tracking using sensors and video, to extract and select the characteristics associated with the acquisition of skills; with this, the path to the design of the system is proposed: better learning support through physiological detection [6,14]. Indeed, MMLA captures, integrates, and analyzes the traces of learning available through various sensors and data collection technologies. The multimodal data sets obtained through these techniques can provide new ideas and opportunities to investigate more complex learning processes and environments [15].

The work by Ochoa et al. [16] references the main challenges of multimodal data collection and includes methodologies, techniques, and tools in order to capture, process, and analyze multimodal learning traces. For smart classrooms, the work by Aguilar et al. [17] identifies learning analytics tasks as a set of tools used to collect and analyze data from these classrooms, thus studying the impact on ambient intelligence (AmI) in education.

The adoption of learning analytics by online courses is exemplified by works such as that of Lu et al. [18], who created a tool to produce monthly reports to highlight at-risk students that required a timely intervention to prevent dropping out of MOOCs related to programming. This improved students’ learning and increased the engagement in the course. The work by Drachsler et al. [19] discusses the interaction between MOOCs and learning analytics by introducing a conceptual framework (MOLAC) that focuses on key areas required to be enabled, such as data-sharing facilities across institutions and standardized evaluation approaches. Indeed, the work by Rui Pérez-Valiente et al. [20] introduced a learning analytics tool implemented for Open edX, called ANALYSE, which provides useful visualizations for teachers’ feedback backed by pedagogical foundations. The work by Er et al. [21] uses learning analytics to design a predictive analytics solution to involve instructors in the design process of an MOOC. Finally, the work by Tabaa and Medouri [22] focuses on creating a learning analytics system for massive open online courses based on big data techniques, such as Hadoop, in order to target “at-risk” students.

However, there are few works in the literature that support learning analytics by means of cloud computing techniques. For example, the work by Shorfuzzaman et al. [23] presents a cloud based mobile learning framework that uses a Big Data analytics technique to extract values from mobile learners’ data. The proposed big learning data analytic model uses cloud computing to provide scalable computing and data storage resources. The work by Klačnja-Milićević et al. [24] identified the importance of Big Data in performing efficient processing of learning analytics, and an abstract architecture framework was envisioned that involves the usage of cloud computing.

There exist some commercial platforms that can ingest logs coming from CloudTrail in order to achieve further visibility of the activity taking place in an AWS account. This is the case of Loggly [25], a unified logging system that can produce enhanced dashboards from the processed data. However, exploiting this information to produce learning analytics in order to achieve additional insights into the degree of performance of students doing the cloud-based activities is the key point of our contribution.

From the analysis of the state of the art, we can notice that one of the fundamental challenges is to characterize, using the data obtained through different techniques, the population of students who take the courses. To achieve this goal, the main contribution to the state of the art of this work lies in the automated compilation of learning analytics from hands-on activities in AWS via a cloud-based architecture and the subsequent analysis of said data together with the academic results from the LMS for students from an online course on cloud computing with Amazon Web Services. The analysis provides further insights in order to steer the course design in light of the activities carried out by students during the time frame allocated to its completion.

3. Educational Data Analytics

The online course in cloud computing with Amazon Web Services (CursoCloudAWS (in Spanish): <https://www.grycap.upv.es/cursocloudaws>) was the first Spanish-speaking online course on AWS offered worldwide, and has trained more than 1000 students from 10 countries (mainly Spain and Latin America) since 2013. It is a fully online self-paced experience that involves multiple learning materials, such as video lessons, hands-on lab guides, remote virtual labs for student to self-deploy their own virtual infrastructures, and self-assessment tools, as described in the work by Moltó et al. [26]. It cannot be considered an MOOC, since it is not offered for free, but it shares many of the features of these courses with respect to the challenges of online instruction.

Figure 1 shows the different types of data collected from a student during the course. Students are provided with user credentials with limited privileges linked to the teacher's AWS account. A Lab Machine configured with multiple Linux user accounts is automatically deployed in AWS for students to carry out the hands-on labs defined in the course. This machine is automatically deployed and configured for each academic course using the Infrastructure Manager (IM) (Infrastructure Manager (IM): <https://www.grycap.upv.es/im>) [27], as described in the work by Segrelles et al. [28].

