

Analysis of Learning Records to Detect Student Cheating on Online Exams: Case Study during COVID-19 Pandemic

Antonio Balderas

University of Cadiz, Av. Universidad de Cádiz 10, CP
11519, Puerto Real, Spain
antonio.balderas@uca.es

Juan Antonio Caballero-Hernández

University of Cadiz, Av. Universidad de Cádiz 10, CP
11519, Puerto Real, Spain
juanantonio.caballero@uca.es

ABSTRACT

In March 2020, due to the Covid19 pandemic, higher education had to switch from face-to-face to exclusively virtual mode overnight. In this unexpected scenario, supervisors also had to adapt the assessment procedures, including the exams. This caused a significant controversy, as, according to many students, supervisors were more concerned about how to prevent students from cheating, than actually measuring their learning. This paper introduces an experience that implemented several of the students' requests in an online exam and conducts a comprehensive analysis of students' behavior according to the virtual learning environment records. Different existing software tools are used for the analysis, complemented with a Python application ad-hoc developed. The objective indicators gathered provide evidence that some students cheated and invite focusing on evidence-based assessment.

CCS CONCEPTS

• Applied computing; • Education; • Learning management systems;

KEYWORDS

Learning analytics, cheating, online exams, virtual learning environments

ACM Reference Format:

Antonio Balderas and Juan Antonio Caballero-Hernández. 2020. Analysis of Learning Records to Detect Student Cheating on Online Exams: Case Study during COVID-19 Pandemic. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'20)*, October 21–23, 2020, Salamanca, Spain. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3434780.3436662>

1 INTRODUCTION

The challenges that formal education faced as a result of the COVID-19 pandemic were numerous and unexpected. In Spain, higher education students suddenly stopped attending classes in March 2020 and universities were forced to switch from face-to-face to virtual teaching overnight [10]. This posed multiple issues for universities: updating and expanding servers to provide greater capacity for

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

TEEM'20, October 21–23, 2020, Salamanca, Spain

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8850-4/20/10...\$15.00
<https://doi.org/10.1145/3434780.3436662>

virtual learning environments (VLE), providing support for teaching staff to adopt the use of technologies for online education, and adapting teaching and assessment methods to this new scenario, among others.

While switching from face-to-face to virtual education posed several problems, the assessment process had to additionally face the cheating menace. This work arises from the claims expressed by many students on social networks and in the media concerning an excessive focus on students cheating [13]. According to the students, this concern was leading supervisors to set extremely demanding exams. Firstly, they considered they were given too little time to answer the questions on the exams. Secondly, students reported problems with connectivity to the institutional VLE when they were all online at once, so they demanded that the exams be asynchronous, i.e., made available for a certain period of time, so they could connect little by little to take the exam. Taking these students' assertions into consideration, was this concern on the side of the supervisors justified?

In order to answer this question, this paper introduces how an undergraduate course collected both requests from students (longer time to take the exam and asynchronous availability) and implemented them in an exam. Then, the results obtained and the records of the VLE are analyzed to check evidence of students cheating in the exam. For this purpose, a Python tool was developed to help process the records of the VLE. This tool was complemented with other software systems to visualize and analyze students' records. The objective indicators gathered provide evidence that some students cheated and invite focusing on evidence-based assessment.

The rest of the paper is structured as follows. First, the context in which this work took place is introduced. Secondly, the software tools used are described. Thirdly, the case study is presented. Fourthly, the results are analyzed. In the fifth section, the results are discussed. Finally, the conclusions and future work are outlined.

2 BACKGROUND

Extensive literature has been written about the cheating of online assessment and the need for progress in the technical development of countermeasures [3, 15, 16]. Although the use of VLEs provides supervisors with the opportunity to use different activities to assess their students, some of them even automatically, final exams are still given in person. In fact, even online universities usually call their students to different locations for final exams.

The COVID-19 pandemic has also led to significant changes in assessment in higher education worldwide [7], requiring all assessment to be conducted through online methods. Although many supervisors are used to working in online environments,

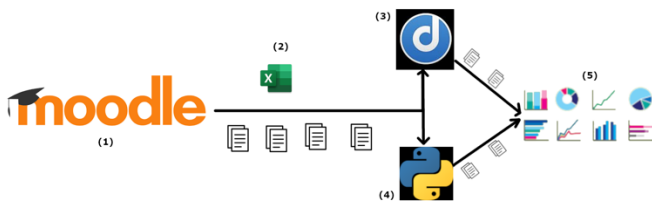


Figure 1: Process followed to analyze the VLE learning records

they are reluctant to assess their students in these environments because students cheating are difficult to detect [17].

