

# Analytics of the Digital Behavior of Russian First-year University Students: Case Study\*

Ekaterina Terbusheva<sup>1</sup>  
terbushevae@gmail.ru

Xenia Piotrowska<sup>1</sup>  
krp62@mail.ru

Svetlana Kalmykova<sup>2</sup>  
kalmykovas@mail.ru

<sup>1</sup> Herzen State Pedagogical University of Russia,  
<sup>2</sup> Peter the Great St.Petersburg Polytechnic University  
St. Petersburg, Russian Federation

## Abstract

The goal of research is to determine whether modern analytical tools for Moodle courses data can be useful for ordinary educators. A considerable amount of literature is examined. Possible analytical tools were classified by its suitability from the teachers' point of view. Presented case study demonstrates the potential of analytical tools can be used by teachers. By authors' approach, it is shown that even simplest analytic tools allow to learn a lot about the online activity of students in the online course and get interesting results.

**Keywords:** *learning management systems, digital behavior, data mining, educational data mining, learning analytics, Moodle data analysis, analytical tools, student's activities, open educational platforms*

## Introduction

Today, the main features of modern students' requests are becoming participants in the organization of their training, creating an individual educational route, personification of the educational process, providing opportunities for initiative, taking into account the social mechanisms of interaction: competition, cooperation, mutual learning and assessment, as well as building a creative space for the student.

Modern students' activity is organizing by the digital educational environment as learning units, that should be provided to the teacher at the same time in a convenient, compact and demonstrative forms, such as mind maps, graphs and diagrams that should reflect the relationships and patterns between identified learning activities. Huge amount of information about students "digital traces" are storing in open education platforms ([Coursera](#), [edX](#), [NPTEL](#), [FutureLearn](#), [Open Education](#), [Universarium](#), [Lectorium](#), etc.) and learning management systems ([Moodle](#), [iSpring](#), [Mirapolis](#), [ShareKnowledge](#), [TeachingBase](#), [WebTutor](#), etc.).

Nowadays an actual task become studying the capabilities of the educational analytics to build a high-quality educational process in a digital learning environment. The information contained in system log files is very useful in making predictions of learning processes. Mostly it should help in marking about the success of training (*predictive analytics*), in describing what has it happened (*descriptive analytics*) or in making conclusions about why have it happened (*diagnostic analytics*).

---

\*Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The main analytics' aim is to get informed conclusions about what needs to be done to improve the educational process.

LMS Moodle system is one of the most powerful, popular and, that's why, widely used systems in the world for organizing distance learning scaffolding for courses. In the Russian Federation, as well as in some developed countries, the Moodle platform is actively used in official educational institutions (Table 1). The popularity of Moodle is because it is free and has detailed technical documentation, widest possibilities of customizing the interface and functions, as well as a flexible system of statistics and reports in comparison with other free LMS (such as [ATutor](#), [Ilias](#), [Diskurs](#)).

Table 1: The top countries on the use of LMS Moodle

Country	Number of sites based on Moodle	Percentage of educational institutions sites (%)
USA	11075	23,7
Spain	10237	26,1
Mexico	8763	23,9
Germany	6165	19
Brazil	5956	35,1
France	5586	41
Russia	4573	31,3
Italy	4088	26
Great Britain	4057	24,6
India	3392	24,7

Further in the paper, the world experience of data analysis from Moodle is given in Section 1.2, a different groups of tools for Moodle data analytics are identified in Section 1.3, and a case study on the analytics of the digital behaviour of the first-year students in Moodle-course "History" is presented in Section 2. An experimental research has made in the Herzen State University of Russia in collaboration with Peter the Great St.Petersburg Polytechnic University in 2019. The foregoing indicates the relevance of acquiring and transferring experience in analyzing the data accumulated in the LMS Moodle.