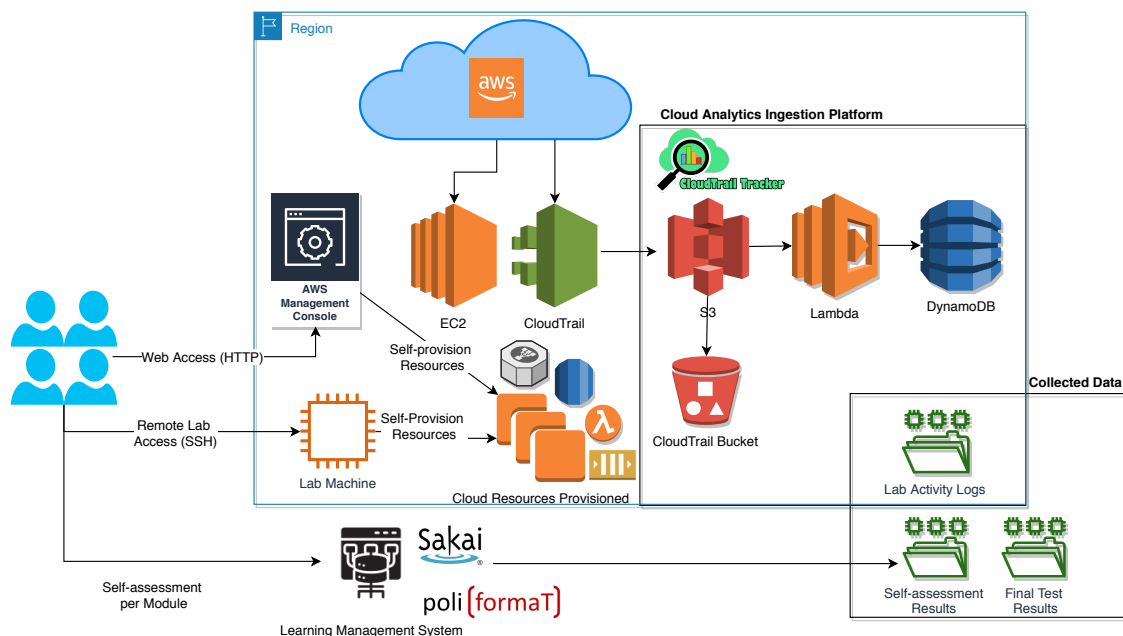


Figure 1. Data gathered throughout the cloud course.

Once the students start using the cloud services, AWS CloudTrail [29] periodically delivers activity logs of the services used into a permanent storage bucket in Amazon S3 [30]. These data are automatically collected, parsed, and ingested by the CloudTrail-Tracker (CloudTrail-Tracker: <https://www.grycap.upv.es/cloudtrail-tracker>), an open-source platform that provides a graphical dashboard for teachers to visualize progress across multiple lab activities for each student, as described in the work by Naranjo et al. [31]. For the sake of completeness, a summary of the role of CloudTrail-Tracker is included in Section 3.1.

After each module, students are encouraged to undertake an optional self-assessment test in order to determine their level of knowledge with respect to both the theoretical and practical concepts studied in each module. These tests are implemented in PoliformaT, the Sakai-based learning management system (LMS) used at the Universitat Politècnica de València in Spain. In addition, after each module, the student is asked to provide feedback on the level of satisfaction with respect to the corresponding hands-on lab activities performed. This way, we can match the perception of the student with respect to the actual development of activities within the cloud platform, though this study is outside the scope of this paper.

3.1. Gathering and Processing Activity Logs for Students in the Cloud

Figure 2 describes the architecture of the platform created in order to automatically process the activity logs that trace the actions performed by the students in AWS while carrying out the lab activities.