Some institutions have provided e-proctoring tools for their teaching staff to support these exams. E-proctoring tools ease the monitoring of the assessment process at a distance through telematic resources [11]. Although these tools are growing in trust for online learning, most face-to-face higher education institutions had to cope with online assessment in the context of the pandemic without such tools, due to technical, economical or legal limitations.

A different approach would be to use evidence-based assessment and learning analytics [9]. While working in the VLE, students leave evidence of their work in the records of the environment [1]. The assessment process involves the analysis and revision of that evidence. In previous work, learning records have been successfully used as evidence to assess students' performance in various generic competences from their work in VLEs [4, 5, 8], wikis [6, 14] and virtual worlds [2].

3 PROPOSAL

This paper analyses the records of the VLE to detect evidence of cheating when students take an online exam.

To analyze the learning records, the process shown in Figure 1 is followed: the records are obtained from the institutional Moodle VLE (1). These records are formatted via a spreadsheet (2) and are processed by the Disco tool (3) and Py-Cheat program (4). From the output of these systems, different reports and graphics are obtained to represent the information stored (5).

The information that known about the exams taken by a student through the VLE is the following:

- Start time
- Time of completion
- Grade
- IP address from which he took the exam
- Time at which the student answers each problem
- Resources accessed during the examination

3.1 Disco

Disco is a process mining tool that manages large event logs and complex models so their conversion and filtering are made easy. Disco takes a comma-separated values (CSV) file and creates a state machine that shows the evolution of the activity it stores. In the case study introduced, these are the resources that a student goes through during his/her connection to the VLE [12].

This paper uses Disco to detect whether students access different resources while taking the exam or whether they simply answer the problems. This way, it is intended to address points 5 and 6 of the previous list.

3.2 Py-Cheat

Py-Cheat is a Python tool developed to detect groups of students who may have collaborated to take the exam. For this purpose, the following variables are taken into account:

- Start time
- Time of completion
- Grade
- Student's class group

The tool returns lists of students who took the exam sequentially, i.e. a second student does not start until the first one finishes. To determine the groups, it is based on the features:

- The second student takes less time than a previous to take the exam
- The second student obtains a similar or higher grade than the previous one
- They are from the same class group.

This tool only makes sense when the exam is set up asynchronously, like the one analyzed in the case study. Py-Cheat is licensed under the GNU General Public License v3.0 and is available in Github¹.

4 CASE STUDY

This research took place in "Database", a second year compulsory course in the Degree in Computer Science and Engineering of the University of Cadiz, Spain. The course was designed following a blended learning methodology, using the institutional Moodle-based VLE. This way, when the face-to-face sessions were not allowed, the switch to online learning was not especially complex. Anyway, although the course could be followed in an online format from the first day of confinement, student participation was lower than in the face-to-face previous period of the course. The course had 123 students enrolled, each one attending a group for theoretical lessons and a lab group. For the lab sessions, they were divided into 6 groups of up to 22 members. When the semester started, students could freely join the group that best fitted their schedule until the group was full. The groups are labelled in this paper as G1, G2, G3, G4, G5 and G6.

This study takes place in the first lab exam of the course, in which 103 students participated. This exam consisted of a multiple choice test of 10 problems. For each problem, 4 possible answers were given, and only one was correct. The 10 problems were drawn randomly from a pool of 100 categorized problems. Each correct answer added up to 1 point, while each incorrect problem subtracted 0.25 points. The examination had the following characteristics:

- 10 problems presented in sequence (no going back).
- Once started, the exam cannot be paused.
- Asynchronous exam: 25-minute exam available for 3 hours: from 11 to 14.

¹<https://github.com/abalderas/Py-Cheat>

Table 1: Summary of marks for the last four years.

Marks	2016-17	2017-18	2018-19	2019-20
A's	0 %	0 %	3 %	11 %
B's	9 %	2 %	12 %	40 %
C's	32 %	27 %	40 %	26 %
D's	59 %	71 %	45 %	23 %

The anonymized data collected from the examination can be consulted also in Github². It should be noted that, according to the policies established by the institution, some had an additional 25% of time to compensate for their disabilities.