## 1 Materials and methods

### 1.1 Materials of investigations

We conducted some experimental study on the analytics of data collected in the distance course "History" in LMS Moodle in the period of investigations from February 1 to June 30, 2019 (the 2nd semester of the 2018-2019 academic year). Number of respondents was more than 3 thousand people. The course is intended for all the 1st year students and lasts for one semester. The course contained 8 video lectures, with presentation and self-control test for each one, description of tasks for seminars, 3 control tests and 1 final test.

### 1.2 Related works

Educational analytics implies the application of knowledge from various fields, such as statistics, recommendation systems, data mining, psychometry, learning technologies, etc. The main purpose it applied is to analyze data from the educational environment:

- in [Nilgün, 2018] and [Noskova et al., 2018], by using the correlation and cluster analysis, there were studied the strategies of educational activities of students and security in social networks,
- in [Almutairi et al., 2019] and [Nesterov et al., 2019] predicting students' academic performance and main behavioural features of learning outcomes were evaluated from distance courses by using data mining methods,
- in [Spatiotis et al., 2019] and [Fomin et al., 2019], text-mining technologies were used to evaluate educational information resources.

Nowadays special actively developing area is highlighted by science - the educational data mining / EDM, which is aimed at developing data mining methods for making decisions in the field of education. The possibilities of EDM are broadly considered in the science community [Bakhshinategh et al., 2018]. Approaches to training educators in data analysis are considered in [Piotrowska & Terbusheva, 2019]. As we are working with Moodle-courses we identified the main areas of the research in the field of EDM conducted in the scientific community directly with the data of the Moodle environment:

- study of the relationship between the activities of students in Moodle and their successes [Kadoić et al., 2018], [Mogus et al.], [Stiller et al., 2018];
- identification of behavioral strategies in online learning [Akcapinar et al., 2015], [Bogarín et al., 2015];
- analysis of gender differences in education in LMS [Kadoić et al., 2018];
- visualization of the extracted data [Rybanov et al., 2013], [Aguilar et al., 2008];
- instructions for preprocessing Moodle data and examples of the application of mining techniques (visualization, clustering, classification, associative rules, etc.) to such data [Romero et al., 2008].

In these papers, the analysis was carried out mainly on the data from questionnaires and a system log files. The following tools were used for analysis: Excel, Statistica (for calculating statistical indicators and correlation analysis), RapidMiner, Weka, Deductor (for clustering and other data mining methods), self-developed tools that include Moodle data processing and some analysis and visualization capabilities). Many works use several kinds of different tools and methods to perform the study [Lahbi & Sabbane, 2020].

A more complete overview of the possible tools for Moodle data analysis is given in the next section.

### 1.3 Tools for Moodle data analysis

**Standard Moodle tools.** These tools are reports (competency breakdown report, logs, participation report, activity report, grade report, statistics) and simple analytical models (such as students at risk of dropping out, upcoming activities due) that are included in the basic version of Moodle and available to all teachers. Standard Moodle tools provide some statistics on student activity in the course (for example, students viewing various elements of the course). Reports are easy to get, but they often don't have sufficient functionality and visibility to make decisions. Detailed information on the capabilities of these tools is available in the Moodle documentation [Moodle Reports].