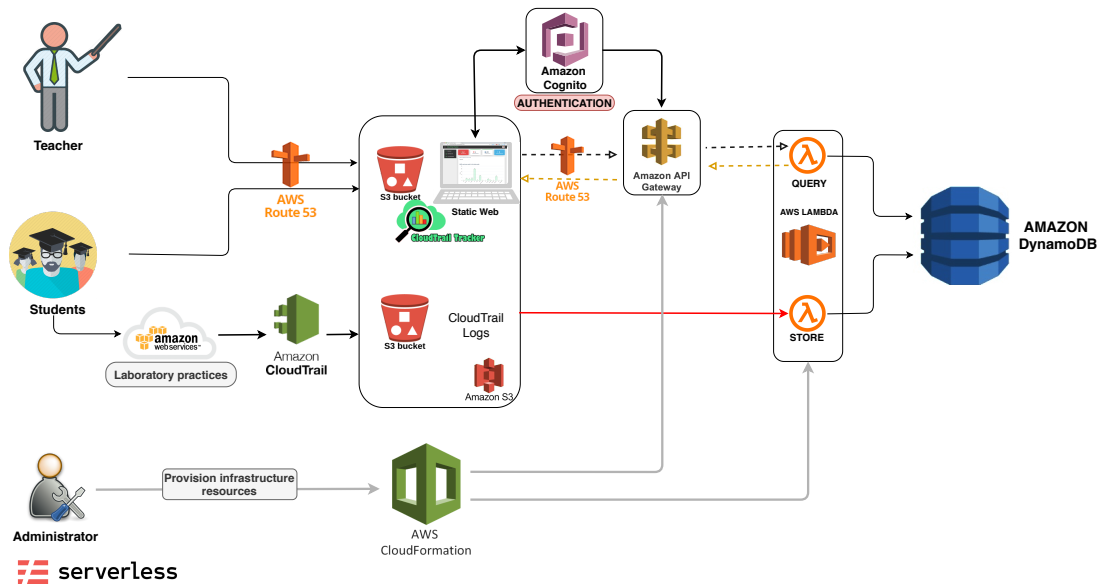


Figure 2. Distributed platform to collect, store, process, and visualize educational analytics for cloud training in Amazon Web Services (AWS).

The platform is defined as a completely serverless application, where the cloud provider manages the capacity allocation for the underlying services employed [32]. The activity logs of the students performing the lab sessions are automatically generated by AWS CloudTrail and delivered up to 15 min after the actions are carried out. These are compressed JSON files that include records about each time an API (Application Programming Interface) is invoked, that is, every time a user performs an action with an AWS service that is supported by CloudTrail. Each record contains valuable information that includes:

- **WHO.** The AWS IAM (identity and access management) user—in our case, the student—that made the action.
- **WHAT.** The specific action performed, that is, the API call invoked for each AWS service.
- **WHEN.** The timestamp that indicates when the action was carried out.
- **WHERE.** The client tool used to perform the action, typically a web browser or a command-line tool.

These logs are automatically uploaded as files to an Amazon S3 bucket, which triggers the execution of a Lambda function (*Store*) responsible for storing the events (actions carried out by the student) in a NoSQL table in AWS DynamoDB. In order to minimize the amount of information stored in the database, actions that do not modify the infrastructure are discarded. Only those actions that involve creating or modifying a resource in AWS are logged, as required in the hands-on activities proposed. A REST API to query this information is created via the API Gateway, which, upon a request, triggers a Lambda function (*Query*) responsible for querying the DynamoDB table and formatting the results in JSON.

The application has a web-based graphical front end coded as a Vue.js application that is compiled into a static website (only HTML, CSS, and JavaScript), which is hosted in an Amazon S3 bucket and exposed via a custom DNS (domain name service) entry through Amazon Route 53. The web application, running in the user's browser, relies on Amazon Cognito to perform user authentication

via the user's credentials. This is also employed to obtain valid access tokens to query the REST API to provide programmatic access to the collected data.

This application architecture has been fully released as an open-source platform to gather activity logs, which can be exploited as a learning dashboard aimed at self-regulation of students, as demonstrated in the work by Naranjo et al. [31]. This work analyzes the insights obtained from the course-related data collected through both CloudTrail-Tracker and the LMS.

3.2. Statistical Variables Defined

The following variables are defined:

- *CTOCA*, which represents the mark obtained in the final timed exam that students undertake. They need to achieve a mark greater than or equal to 5 in order to obtain the certificate of achievement. These data are collected from the LMS.
- *CTAM1* through *CTAM7*, which indicate the results of the optional self-assessment test carried out by the student after each one of the seven modules, ranging from 0 to 10. This value may be absent for students that did not perform the test. These data are collected from the LMS. The variable *CTAMT* is the average of the seven aforementioned variables, where 0 is assigned for a non-taken self-assessment test for the corresponding module.
- *PL_X* where $X = EC2, EC2_S3, RDS, APP, CF, VPC,$ and *LAMBDA_SQS*, which indicate the percentage of progress of a student in each lab activity for each student (in order of appearance during the course). These data are collected by CloudTrail-Tracker. The variable *PL_TOTAL* is the average of the seven aforementioned variables, where 0 is assigned for a lab activity not carried out by the student.

For the sake of completeness, a brief summary of the lab activities is included:

- *PL_EC2*: Students deploy and configure virtual machines (VMs) with different operating systems using Amazon EC2.
- *PL_EC2_S3*: Students create fleets of VMs that can grow and shrink depending on the workload being processed. They also interact with Amazon S3 to perform object storage for the management of files.
- *PL_RDS*: Students use Amazon RDS to manage the lifecycle of relational database management systems (RDBMS), such as MySQL, using fault-tolerant deployment of databases.
- *PL_APP*: Students deploy a highly available architecture to support the Wordpress blogging platform by replicating and distributing the internal components required.
- *PL_CF*: Students use Amazon CloudFormation to deploy complete application architectures described in template files using an Infrastructure as Code (IaC) approach.
- *PL_VPC*: Students create isolated networking sections in AWS to better increase the security of the deployed applications by using Amazon VPC (Virtual Private Cloud).
- *PL_LAMBDA_SQS*: Students create serverless event-driven applications to process files using queue services such as Amazon SQS and Functions as Service (FaaS) approaches via AWS Lambda.

