

Mathematical Modeling Course Creation and Implementation Using Cocalc

V.N. Kozhukhova^{1*}, A.A. Korobetskaya¹ and V.K. Semenychev¹

* Corresponding author: kvn505@yandex.ru

¹ Samara State University of Economics, Samara, Russia

Abstract. Being commonly used worldwide, e-learning courses are only starting to evolve in regional Russian universities. Unlike the most Moodle-using colleagues, the authors created mathematical modeling e-course with not so well-known in Russia learning management system Cocalc. The biggest advantage of Cocalc is the ability to write and compile program code inside, with a lot of programming languages supported. Using statistical methods, the authors assessed the results of the course implementation, revealing that students began to perform tasks faster, learned to program fluently in R and Latex and increased their English level knowledge.

Keywords: Cocalc, learning management system, e-learning, Latex, R, mathematical modeling.

1. Introduction

Due to Russia's digital economy developing, learning process is considered not only as digital data transmission by means of ICT but also as skills acquisition in order to form knowledge concerning object analysis.

Big data processing at mesoeconomics levels requires mass training in qualitative mathematical knowledge. Thus includes predictive modeling and fluent program code writing. Such requirements can be met with the latest learning management systems (LMS).

Whereas in developed countries education institutions have been using learning management systems for years, Russian education system is only at the stage of their mass implementation. Currently, many electronic services are now starting to implement learning management systems. As a rule, they include the exchange of materials between tutors and students, messaging, homework assessments and the test system support.

In the Russian-speaking educational segment, the most popular LMS is undoubtedly Moodle. We have studied query results at elibrary.ru by keywords "LMS" and "Moodle": over the past five years, Russian Science Citation Index has published 1123 articles with the keyword "LMS" in the category "Pedagogy", of which 833 are devoted to Moodle. However, the authors consider Moodle not meeting the requirements and purposes of their mathematical modeling course and suggest implementing LMS Cocalc into educational process.

2. Problem Statement

This article is devoted to current Russian education situation for mathematical and IT disciplines teaching. The authors are engaged in the mathematical modeling courses development based on modern mathematical statistics and econometrics. These disciplines are now fully computerized within the educational systems of developed countries. Students should use the appropriate software for statistical analysis, optimally combining theoretical knowledge of mathematical models with the computer calculations practice. Implementation of such courses in Russia involves extensive use of electronic resources: Office applications, various high-level programming languages based upon the learning management system. Laboratory classes should take place in specialized computer-equipped rooms and require stable high-speed Internet access.

Unfortunately, in most Russian regional universities (thus excluding Moscow and Saint Petersburg) laboratory tasks for statistics disciplines are usually provided using a table processor Microsoft Excel. This choice is not accidental, but much related to the fact that this software is the main (and sometimes the only) thing students face when they start working. Many regional companies' calculations are carried out using this software product. However, modern digital economy considers that data processing requires R or Python, widely used by the world's largest companies.

The authors see the current problem as the need for educational process reorganization using such modern LMS as Cocalc with free access to R, Latex, Python etc.

3. Research Questions

The article aims to address the following issues:

- advantages of LMS Cocalc implementation in educational process;
- handouts and assignment materials provided;

- problems that students may face while learning to work with new course management system;
- tracing the effect of using new LMS for mathematical modeling course teaching.

4. Purpose of the Study

The purpose of this study was to share the authors' e-learning course creation and implementation experience. We aimed to assess the choice of LMS CoCalc as a new learning tool for a typical regional university in Russia. One of the main objectives of this paper was to calculate the grade and time effect of using this e-course.

5. Research Methods

In order to measure the e-course implementation effect, we used test and control students' groups studying with and without LMS Cocalc respectfully. Both groups wrote three exam works. The first was an entering exam assessing the level of mathematical training. The second was a mid-term exam and the third one was final exam. For our course and further experiment we used traditional Russian grading scale: 5 – excellent, 4 – good, 3 – satisfactory, 2 – unsatisfactory.

Control group performed all tasks in a traditional way calculating and then reporting the results using Microsoft Office applications. The test group fulfilled the entrance test in the same way with Microsoft Office, but the second and third exam tasks were fulfilled generating online rnw reports within R and Latex using LMS Cocalc. The pedagogical research usually involves various statistical methods [0-3]. We used the following statistical tests to verify the results of this experiment [4]:

- chi-squared test for categorical data – for testing the hypothesis of no significant difference between grades of two independent groups;
- Mann–Whitney U test – for testing the hypothesis of no significant difference between exams writing time in two independent groups;
- Friedman test – to detect differences in writing time inside each group when fulfilled with and without programming in Cocalc.