Table 2: System of learning analytics tools for Moodle data

Tools for Moodle data analysis		Learning Analytics Tools Features (level on average)					
		Functionality	Statistics	Machine Learning & Data Mining	Free	Availability / Easy to Install for the teacher	Simplicity of use for the teacher
Standard Moodle Tools	<b>Reports:</b> Logs, Participation report, Activity report, Grade report, Statistics, Competency breakdown report	low	middle	low	high	high	high
	<b>Moodle analytics models:</b> students at risk of dropping out, upcoming activities due, no teaching	low	low	low	high	high	middle
Moodle Plugins	<b>Moodle plugins:</b> SmartKlass Learning, Analytics Moodle, Heatmap, IntelliBoard	middle	high	middle	high	low	middle
Specialized External Tools for Moodle data	<b>External tools for Moodle data:</b> KEATS Analytics, MoodleMiner, SNAPP, LAPA, Loop, ViMoodle	middle	middle	middle	middle	low	middle
Multifunctional External Tools	<b>Computing environments:</b> Statistica, SPSS Statistics, Matlab, Octave, MS Excel	high	high	middle	middle	middle	middle
	<b>Data Mining programs:</b> Orange, Knime, RapidMiner, Weka, Deductor, KEEL	high	middle	high	high	middle	middle
	<b>Programming languages and development environments:</b> R (RStudio), Python (PyCharm, Spyder), Java (Eclipse, NetBeans)	high	high	high	high	low	low
	<b>Database management systems:</b> Oracle, MySQL, Microsoft SQL Server, PostgreSQL	high	high	middle	middle	low	low
	<b>Tools for web-site analysis:</b> Webalizer, AWStats	middle	middle	low	middle	low	low

**Moodle plugins.** Such plugins are additional functional extensions of Moodle, which can be installed only by the site administrator. There are paid solutions, but at the same time a large number of free solutions are available in the Moodle plugins directory. Examples of various analytic plugins: for additional statistics (IntelliBoard, Courses Usage Statistics), for visualization (Heatmap, Analytics graphs, Daily usage), for pattern discovery or student at risk predictions (SmartKlass, Inspire) and others. But individual teachers are not able to select, install and research plugins, which has great limitations in their use and usefulness.

**Specialized external tools.** These tools are designed specifically for analyzing data from Moodle courses. Despite the number of descriptions of such tools in the literature ([Akçapınar et al., 2019], [Aguilar et al., 2008], [Corrin et al.], etc.), it is difficult to find, install and use these tools for teachers on their own. Some of the free tools (MoodleMiner, Moodler, MoodleLogAnalyse) are available as source code in the github repository, but a teacher without IT skills won't be able to use them. The exception in the current category is KEATS Analytics tool [Konstantinidis et al., 2013], that can be freely downloaded from a Google Drive [KEATS analytics]. Only Excel is required to use this tool.

**Multifunctional external tools.** These are various data analysis environments that provide wide functionality and capabilities for high-quality and multi-aspect data analysis. But to use them, it is often necessary to have specialized knowledge in the field of IT and Math (statistical methods, data mining, programming languages, etc.). In this category we distinguish 5 subgroups: computing environments, data mining programs, development environments, database management systems and web-site analytic tools. Among these tools, there are freely distributed ones with a fairly simple interface, for example, the Weka program that implements data mining methods.

In Table 2, we tried to compare the different groups of tools of data analysis by various parameters from the point of view of the teacher's convenience. All these solutions fulfill their role of big educational data transform into some more understandable information. However, they are limited in terms of customization and do not provide sufficient resources for various situations. Our approach described below in this paper expands opportunities through an open solution that enables the development of resources to meet needs identified by specific educational contexts.

## 2 Description of authors' approach

To show how ordinary educator can study students digital behavior with the help of simplest tools we provide the further case study. The general scheme for solving the paper problem is shown in figure 1.

It's possible to understand the intensity of the course usage using the standard Moodle report "Activity report". To get data on views of course elements for a certain time, one can set a filter. In this purpose we set the period from February 1 to June 30, 2019, that corresponds to the spring semester of the 2018-2019 academic year. Already at this stage of the study, a significant shortcoming in the design of the course regarding further analytics was revealed: the activity report does not contain data on viewing video lectures, because the videos were not included as Moodle elements of the course.

Visualization (using MS Excel program) of the activity report information allows us to see that the educational activity of students decreases throughout the semester (Fig. 2). The number of users viewing presentations for video lectures and doing tests for self-control decreased by the end of the semester by 3,3-3,8 times (by 70-74%). At the same time, control tests are performed by students. The number of viewers of control tests also decreases, but not as significant as the number of viewers of the rest of the digital content. The third test out of four was viewed by 7% less students than the first test, and the final test – by 24% less students than the first.