It is important to point out that these are not the only lab activities carried out throughout the course, but those from which automated metrics can be obtained via CloudTrail-Tracker. However, these lab activities are spread across the course and, therefore, represent a good proxy for the amount of work carried out by the students when performing their practical educational activities in AWS.

4. Insights from Data Analytics

This section analyzes the data gathered across three academic courses (from 2016/2017 to 2018/2019) from an initial population of 427 students (380 males and 47 females) that took the online course in cloud computing with Amazon Web Services (101 students in 2018/2019, 160 students in

2017/2018, and 166 students in 2016/2017). The students were mainly from Spain, with some students from Latin America. The full raw data to reproduce the statistical results are available as a publicly available Google Spreadsheet (<https://bit.ly/as-cloudylitics>).

4.1. Statistical Overview

First, we focus on the aggregated statistics of the students. With this aim, Table 1 shows the number of students (*COUNT*) that undertook the voluntary assessment test for each one of the seven modules of the course, together with the average mark on a 0–10 scale (*AVG*) and the standard deviation (*STD*). The same information is obtained for the final timed test (*CTOCA*). The last line (*DROP*) indicates the percentage of students that did not take the corresponding test with respect to the total population of students.

Table 1. Statistical overview of the students' results for the theoretical activities across three academic years ($N = 427$).

STAT	CTAM1	CTAM2	CTAM3	CTAM4	CTAM5	CTAM6	CTAM7	CTOCA
COUNT	386	346	336	323	314	276	230	321
AVG	9.24	8.59	8.72	8.93	8.49	8.56	8.79	8.70
STD	0.97	1.32	1.36	1.25	1.51	1.49	1.58	1.05
DROP (%)	9.6	19.0	21.3	24.4	26.5	35.4	46.1	24.8

The results indicate that 75.17% of the students took the final test, obtaining an average grade of 8.70 with a standard deviation of 1.05. The dropout rate of students with respect to the voluntary tests starts at 9.6%, which stands for the percentage of students that did not take the test after module 1, and increases throughout the different modules, up to a total 46.1% of students that did not take the test in module 7. Notice, though, that the dropout rate of students that did not perform the final test goes down to 24.8%, since students do not typically want to miss the opportunity to achieve the final course completion certificate.

The rationale underlying this behavior may be two-fold. On the one hand, the students may be facing difficulties with keeping up with the amount of material of the course. Even if the course is self-paced, where students are provided with all of the material and the cloud environment required to perform all the activities at their own speed, a tentative schedule is suggested, though not enforced. The length of the course, which is planned for a four-month period (120 days), may impact the dropout rate of the students. Previous studies on attrition in online learning, such as the one carried out by Diaz and Cartnal [33], indicated that shorter term length was a key variable affecting student attrition, since many students argue that time requirements represent a barrier to enroll in an online course.

Table 2 shows statistics of the percentage progress of students with respect to each lab activity in the cloud, visually depicted in Figure 3. Notice that we provide average values that both include students that did not perform a specific lab activity (i.e., the percentage of progress is 0) and exclude them in order to focus, in this latter case, on students that carried out the lab activities. Indeed, as the course develops, the average percentage of progress for each lab activity decreases. The difference between the blue and red bars increases more towards the end of the course. This means that fewer people carry out these lab activities towards the end of the course, but those who do achieve a good percentage of progress and almost fulfill all the activities within each lab activity in the cloud. Note also the decrease in the number of students for certain lab activities (e.g., *DROP* value for *PL_RDS*), showing the preference of students for the other lab activities.

Table 2. Statistical overview of the students’ results for the lab activities in the cloud across three academic years ($N = 427$).

STAT	PL_EC2	PL_EC2_S3	PL_RDS	PL_APP	PL_CF	PL_VPC	PL_LAMBDA_SQS
COUNT	328	319	253	310	311	236	102
AVG (>0)	93.67	72.88	78.77	59.18	53.63	64.16	82.45
AVG (>=0)	72.12	54.58	46.78	43.07	39.15	35.54	19.74
DROP (%)	23.2	25.3	40.7	27.4	27.2	44.7	76.1

Notice that this information could not have been gathered and analyzed without using CloudTrail-Tracker, thus highlighting the benefits of collecting educational analytics for cloud instruction using cloud techniques as well. The results indicate a similar behavior when compared to those obtained when carrying out the theoretical activities. However, the drop rate is substantially increased for the latest lab activities. This indicates that students may be facing difficulties in dedicating the required amount of time to perform all the activities and, therefore, they prioritize the theoretical activities that would increase their chances of passing the final test.