6. Findings

6.1 The features and advantages of LMS Cocalc

To meet the current international requirements for the specialists training in the field of statistical calculations, the authors decided to include in their course the study of Latex integrated with R using knitr. Such a “bundle” of Latex with R can be easily installed on a personal computer. However, when it comes to learning management systems (such as the Moodle system), most of them do not support writing code, checking it and compiling it. Latex and R tasks inclusion (and, in general, all tasks for writing program code) into the electronic course assumes the following steps. Student performs the task on his computer and attaches the sources to the learning management system. The tutor saves them, checks on his computer and grades the work in LMS.

More promising is the idea implemented on cocalc.com, which can be used as a LMS specifically for math and IT disciplines. This site allows, like any LMS, to create a course, to exchange materials with students, to set deadlines for the tasks completion, to grade homework, etc.

Right in the cloud it is possible to create projects in R, Python, bash, sage, tex, Octave, Python, Julia, etc. [5]. On the CoCalc server the Linux terminal can be run, and through it students can edit and compile code in many other languages, including Java, C/C++, Perl, Ruby. The created projects can be accessed by several collaborators at the same time. Thus, the student and the tutor have the opportunity to write and correct the program code of the project at the same time, and each user sees the work of his collaborators.

Another feature of CoCalc is the compilation of written code in the cloud [6]. The system allows integrating interactive program code into the text of rnw-document. This code is compiled “on the fly” and the result is a finished report or article in pdf format [7].

All software products are installed on CoCalc servers. Users can get both free and paid access to these servers. Paid access allows increasing the speed of work and getting technical support with a higher priority of answers.

At the same time, free projects cannot download data from the Internet. For example, when used in R, this means that statistical data download “on the fly” will not be available. However, for training tasks it is not usually required. Let's note, for comparison, that the restrictions for the service use of Moodle Cloud are much more serious [8].

The above means that the university computer classes engineers do not need to install special software, but only need to provide the latest version of the browser and high-speed Internet access.

The main advantage of LMS includes the fact that the course interface is intuitive. The chat between students and teacher is adapted to the needs of mathematics and programming. It supports html markup languages, markdown and

Tex formula syntax [9]. The platform allows integration with GitHub. The portal also implements the function of peer grading.

A certain drawback for Russian students is that the portal is not yet russified and is available only in English. CoCalc servers are also running Linux, so it is not possible to work in the cloud with Microsoft Office (so common for Russian students unlike Latex). At the same time, with paid access, it is possible to install any cross-platform applications on CoCalc servers (if there are appropriate licenses), as well as increase the available server resources.

6.2 Course materials creation

Let us consider the experience of creating a course on the CoCalc platform on the example of the discipline “Mathematical modeling”. We have divided all teaching materials into handouts (lectures, slides etc.) and assignments (homework) and invited students to the course by sending e-mail through the system interface. Hangouts are immediately provided to students for review. Assignments are sent with a deadline after which the task is not accepted. Students register on the portal and join the course. After that, they can perform tasks using the rnw source code and make online reports with R. Finished works are sent to the tutors’ account.

At the same time, students can, as in the Moodle system, perform part of the work not in the cloud, but, for example, in the desktop version of Excel or other software and attach the files of the completed work in the personal account. Such integration, as authors think, is very important in transition period when students just are getting used to new LMS.

Firstly, the time-tested method of using LMS is preserved, and, secondly, students get acquainted with an alternative opportunity to work simultaneously with the tutor in the cloud and make scientific calculations using modern programming languages – R, Python and Latex. Writing articles in Latex template is a requirement in many leading scientific journals of the world.

In the mathematical modeling course, assignments are bound to the “Labs” and “R” folders.

The “Labs” folder contains tasks for laboratory work in Excel, which is the first part of the course. For students, in fact, there is no difference from working with well-known Moodle. The second part of the course involves using R and Latex (knitr) to perform several homework tasks:

- basics of work in R;
- linear pair and multiple regression;
- time series analysis;
- non-linear pair regression using genetic algorithms.