5 RESULTS

The analysis of the results is divided into different subsections depending on the source of the findings.

5.1 Exams results

There is a significant increase in the exam grades compared to previous editions of the course, especially in terms of high grades. Table 1 Summary of marks for the last four years. shows that in previous academic years the failing students were between 45% and 59%, while in this exam, only 23% of the students failed. Moreover, between B's and A's there are 51% of students, whereas in previous years they were between 2 and 15%.

A priori this result should not be expected. When the course switched to online format, student participation was lower than in previous courses. In addition, another evidence to take into account, and that contradicts the results of the exam, was the submission of assignments during the course. During lab sessions, but face-to-face and online, students had to perform weekly assignments on an online server. In the previous year, the average student's grades on the assignments were 7.66 out of 10. For the 2019-20 academic course the average of the course assignments decreased to 3.56 out of 10.

5.2 Access to VLE resources during the exam

Using the Disco tool, supervisors were able to review students' behavior in the VLE during the exam. Unfortunately, the use of this tool did not help supervisors detect evidence of cheating. After processing the records, this tool returned a set of states (activities) that students went through while taking the exam. Figure 2 shows this set of states. Although the VLE presented a lot of activity from the beginning of the exam, the students hardly moved among the exam, the forum (where there was a message with instructions for the exam) and the grade book (to check their grades).

The supervisors observed that many students did not start the exam at the stipulated time, but they were connected to the VLE. This could be evidence that they did not start because they did not want to, but not because there was any problem with the VLE or their Internet connection (these probably were the reasons why they claimed that the exams were asynchronous).

However, once a student started the exam, there was little or no navigation to other resources on the VLE. This means that, if there was any collaboration among students or any kind of cheating, it was with some tool external to the VLE.

5.3 Exam timeline

Figure 3 shows that the participation of the students in the exam was evenly distributed during the time it was activated, from 11:00 to 14:00. It is noteworthy that when there is a synchronous exam at a given time, all students usually take the exam on time without any problem. In this case, only 15 students started the exam between 11:00 and 11:05, although the VLE log recorded activity by 55 students between 10:50 and 11:05.

One of the reasons why supervisors do not usually prefer exams to be taken asynchronously is because students can take them collaboratively more easily, i.e. several students organize themselves to take the exam. To analyze if this could happen the Py-Cheat tool was developed. This tool receives the exam records from the VLE and returns different relationships of students who were suspected to take the exam as a team. Table 2 shows a group of students who took the exam sequentially, one student did not start until the previous one had finished. Figures show that, as the group of students progresses, they take less time and get similar or higher grades. These two values are combined to obtain the grade point per minute spent in the exam (*grad/min* column), a value that increases as students take the exam later.

Table 3 shows a group of students in the same situation as seen in table 2, with the additional detail that all the students belonged to the same practical group (G2). It should be noted that the time of the exams ranges from 27 or 25 minutes for the first student to 12 or 10 minutes for the last one in a series of 6 students, shown in tables 2 and 3 respectively. It is especially noteworthy that the last student just needed 1 minute per problem to obtain 8.75 points, that is, 9 right answers and only 1 wrong one. In that period of time it is very difficult to answer the exam with such success, even for the supervisor who created the problems. It should also be noted that the high grades of the students who started their exams just at 11:00 required 22 or more minutes

Finally, below a graph represents sequences of students taking the exam (Figure 4). By using the Py-Cheat tool possible sequences of 6 or more students taking the exam in a row are obtained. The students' grades are represented by the width of the arrow, while the color represents students' paths. Some of them overlap, for example, user043 is reached from user083 and from user057. This may be evidence that both students were waiting for the previous one to finish before performing.