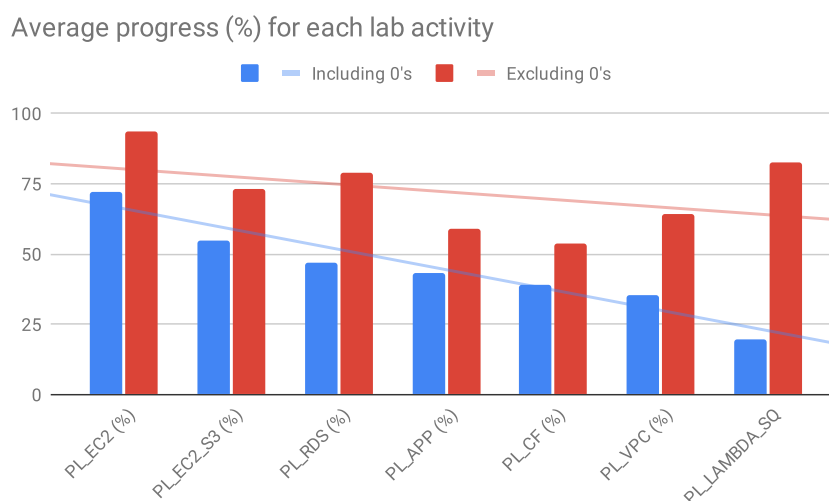


Figure 3. Average progress (percentage) for each lab activity across the student population ($N = 427$).

4.2. Gender Analysis: An Statistical Approach

Unfortunately, gender inequality in STEM subjects (science, technology, engineering, and mathematics) is a reality in universities around the world, as reported in the UNESCO (United Nations Educational, Scientific and Cultural Organization) report [34], which indicates that only 35% of students registered in STEM studies and 28% of researchers are women. We wanted to contribute with this work to the elimination of gender stereotypes and prejudices that compromise the quality of the students’ learning experience and limit their educational options [35].

With this aim, three statistical studies were carried out to compare the populations of women and men. To determine if the distributions are normal, the Kolmogorov–Smirnov (K-S) test was used, since the number of values in the distributions was greater than 30 [36]. Because the results point to non-normal distributions, the Mann–Whitney U [37] non-parametric test was used to compare two distributions, and the Kruskal–Wallis (K-W) one-way analysis of variance (ANOVA) [38] was used to compare more than two distributions using IBM SPSS [39].

Female advantage in school marks is a common finding in education research, as indicated in the meta-analysis work by Voyer et al. [40]. Therefore, the target of the first statistical study was to compare the final test results between women and men. The study included 321 students (10.59% women and 89.41% men) who achieved a final grade (variable *CTOCA*) and excluded 106 students (12.26%

women and 87.74% men) who had not taken the final test. The K-S test shown in Table 3 indicates that the distribution of men is not normal ($p(K-S) < 0.05$). Therefore, the Mann–Whitney U test was performed ($p(U-M-W) = 0.460$), retaining the null hypothesis (the distribution of *CTOCA* is the same between genders), and thus concluding that there are no significant differences in the final grades between genders.

Table 3. Comparative study of learning between women and men.

Gender	N	AVG	STD. DEV.	p(K-S)	p(U-M-W)
Male	287	8.689	1.0455	0.004	0.460
Female	34	8.800	1.0714	0.285	

The second study included 386 students (10.36% women and 89.64% men) who had performed the self-assessment tests (variable *CTAMT*), and it excluded 41 students (17.07% women and 82.93% men) because they had not taken these optional tests. The K-S test shown in Table 4 indicates that both distributions are not normal ($p(K-S) < 0.05$). Therefore, the U-M-W test was performed ($p(U-M-W) = 0.025$), rejecting the null hypothesis (the distribution of *CTAMT* is the same between the categories of gender). This allows the conclusion that there are statistically significant differences between women and men with respect to the results of the optional self-assessment tests. The rationale behind these results requires further analysis. Previous works by Ellemers et al. [41] revealed no gender differences in work commitment from a population of doctoral students in the Netherlands as they attempted to find possible explanations for the underrepresentation of women among university faculty. However, few studies are available to contrast the results in terms of commitment to a course between genders. For example, in the work by Sheard et al. [42], female students successfully surpassed male studies in the academic evaluation criteria and reported higher mean scores on commitment compared to their male counterparts.

Table 4. Comparative study of commitment to the course between women and men.