Analysis of the time activities on the course of 36 students of one group with KEATS Analytics tool [Konstantinidis et al., 2013] showed that mainly students entered the course in the middle of the week; the time from Tuesday to Friday was the most active (Fig. 3). On Saturday and Monday, there was a decline in activity, on Sunday it was average. At the same time, students most often worked in

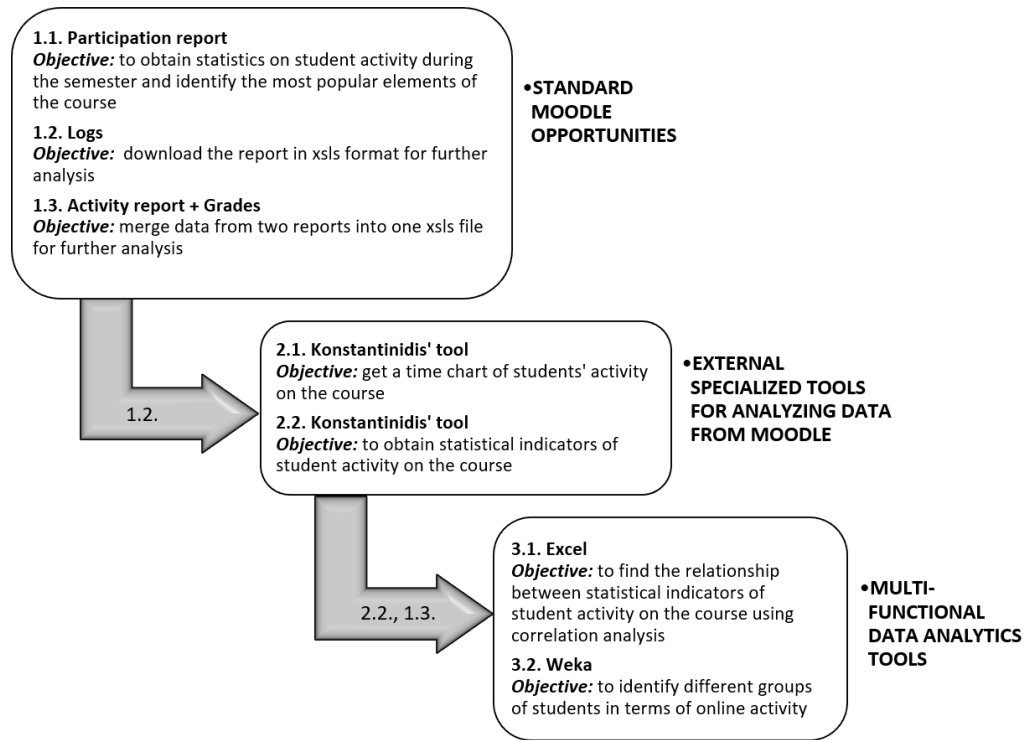


Figure 1:

the course during the morning hours from 8 to 12 and in the evening between 17 and 23 hours. This information can be used in planning educational activities. For example, it is better to plan online testing or discussion during hours of high students' activity. At this time most students are likely to be able to take part in online activities.

We also carried out a correlation analysis between some statistical indicators of activity of this group of students on the course and their final scores (Table 3). To do this, we first combined the values of students' activity indicators obtained in KEATS Analytics tool with their total score in one file and then we used Excel to calculate the correlations. Among such indicators: number of total views of the course (i.e. any its element); number of various actions; the number of different course pages that have been viewed; the number of different dates a student was active on the course. The results of the analysis allowed us to conclude that the final score depends more on the number of different actions on the course (correlation coefficient 0.77) than on the number of total course views (0.51), the number of different pages visited (0.54) or the number of active days (0.58). For further

Table 3: Correlation between student activity indicators in an online course

	Total views	Actions	Pages	Dates
Actions	0.68	-	-	-
Pages	0.86	0.65	-	-
Dates	0.86	0.73	0.67	-
Final score	0.51	0.77	0.54	0.58

analysis, we combined information on the activities and performance of student by manually copying the data from the "Participation report" and "Grade Report" reports into the Excel spreadsheet. Since