Rnw source code listings to these tasks are located in the “R” course folder in the corresponding subfolders. Interested readers may request a link to the e-course by contacting the authors of this article.

Even if students do not submit the work for review, the tutor can push the Collect button to collect all works. It is possible to return work to a student with notes about errors. We also exported grades to a csv file to use this data in our statistical analysis.

6.3 Pedagogical experiment

For an experiment the authors took the second year education students studying the bachelor’s degree “Applied computer science” of two different programs “Applied computer science in electronic economy” and “Electronic services and artificial intelligence”. The number of students involved is 58.

We randomly divided students into two groups of 26 and 32 members. Each group wrote the same mathematical exam entering the course. The grading scale has been discussed in section 6.2. The descriptive statistics for each group is given in Tables 1 and 2.

Table 1. Students’ marks for entering exam

| Grade | Number of Students | |
|------------|-----------------------|-----------------------|
| | 1 st group | 2 nd group |
| 5 | 5 (19,23%) | 5 (15,63%) |
| 4 | 15 (57,69%) | 16 (50,00%) |
| 3 | 6 (23,08%) | 11 (34,38%) |
| 2 | 0 (0,00%) | 0 (0,00%) |
| Sum | 26 | 32 |

Source: compiled by the authors.

Table 2. Descriptive statistics of two students’ groups

| Statistics | 1 st group | 2 nd group |
|------------|-----------------------|-----------------------|
| mean | 3,962 | 3,813 |
| SEM | 0,129 | 0,122 |

| | | |
|-----------------|--------|--------|
| median | 4 | 4 |
| mode | 4 | 4 |
| sd | 0,662 | 0,693 |
| kurtosis | -0,501 | -0,804 |
| skewness | 0,039 | 0,267 |

Source: compiled by the authors.

We performed chi-squared test for categorical data in Table 1 testing the hypothesis of no significant difference between 2 groups:

$$\chi^2 = NM \sum_{i=1}^L \frac{\left(\frac{n_i}{N} - \frac{m_i}{M}\right)^2}{\frac{n_i}{N} + \frac{m_i}{M}}$$

where $N = 26$, $M = 32$, $L = 3$, n_i, m_i are number of students with i -th grade (columns 2 and 3 of Table 1 excluding grade = 2). The test statistic is 0,892 with chi-squared critical value 5,991 at 0,05 significance level with 2 degrees of freedom so the null hypothesis is not rejected.

As a result, we get two groups with similar mathematical knowledge level for the future tests. The 1st group of 26 students is the control group. They studied mathematical modeling course traditionally, without using LMS Cocalc. The 2nd group of 32 students is the test one. This group studied the course working with LMS Cocalc.

As a result of using new LMS and programming languages we haven't expected significant differences in the level of mathematical models knowledge. But we expect significant time reduction of exams writing because R programming and report generation allow calculating and representing the same things faster.

Firstly, we graded mid-term and final exam works with the same scale as the entering one and performed chi-squared test at the same significance level. The test statistics are 0.683 and 2.744 for the control group and the test group respectfully. The null hypothesis of no significant differences in the level of mathematical models knowledge is not rejected.

Secondly, we have measured time of the exams writing for each student (see Table 3 for descriptive statistics). The time limit for each exam was 80 minutes.

Table 3. Descriptive statistics of two students' groups exams time performances

| Statistics | Control group | | | Test group | | |
|-----------------|---------------|----------|--------|------------|----------|--------|
| | Entering | Mid-term | Final | Entering | Mid-term | Final |
| mean | 72,462 | 69,654 | 66,962 | 70,375 | 61,219 | 58,969 |
| SEM | 1,447 | 1,418 | 1,708 | 1,531 | 1,762 | 2,230 |
| median | 75 | 69,5 | 68 | 73 | 63 | 55,5 |
| mode | 77 | 68 | 68 | 78 | 69 | 55 |
| sd | 7,377 | 7,233 | 8,711 | 8,661 | 9,970 | 12,612 |
| kurtosis | 2,710 | -0,949 | -0,017 | 1,437 | -0,976 | -1,296 |
| skewness | -1,718 | -0,281 | -0,514 | -1,273 | -0,378 | 0,268 |
| min | 51 | 56 | 45 | 44 | 41 | 42 |
| max | 80 | 80 | 80 | 80 | 78 | 80 |

Source: compiled by the authors.