Gender	N	AVG	STD. DEV.	p(K-S)	p(U-M-W)
Male	346	7.107	2.6999	0.000	0.025
Female	40	7.811	3.6764	0.007	

As shown in Figure 3, the percentage of progress decreases as the course progresses through the different lab activities. The results of Table 5 indicate significant differences ($p(K-W) < 0.05$) between the first lab activity (*PL_EC2*) and the following ones with respect to the degree of progress. This can be motivated by the relinquishment associated with students with low involvement and commitment, who usually drop out in the first sessions or activities. In the third and fourth lab activities, there are no significant differences in the progress with respect to the subsequent lab activities, probably because the least motivated students had already left the course in the first activity. Finally, there are always significant differences with respect to the last lab activity (*PL_LAMBDA*), which is carried out at the end of the course, and many students fail to reach that point in the course. This again stresses the need to further allocate time extensions to the course design or split it into two courses.

Table 5. Comparative study of progress in all the lab activities with respect to the subsequent ones.

Sample1	Sample2	p(K-W)	Sample1	Sample 2	p(K-W)
PL_EC2	PL_EC2_S3	0.000	PL_EC2_S3	PL_RDS	0.380
	PL_RDS	0.000		PL_APP	1.000
	PL_APP	0.000		PL_CF	0.020
	PL_CF	0.000		PL_VPC	0.000
	PL_VPC	0.000		PL_LAMBDA	0.000
	PL_LAMBDA	0.000			
PL_RDS	PL_APP	1.000	PL_APP	PL_CF	1.000
	PL_CF	1.000		PL_VPC	0.149
	PL_VPC	0.020		PL_LAMBDA	0.000
	PL_LAMBDA	0.000			
PL_CF	PL_VPC	1.000	PL_VPC	PL_LAMBDA	0.000
	PL_LAMBDA	0.000			

The third and last statistical study aims to compare the degree of progress in the lab activities planned in the course. This study included 427 students (11% women and 89% men) who had carried out lab activities and, therefore, had an average percentage of progress identified in the variable *PL_TOTAL*. The K-S test shown in Table 6 indicates that the male distribution is not normal ($p(K-S) < 0.05$). Therefore, the U-M-W test was performed ($p(U-M-W) = 0.872$), retaining the null hypothesis, and thus allowing the conclusion that there are not significant differences between women and men.

Table 6. Study of the degree of progress in lab activities.

Gender	N	AVG	STD. DESV	p(K-S)	p(U-M-W)
Male	380	44.46	34.838	0.000	0.872
Female	47	39.215	39.215	0.217	

5. Conclusions and Future Work

This paper has focused on collecting learning analytics from an online course on cloud computing. The ability to automatically track the progress of the students across the course has enabled the identification of higher dropout rates in the lab activities towards the end of the course. This corroborates the results from other authors on the impact of the length of the course on attrition. In addition, the results of the gender analysis showed similar academic results between genders, but increased commitment from women. The dataset has been made available to reproduce the results of the paper.

To enable this analysis, a cloud-based architecture for automated data ingestion from the activity logs of the students in AWS has been used. This has allowed transformation of the logs into meaningful learning analytics related to the degree of completion for the main hands-on lab activities in the course. By adopting a serverless computing strategy, the platform can operate at zero cost under the free tier of the involved services (mainly AWS Lambda, API Gateway, and Amazon DynamoDB).

The ability to systematically monitor how students behave in the course paves the way for rethinking the strategy of how a course is offered for subsequent editions. This is specially important for online instruction in which other sources of information from students are available apart from the educational trails that students leave during their learning procedure. This can be used for teachers to automatically capture them in order to extract information from the data.

The platform introduced here automatically obtains the percentage of progress for each student in each hands-on activity. Without such a platform, the instructor has no perception of how the students are progressing through their activities. In addition, this is entirely achieved without any

active involvement by the students, who are passively monitored; therefore, they do not need to spend additional time producing an activity report to justify that they carried out the lab activities.

Future work includes further evolving the CloudTrail-Tracker with predictive modules in order to anticipate the dropping out of students and to alert professors to introduce corrective countermeasures, such as extending the allocated timeframe for accessing the course or unblocking additional material in order to minimize the knowledge gap required to undertake the latest course modules.