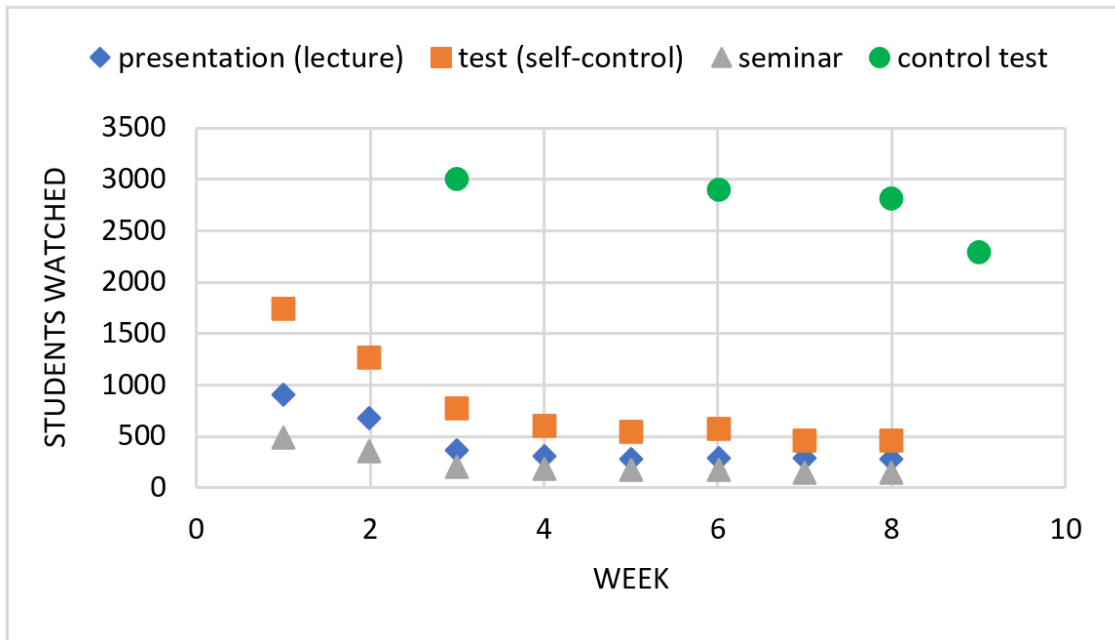


Figure 2:

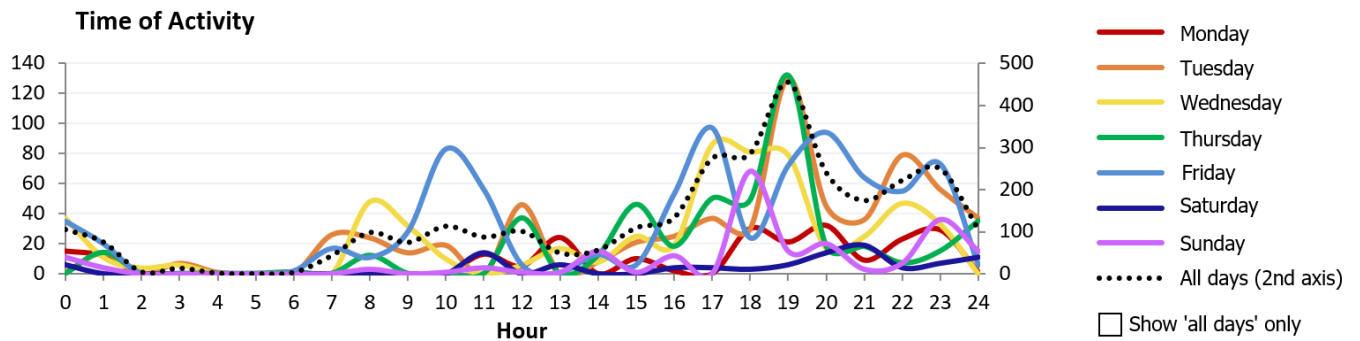


Figure 3:

the report “Participation in the course” is built in Moodle for each course element separately, it is necessary to make this report as many times as there are elements in the course for which we would like to take into account the information. Data from all built reports were copied to one spreadsheet. Further, data on students’ grades from another report were added to the same table.