We have conducted Mann–Whitney U test for the hypothesis of no significant difference between exam time in control and test groups. The results are given in Table 4. As the samples sizes are higher than 20, the test statistics at 0,05 significance level is 1,96.

Table 4. Mann–Whitney U test results

| Exam | Entering | Mid-term | Final |
|------|----------|----------|-------|
| U | 1,173 | 3,268 | 2,674 |

Source: compiled by the authors.

The test results show that except the entering exam there is a significant difference in exam time performance between test and control groups. Median statistics for final exam in control group is 68 minutes while in the test group the median time decreases to 55.5 minutes.

Finally, we conducted a Friedman test inside each group of students for a null hypothesis of no time difference between three exams. Q statistics of Friedman test is given in Table 5.

Table 5. Friedman test results

| Statistics | Control group | Test group |
|------------|---------------|------------|
| Q | 4,846 | 18,2 |

Source: compiled by the authors.

Critical values for 3 blocks at 0,05 significance level are 6.077 and 6.063 for 26 and 32 sample sizes respectfully. The results show that in control group students have performed all exams at nearly the same time. The test group up to the end of the course has learned some new programming tips and performed the same tasks faster than the control group.

7. Conclusion

Experience of Cocalc e-course implementation was very useful for both authors and students. As a result of learning new e-course, our test group of students have got the following skills:

- they learned how to work with new LMS;
- they improved their English knowledge level, as the portal cocalc.com is not russified;
- they learned the basics of programming in R and generating reports in Latex;
- they started to pass exams earlier as they mastered writing program code and making reports.

The authors also want to encourage their colleagues to create programming and mathematical courses using this LMS. Mathematicians can use Sage, Python or R, generate tasks using sagemath package. All languages that are supported by Cocalc can also be used for various programming courses. For courses based on R making homework reports is much easier using knitr with rnw format support. We plan to make new e-courses using cocalc.com.

References

1. S.V. Chumakova, Variability of pedagogical objects studies from the point of mathematical and statistical methods. *Journal of Scientific Conferences*, **1-3(5)**, 117-118 (2016). [in Rus.].
2. P.I. Frolova, Application of mathematical statistics methods in educational research. *Actual Problems of Teaching Mathematics at a Technical University*, **5**, 137-141 (2017). [in Rus.].
3. M.V. Mirzoyan, O.D. Rozhenko, The application of mathematical statistics methods in modern pedagogical research. *A New Word In Science and Practice: Hypotheses and Approbation of Research Results*, **9**, 81-86 (2014). [in Rus.].
4. L. Cohen, L. Manion, K. Morrison, *Research methods in education* (7th ed.). New York: Routledge (2011).
5. O.M. Markova, S.O. Semerikov, M.V. Popel, Cocalc as a learning tool for neural network simulation in the special course "Foundations of Mathematic Informatics". In V. Ermolayev, V. Peschanenko, A. Spivakovsky, M.C. Suarez-Figueroa, V. Yakovyna, V. Kobets, Y. Prytula, H. Kravtsov, V. Kharchenko, M. Nikitchenko (Eds.), *14th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2018* (pp. 388-403). Kyev: CEUR-WS (2018).
6. R.F. Miftakhov, A math homework bank development in the cloud computing system Cocalc. *Systems of Computer Mathematics and Their Applications*, **19**, 66-70 (2018). [in Rus.].
7. P. Szabó, J. Galanda, Sage math for education and research. In F. Jakab (Ed.), *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, (pp. 1-4). Sary Smokovec: Institute of Electrical and Electronics Engineers. DOI: 10.1109/ICETA.2017.8102535 (2017).
8. T.V. Zykova, V.A. Shershneva, Y.V. Vainshtein, I.F. Kosmidis, A.A. Kytmanov, S.A. Tikhomirov, Teaching mathematical disciplines in conditions of applying cloud technologies on the basis of LMS Moodle. *Bulletin of Krasnoyarsk State Pedagogical University*, **4(42)**, 58-65. DOI: 10.25146/1995-0861-2017-42-4-21 (2017). [in Rus.].
9. J. Rogness, Mathematical Visualization. *Journal of Mathematics Education at Teachers College. Technology Issue*, **2(2)**, 1-7 (2011).