Author Contributions: Conceptualization, G.M.; methodology, G.M. and D.M.N.; software, D.M.N.; investigation, G.M., D.M.N., and J.D.S.; resources, G.M.; data curation, D.M.N. and J.D.S.; writing—original draft preparation, G.M.; writing—review and editing, G.M., D.M.N., and J.D.S.; visualization, G.M., D.M.N., and J.D.S.; funding acquisition, G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish “Ministerio de Economía, Industria y Competitividad through grant number TIN2016-79951-R (BigCLOE)”, the “Vicerrectorado de Estudios, Calidad y Acreditación” of the Universitat Politècnica de València (UPV) to develop the PIME B29 and PIME/19-20/166, and by the Conselleria d’Innovació, Universitat, Ciència i Societat Digital for the project “CloudSTEM” with reference number AICO/2019/313.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Motiwalla, L.; Deokar, A.V.; Sarnikar, S.; Dimoka, A. Leveraging Data Analytics for Behavioral Research. *Inf. Syst. Front.* **2019**, *21*, 735–742. [[CrossRef](#)]
2. Cooper, A. What is Analytics? Definition and Essential Characteristics. *CETIS Anal. Ser.* **2012**, *1*, 1–10.
3. Van Barneveld, A.; Arnold, K.; Campbell, J. Analytics in Higher Education: Establishing a Common Language. *Educ. Learn. Initiat.* **2012**, *1*, I–II.
4. Siemens, G.; Baker, R.S. Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the ACM International Conference Proceeding Series*; Association for Computing Machinery: New York, NY, USA, 2012; pp. 252–254. [[CrossRef](#)]
5. Siemens, G.; Dawson, S.; Lynch, G. *Improving the Quality and Productivity of the Higher Education Sector Policy and Strategy for Systems-Level Deployment of Learning Analytics Society for Learning Analytics Research*; Technical Report; Society for Learning Analytics Research for the Australian Office for Learning and Teaching: Canberra, Australia, 2013.
6. Blikstein, P. Multimodal learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, Leuven, Belgium, 8–12 April 2013. [[CrossRef](#)]
7. Worsley, M. Multimodal learning analytics’ past, present, and, potential futures. In *Proceedings of the 8th International Conference on Learning Analytics & Knowledge*, Sydney, Australia, 6 March 2018; Volume 2163.
8. Hewson, E.R. Students’ emotional engagement, motivation and behaviour over the life of an online course: Reflections on two market research case studies. *J. Interact. Media Educ.* **2018**. [[CrossRef](#)]
9. Kahan, T.; Soffer, T.; Nachmias, R. Types of participant behavior in a massive open online course. *Int. Rev. Res. Open Distance Learn.* **2017**, *18*, 1–18. [[CrossRef](#)]
10. Cross, S.; Whitelock, D. Similarity and difference in fee-paying and no-fee learner expectations, interaction and reaction to learning in a massive open online course. *Interact. Learn. Environ.* **2017**, *25*, 439–451. [[CrossRef](#)]
11. Koller, D.; Ng, A.; Do, C.; Chen, Z. Retention and Intention in Massive Open Online Courses: In Depth. *Educ. Rev.* **2013**, *48*, 62–63.
12. Phil Long, B.; Siemens, G. Penetrating the Fog: Analytics in Learning and Education. *Ital. J. Educ. Technol.* **2014**, *22*, 132–137.
13. Charleer, S.; Klerkx, J.; Duval, E. Learning dashboards. *J. Learn. Anal.* **2014**, *1*, 199–202. [[CrossRef](#)]
14. Worsley, M. Multimodal learning analytics—Enabling the future of learning through multimodal data analysis and interfaces. In *Proceedings of the ICMI’12—ACM International Conference on Multimodal Interaction*, Santa Monica, CA, USA, 22–26 October 2012; pp. 353–356. [[CrossRef](#)]