Manually collecting data for many students is laborious, so it is important to develop specialized software that automate this process.

In order to identify groups of students based on the collected data, we performed a cluster analysis in the Weka program. In order to perform clustering in Weka, it's only needed to load the data on the Preprocess tab and select the clustering algorithm on the Cluster tab. Since we hadn't elaborated an assumption about the number of clusters, we chose an SelfOrganizingMap algorithm that didn't not require setting the number of clusters for execution. Fragment of clusterer's output (for 6 of 36 attributes) is shown in fig. 4.

As you can see, the algorithm divided the students into 3 clusters. The analysis of these clusters allowed to distinguish three groups of students.

1. Students who did not access the course (or accessed only a few times) and did not perform control tests (cluster 3).
2. Students who performed only control tests and showed no other activity in the course (cluster 1).
3. Students who performed some educational activities (such as viewing presentations for video lectures, passing tests after lectures for self-control) and did control tests (cluster 2).

We also tried other algorithms, they showed not identical, but similar results. This requires the development of recommendations for working with each group, and it is additionally necessary to study students' motivation and approaches to teaching students in each group.

In the study group, the first group includes 5 out of 36 students. To reduce the number of such students, you can send them reminders about the course and the upcoming first assessment, if an inactivity was detected. The Moodle functionality allows to select these students in the list and send them messages. The opportunity to send messages is available in the report "Participation in the course". But to identify the student's prolonged inactivity, it needs to look at the reports for a number of elements of the course, that is difficult in cases of large groups and high employment of teachers.

As for the students of the second group, they do not view the course materials, they only pass

	cluster 1	cluster 2	cluster 3
<b>lection6</b>			
value	0	0.3638	0
min	0	0	0
max	0	7	0
mean	0	0.3636	0
std. dev.	0	1.4975	0
<b>test6</b>			
value	0	0.0863	0.5771
min	0	0	0
max	0	1	4
mean	0	0.0909	0.5714
std. dev.	0	0.2942	1.5119
<b>seminar6</b>			
value	0	0	0
min	0	0	0
max	0	0	0
mean	0	0	0
std. dev.	0	0	0
<b>controlTest2</b>			
value	4.8039	6.4781	0
min	2	3	0
max	10	16	0
mean	4.8333	6.5	0
std. dev.	3.0605	4.206	0
<b>scoreTest2</b>			
value	12.7064	19.0408	0
min	0	17	0
max	20	20	0
mean	12.8333	19.0455	0
std. dev.	9.9683	0.8985	0
<b>lection7</b>			
value	0	0.0499	0
min	0	0	0
max	0	1	0
mean	0	0.0455	0
std. dev.	0	0.2132	0
<b>test7</b>			
value	0	0.0416	0
min	0	0	0

Figure 4:

control tests. In the corresponding cluster 1 on Fig. 4 min value for the scoreTest2 is zero. It means that this group also includes students who missed some of the control tests and by default, they have 0 points for it. Other students of the group passed the tests quite successfully, no worse than students of the third active group. But we do not have information on viewing video lectures, since these videos were not included as elements of the course (as opposed to presentations for video lectures, seminars, tests for self-control).



The previous discussion suggests two variants:

- students of the second group did not watch the video lectures. In this case, several questions are arising: “How was the learning material studied?”, “Why was it learned outside the LMS?”
- students of the second group watched the video lectures. In this case, we could conclude that video lectures watching is the main criterion for a successful assessment.

Further research is needed to clarify this situation.