²https://github.com/abalderas/Py-Cheat/raw/master/data/anonymized_data.zip

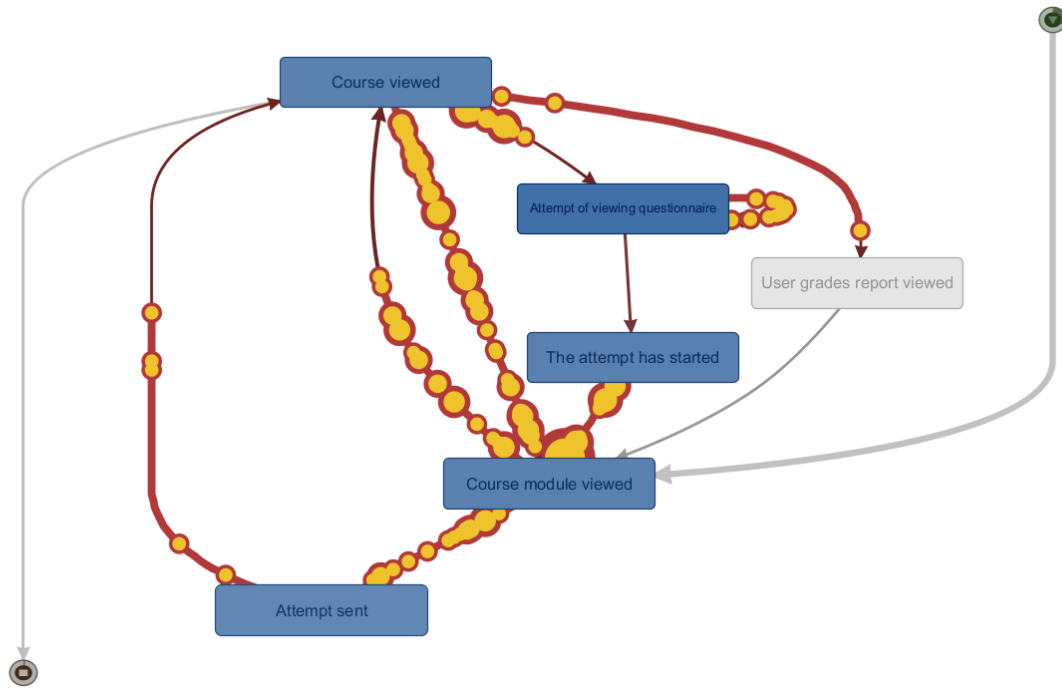


Figure 2: States that students passed through while taking the exam. The arrows represent the students' movement from one resource to another. The students' responses to the questions are represented in the state "Attempt of viewing questionnaire" loop.

Table 2: Group of students taking the exam sequentially

Student	Group	Start	Finish	Minutes	Grade	Grad/min
User029	G2	11:00	11:27	27	7.50	0.27
User116	G6	11:32	11:53	21	8.75	0.41
User042	G3	11:55	12:15	20	7.50	0.37
User028	G1	12:18	12:34	16	6.25	0.39
User038	G3	12:39	12:53	14	10.00	0.71
User054	G2	12:54	13:06	12	8.75	0.72
user061	G3	13:07	13:17	10	10.00	1.00

Table 3: Students of group G2 taking the exam sequentially

Student	Group	Start	Finish	Minutes	Grade	Grad/min
User010	G2	11:41	12:06	25	6.75	0.27
User018	G2	12:08	12:32	24	7.50	0.31
User045	G2	12:32	12:55	23	6.25	0.27
User057	G2	12:58	13:18	20	8.75	0.43
User043	G2	13:19	13:31	12	8.75	0.73
User052	G2	13:31	13:41	10	8.75	0.87

6 DISCUSSION

The results of the analysis suggest that some students were likely to develop cheating behaviors during the exam. The difference among the grades of this year's edition of the course, with respect

to the previous ones, already gave clues that this could be the case. Evidence suggests that students have collaborated using tools outside the VLE, have made the exams in groups, so they have been exploring the answers to the problems among a few. Using the Py-Cheat tool, at least 21 students have been detected who have taken the exam sequentially with up to 6 turns. This would explain why students who take the exam later in the detected series not only get higher grades, but also complete the exam in much less time. It should be noted, for example, that the last student to start the exam (at 13:39), took only 8 minutes to get a 7.5 out of 10 points. This implies spending just an average of 48 seconds on each problem, having 8 correct answers and 2 incorrect ones.