15. Spikol, D.; Prieto, L.; Rodríguez-Triana, M.; Worsley, M.; Ochoa, X.; Cukurova, M. Current and future multimodal learning analytics data challenges. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada, 13–17 March 2017; pp. 518–519. [CrossRef]
16. Ochoa, X.; Worsley, M.; Weibel, N.; Oviatt, S. Multimodal learning analytics data challenges. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh, UK, 25–29 April 2016; pp. 498–499. [CrossRef]
17. Aguilar, J.; Sánchez, M.; Cordero, J.; Valdiviezo-Díaz, P.; Barba-Guamán, L.; Chamba-Eras, L. Learning analytics tasks as services in smart classrooms. *Univers. Access Inf. Soc.* **2018**, *17*, 693–709. [CrossRef]
18. Lu, O.H.; Huang, J.C.; Huang, A.Y.; Yang, S.J. Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. *Interact. Learn. Environ.* **2017**, *25*, 220–234. [CrossRef]
19. Drachler, H.; Kalz, M. The MOOC and learning analytics innovation cycle (MOLAC): A reflective summary of ongoing research and its challenges. *J. Comput. Assist. Learn.* **2016**, *32*, 281–290. [CrossRef]
20. Ruiperez-Valiente, J.A.; Muñoz-Merino, P.J.; Gascon-Pinedo, J.A.; Kloos, C.D. Scaling to Massiveness with ANALYSE: A Learning Analytics Tool for Open edX. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 909–914. [CrossRef]
21. Er, E.; Gómez-Sánchez, E.; Dimitriadis, Y.; Bote-Lorenzo, M.L.; Asensio-Pérez, J.I.; Álvarez-Álvarez, S. Aligning learning design and learning analytics through instructor involvement: A MOOC case study. *Interact. Learn. Environ.* **2019**, *27*, 685–698. [CrossRef]
22. Tabaa, Y.; Medouri, A. LASyM: A Learning Analytics System for MOOCs. *Int. J. Adv. Comput. Sci. Appl.* **2013**, *4*. [CrossRef]
23. Shorfuzzaman, M.; Hossain, M.S.; Nazir, A.; Muhammad, G.; Alamri, A. Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. *Comput. Hum. Behav.* **2019**, *92*, 578–588. [CrossRef]
24. Klačnja-Milićević, A.; Ivanović, M.; Budimac, Z. Data science in education: Big data and learning analytics. *Comput. Appl. Eng. Educ.* **2017**, *25*, 1066–1078. [CrossRef]
25. SolarWinds. Loggly. Available online: <https://www.loggly.com/> (accessed 19 December 2020).
26. Moltó, G.; Caballer, M. On Using the Cloud to Support Online Courses. In Proceedings of the 2014 Frontiers in Education Conference (FIE), Madrid, Spain, 22–25 October 2014; pp. 330–338. [CrossRef]
27. Caballer, M.; Blanquer, I.; Moltó, G.; de Alfonso, C. Dynamic Management of Virtual Infrastructures. *J. Grid Comput.* **2015**, *13*, 53–70. [CrossRef]
28. Segrelles, J.D.; Moltó, G.; Caballer, M. Remote Computational Labs for Educational Activities via a Cloud Computing Platform. In Proceedings of the Information Systems Education Conference (ISECON), Orlando, FL, USA, 5–7 November 2015; pp. 309–321.
29. AWS. AWS CloudTrail. Available online: <https://aws.amazon.com/cloudtrail/> (accessed 19 December 2020).
30. Amazon. Amazon Simple Storage Service (Amazon S3). Available online: <http://aws.amazon.com/s3/> (accessed 19 December 2020).
31. Naranjo, D.M.; Prieto, J.R.; Moltó, G.; Calatrava, A. A Visual Dashboard to Track Learning Analytics for Educational Cloud Computing. *Sensors* **2019**, *19*, 2952. [CrossRef]
32. Baldini, I.; Castro, P.; Chang, K.; Cheng, P.; Fink, S.; Ishakian, V.; Mitchell, N.; Muthusamy, V.; Rabbah, R.; Slominski, A.; et al. Serverless computing: Current trends and open problems. In *Research Advances in Cloud Computing*; Springer: Singapore, 2017; pp. 1–20. [CrossRef]
33. Diaz, D.; Cartnal, R. Term Length as an Indicator of Attrition in Online Learning. *Innov. J. Online Educ.* **2006**, *2*.
34. Chavatzia, T. *Cracking the Code: Girls' and Women's Education in Science, Technology, Engineering and Mathematics (STEM)*; UNESCO: Paris, France, 2017.
35. Barbera, E.; Candela, C.; Ramos, A. Career selection, professional development and gender stereotypes. *Rev. Psicol. Soc.* **2008**, *23*, 275–285.
36. Daniel, W.W. Kolmogorov–Smirnov one-sample test. *Appl. Nonparametric Stat.* **1990**, *2*.
37. Zimmerman, D.W. Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances. *J. Exp. Educ.* **1987**, *55*, 171–174. [CrossRef]
38. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [CrossRef]

39. Field, A. *Discovering Statistics Using IBM SPSS Statistics*; SAGE, London, UK, 2013.
40. Voyer, D.; Voyer Susan D., D. Gender differences in scholastic achievement: A meta-analysis. *Psychol. Bull.* **2014**, *140*, 1174–1204. [[CrossRef](#)]
41. Ellemers, N.; Van Den Heuvel, H.; De Gilder, D.; Maass, A.; Bonvini, A. The underrepresentation of women in science: Differential commitment or the queen bee syndrome? *Br. J. Soc. Psychol.* **2004**, *43*, 315–338. [[CrossRef](#)] [[PubMed](#)]
42. Sheard, M. Hardiness commitment, gender, and age differentiate university academic performance. *Br. J. Educ. Psychol.* **2009**, *79*, 189–204. [[CrossRef](#)] [[PubMed](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).