## Conclusions

Our review of research in the field of Moodle data analysis and possible tools for this domain shows that there is no universal tool that provides extensive functionality. To understand the learning and draw more meaningful conclusions, researchers are forced to use several tools process. It allows the user to conduct a high-quality experiment on the analysis of data accumulated in LMS Moodle. In addition, research is usually carried out by IT professionals, because it is too complex for ordinary non-IT educators.

But our experimental study demonstrates that even the use of the simplest tools, which teachers can understand, allows to learn much about the online activity of students in the online course and get interesting results.

The usefulness of learning analytics and educational data mining methods has been repeatedly shown. Further, it is necessary to ensure the widespread use of these opportunities by teachers. For this purpose, it is necessary on the one hand to train ordinary teachers to work with various methods and tools for data analysis [Piotrowska & Terbusheva, 2019], and on the other hand - to provide teachers with multifunctional tools with a simple interface and clear instructions.

## Acknowledgement

The research was supported by the Russian Science Foundation (RSF), Project “Digitalisation of the high school professional training in the context of education foresight 2035” (No 19-18-00108).

## References

- [Almutairi et al., 2019] Almutairi S., Shaiba H., Bezbradica M. (2019) Predicting Students’ Academic Performance and Main Behavioral Features Using Data Mining Techniques //ICC 2019: International Conference on Computing 2019: Advances in Data Science, Cyber Security and IT Applications, Pp. 245-259.
- [Aguilar et al., 2008] Aguilar D.A., Therón R., García-Peñalvo, F.J. Understanding Educational Relationships in Moodle with ViMoodle. In: Eighth IEEE International Conference on Advanced Learning Technologies, Santander, Cantabria, Spain, July 1st – July 5th, 2008. Pp. 954-956.
- [Akcapınar et al., 2015] Akcapınar G. (2015) Profiling students’ approaches to learning through Moodle logs. //Proceedings of the Multidisciplinary Academic Conference, Prague, 2015.
- [Akçapınar et al., 2019] Akçapınar G., Bayazit A. (2019). MoodleMiner: Data Mining Analysis Tool for Moodle Learning Management System // İlköğretim Online. No. 18. Pp. 406-415. DOI: 10.17051/ilkonline.2019.527645.