Online exams have proven to be useful for supervisors, as they provide exam settings with many options: categorization of problems and randomization in their presentation, automatic correction

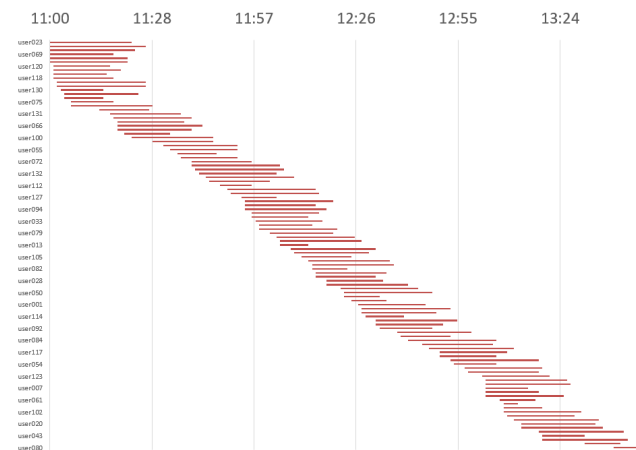


Figure 3: Distribution of students taking the exam

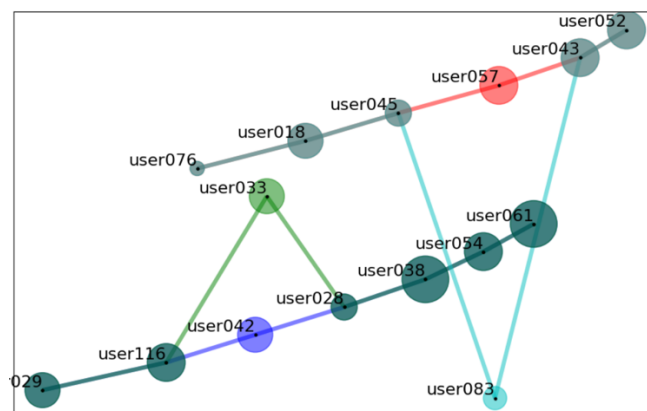


Figure 4: Sequences of students taking the exam, improving their grade from the previous one and taking it in less time

and immediate feedback to students. However, after the face-to-face context, where students attend class together and fellowship is favored, it is difficult to resist the temptation not to collaborate in taking an exam without risk of being detected.

This paper does not intend to accuse students of being cheaters. There are many students who took their exams honestly, as the rules dictate. However, in order to ensure that students achieve the necessary skills and learning outcomes for a course, it would be recommended to be stricter in the configuration of the exams or to explore alternative methods of online assessment, for example, an assessment based on the evidence obtained.

7 CONCLUSIONS

The COVID-19 pandemic forced universities to switch face-to-face teaching into online overnight. Student assessment presented challenges for teaching staff, as they not only had to be concerned in adapting to the new circumstances, but also had concerns that students, without face-to-face supervision, could cheat on their exams.

This concern led to some of them to set up extremely difficult exams that motivated student protests in some cases.

This paper presents a case study in which an exam is prepared following the main requests of the students: that the exam was asynchronous and had a realistic amount of time to answer the problems. Then, by using existing learning analytics tools and a tool ad-hoc developed, supervisors detected evidence that a certain number of students took advantage of these circumstances to cheat.

As future work, this paper proposes to implement an evidence-based assessment for this type of course. Not only because the circumstances can be repeated, but also because a face-to-face exam is susceptible to other types of circumstances that are negative for the student him/herself, such as stress, nervousness or anxiety [18]. For this reason, supervisors should organize their courses so that the student's work is recorded on the VLE. In particular, a focus on the improvement and integration of the developed tool, with the database server and the VLE to favor an evidence-based assessment could be feasible.

ACKNOWLEDGMENTS

Thanks to Professor Manuel Palomo-Duarte, as coordinator of the course in which this research has been conducted, for his collaboration in the development of this work. This work has been developed in the VISAIGLE project, funded by the Spanish National Research Agency (AEI) with ERDF funds under grant ref. TIN2017-85797-R.