- [Bakhshinategh et al., 2018] Bakhshinategh B., Zaiane O. R., ElAtia S., and Ipperciel D. (2018) Educational data mining applications and tasks: A survey of the last 10 years // Education and Information Technologies. Vol. 23. No. 1. Pp. 537–553.
- [Bogarín et al., 2015] Bogarín A., Romero C., Cerezo R. Discovering Students' Navigation Paths in Moodle. In: Proceedings of the 8th International Conference on Educational Data Mining, Madrid, Spain, 26-29 June, 2015. Pp. 556-557.
- [Corrin et al.] Corrin L., Kennedy G., De Barba P., Bakharia A., Lockyer L., Gasevic D., Williams D., Dawson S., and Copeland S. (2015) Loop: A learning analytics tool to provide teachers with useful data visualisations, 2015.
- [Fomin et al., 2019] Fomin V., Osochkin A., Zhuk Y. Frequency and morphological patterns of recognition and thematic classification of essay and full text scientific publications. //CEUR Workshop Proceedings, Vol. 2401 NESinMIS-2019 - Proceedings of the 14th International Conference "New Educational Strategies in Modern Information Space". 2019. URL: [http://ceur-ws.org/Vol-2401/PAPER\\_6.PDF](http://ceur-ws.org/Vol-2401/PAPER_6.PDF)
- [KEATS analytics] KEATS analytics: program for analysis Moodle logs. URL: <https://clck.ru/My8h6>.
- [Kadoić et al., 2018] Kadoić N., Oreški D. Analysis of student behavior and success based on logs in Moodle. In: 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), At Opatija, Croatia, 2018. Pp. 0654–0659.
- [Konstantinidis et al., 2013] Konstantinidis A., Grafton C. (2013) Using Excel Macros to Analyse Moodle Logs 2nd Moodle Research Conference, Sousse, Tunisia October, 4-6, 2013, Pp. 33-39
- [Lahbi & Sabbane, 2020] Zakaria Lahbi Z., Sabbane M. (2020) Learning Analytics Dashboard Model to Supervise a Distance Learning // International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-5, January 2020. Blue Eyes Intelligence Engineering Sciences. DOI: <https://doi.org/10.35940/ijrte.E6675.018520>
- [Moodle Reports] Moodle documentation. Course reports. URL: [https://docs.moodle.org/38/en/Course\\_reports](https://docs.moodle.org/38/en/Course_reports)
- [Mogus et al.] Mogus A. M., Djurdjevic I., and Suvak N., The impact of student activity in a virtual learning environment on their final mark // Active Learning in Higher Education. 2012. Vol. 13. No 3. Pp. 177–189.
- [Nesterov et al., 2019] Nesterov S.A., Smolina E.M. Metody intellektual'nogo analiza dannyh v zadachah ocenki rezul'tatov distancionnogo obucheniya //Sistemnyj analiz v proektirovanii i upravlenii. 2019. No 3. S. 406-412. (In Rus) == Нестеров С.А., Смолина Е.М. Методы интеллектуального анализа данных в задачах оценки результатов дистанционного обучения //Системный анализ в проектировании и управлении. 2019. No 3. С. 406-412.
- [Nilgün, 2018] Nilgün T. (2018) Social Networks as a Learning and Teaching Environment and Security in Social Networks //Journal of Education and Training Studies. Vol. 6, No. 11a. DOI: <https://doi.org/10.11114/jets.v6i11a.3817>
- [Noskova et al., 2018] Noskova T.N., Pavlova T.B., Yakovleva O.V. Study of students educational activity strategies in the social media environment // E-learning. 2018. С. 113-123.

- [Piotrowska & Terbusheva, 2019] Piotrowska X., Terbusheva E. Educational Data Mining for Educational Employees. //CEUR Workshop Proceedings, Vol. 2401 NESinMIS-2019 - Proceedings of the 14th International Conference “New Educational Strategies in Modern Information Space”, 2019. URL:[http://ceur-ws.org/Vol-2401/PAPER\\_4.PDF](http://ceur-ws.org/Vol-2401/PAPER_4.PDF)
- [Romero et al., 2008] Romero C., Ventura S., Garcia E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*. Vol. 51. Pp. 368-384. DOI: <https://doi.org/10.1016/j.compedu.2007.05.016>
- [Rybanov et al., 2013] Rybanov A. A., Serzhantova E. O., Makushkina L. A. Analiticheskaya platforma Deductor kak sredstvo analiza rezul'tatov aktivnosti pol'zovatelej sistemy distancionnogo obucheniya Moodle // *Molodoj uchenyj*. 2013. No 5. S. 173-176. (In Rus) == Рыбанов А. А., Сержантова Е. О., Макушкина Л. А. Аналитическая платформа Deductor как средство анализа результатов активности пользователей системы дистанционного обучения Moodle // *Молодой ученый*. 2013. No 5. С. 173-176.
- [Stiller et al., 2018] Stiller K., Bachmaier R. Identifying learner types in distance training by using study times. In: *Proceedings of the European Distance and E-Learning Network Conference*, Genoa, Italy, June 17-20, 2018. Pp. 78-86.
- [Spatiotis et al., 2019] Spatiotis N., Mporas I., Perikos I., Paraskevas M. (2018) Evaluation of an Educational Training Platform Using Text Mining. In *Proceedings of the 10th Hellenic Conference on Artificial Intelligence (SETN '18)*, July 9–12, 2018, Patras, Greece. Association for Computing Machinery, New York, NY, USA, Article 42, 1–5. DOI:<https://doi.org/10.1145/3200947.3201049>