REFERENCES

- [1] Adaptación de la evaluación presencial a evaluación online: 2020. <https://innovacioneducativa.wordpress.com/2020/04/15/adaptacion-de-la-evaluacion-presencial-a-evaluacion-online/>. Accessed: 2020-09-29.
- [2] Balderas, A., Berns, A., Palomo-Duarte, M., Dodero, J.M. and Ruiz-Rube, I. 2017. Retrieving objective indicators from student logs in virtual worlds. *Journal of Information Technology Research*. 10, 3 (2017). DOI:<https://doi.org/10.4018/JITR.2017070105>.
- [3] Balderas, A., Capiluppi, A., Palomo-Duarte, M., Malizia, A. and Dodero, J.M. 2019. Reducing Procrastination while Improving Performance: A Wiki-powered Experiment with Students. *Proceedings of the 15th International Symposium on Open Collaboration - OpenSym '19* (New York, New York, USA, 2019), 1–10.
- [4] Balderas, A., de-la-Fuente-Valentin, L., Ortega-Gomez, M., Dodero, J.M. and Burgos, D. 2018. Learning Management Systems Activity Records for Students' Assessment of Generic Skills. *IEEE Access*. (2018). DOI:<https://doi.org/10.1109/ACCESS.2018.2816987>.
- [5] Balderas, A., Dodero, J.M., Palomo-Duarte, M. and Ruiz-Rube, I. 2015. A Domain Specific Language for Online Learning Competence Assessments. *International Journal of Engineering Education - Special issue on Innovative Methods of Teaching Engineering*. 31, 3 (2015), 851–862.
- [6] Balderas, A., Palomo-Duarte, M., Dodero, J.M., Ibarra-Sáiz, M.S. and Rodríguez-Gómez, G. 2018. Scalable authentic assessment of collaborative work assignments in wikis. *International Journal of Educational Technology in Higher Education*. 15, 1 (Dec. 2018), 40. DOI:<https://doi.org/10.1186/s41239-018-0122-1>.
- [7] Bao, W. 2020. COVID-19 and online teaching in higher education: A case study of Peking University. *Human Behavior and Emerging Technologies*. 2, (2020), 113–115. DOI:<https://doi.org/10.1002/hbe2.191>.
- [8] Fidalgo-Blanco, Á., Sein-Echaluce, M.L., García-Peñalvo, F.J. and Conde, M.Á. 2015. Using Learning Analytics to improve teamwork assessment. *Computers in Human Behavior*. 47, (2015), 149–156.
- [9] García-Peñalvo, F.J. 2020. Learning Analytics as a Breakthrough in Educational Improvement. *Radical Solutions and Learning Analytics*. D. Burgos, ed. Springer, Singapore. 1–15.
- [10] García-Peñalvo, F.J., Corell, A., Abella-García, V. and Grande, M. 2020. Online Assessment in Higher Education in the Time of COVID-19. *Education in the Knowledge Society*. 21, 12 (2020). DOI:<https://doi.org/10.14201/eks.23013>.
- [11] González-González, C.S., Infante-Moro, A. and Infante-Moro, J.C. 2020. Implementation of E-Proctoring in Online Teaching: A Study about Motivational Factors. *Sustainability*. 12, 8 (Apr. 2020), 3488. DOI:<https://doi.org/10.3390/su12083488>.
- [12] Günther, C.W. and Rozinat, A. 2012. Disco: Discover Your Processes. *Demonstration Track of the 10th International Conference on Business Process Management*

- (BPM 2012) (2012), 40–44.
- [13] López, J.M. 2020. El alumnado de la UCA estalla y denuncia públicamente “el caos” de las últimas semanas. Portal de Cádiz.
- [14] Palomo-Duarte, M., Doderó, J.M., García-Domínguez, A., Neira-Ayuso, P., Sales-Montes, N., Medina-Bulo, I., Palomo-Lozano, F., Castro-Cabrera, C., Rodríguez-Posada, E.J. and Balderas, A. 2013. Scalability of assessments of wiki-based learning experiences in higher education. *Computers in Human Behavior*. (Aug. 2013). DOI:<https://doi.org/10.1016/j.chb.2013.07.033>.
- [15] Rowe, N.C. 2004. Cheating in Online Student Assessment: Beyond Plagiarism. *Online Journal of Distance Learning Administration*. 7, 2 (2004).
- [16] Ruiperez-Valiente, J.A., Muñoz-Merino, P.J., Alexandron, G. and Pritchard, D.E. 2019. Using Machine Learning to Detect ‘Multiple-Account’ Cheating and Analyze the Influence of Student and Problem Features. *IEEE Transactions on Learning Technologies*. 12, 1 (Jan. 2019), 112–122. DOI:<https://doi.org/10.1109/TLT.2017.2784420>.
- [17] Sariasih, W. and Tisnawijaya, C. 2019. Academic Dishonesty: How Students Do the Learning Assessment and Project. *Proceedings of the Sixth of International Conference on English Language and Teaching (ICOELT 2018)* (Paris, France, 2019).
- [18] Trifoni, A. and Shahini, M. 2011. How does exam anxiety affect the performance of university students. *Mediterranean Journal of Social Sciences*. 2, 2 (2011), 93–100.