Collaboration Based Simulation Model for Predicting Students' Performance in Blended Learning

Zhengbing Hu[1], Jun Su[2], and Yurii Koroliuk[3](✉)

[1] School of Educational Information Technology, Central China Normal University, Wuhan, China
hzb@mail.ccnu.edu.cn
[2] School of Computer Science, Hubei University of Technology, Wuhan, China
sjhosix@gmail.com
[3] Chernivtsi Institute of Trade and Economics, Kyiv National University of Trade and Economics, Chernivtsi, Ukraine
yu_kor@ukr.net

Abstract. Due to the positive influence on students' learning outcomes, the interests of studying effective knowledge management has been risen recently. Developing and implementing effective strategies ensures to promote learning outcomes. By reviewing and examining various influence factors, this research study has predicted the major factors that may influence of learning outcomes in blended learning environment. A series of simulation experiments and factor analyses have been conducted in order to investigate collaboration during group learning process. The simulation model for blended learning environment employed in this research has drawn on the characteristics of the Structural Equation Model (SEM) of the blended learning process of 128 students. Both randomness of those student learning behaviors and the reaction to information overload have been considered during simulation modeling. The simulation model enables for greatly increasing statistical samples of student learning behavior analysis. Besides, this research has studied the impact of multiple factors to blended learning mode, these factors include: the size of learning group, the group composition according to previous performance, teaching material amount, and the teacher influence. Experimental results predict that the factors mentioned above can enhance collaborative interaction among students during writing and reading activity. The research results of the optimization restriction factors for blended learning environment achieved in this study can be useful reference for a teacher who are facing the similar challenges. The results of this paper can also be used to reveal and eliminate the problem of inefficient collaboration and poor student performance in blended learning environment. The model proposed in this paper can be integrated with most of decision support systems of universities. AQ1

Keywords: Collaborative learning · Interactive learning environments · Simulations · Teaching strategies · Predicting of academic performance

Z. Hu et al. (Eds.): ICCSEEA 2020, AISC 1247, pp. 1–15, 2021.
https://doi.org/10.1007/978-3-030-55506-1_58

1 Introduction

The informationization of the learning process has optimized the analysis results of student interaction learning, and it has implemented the iteration to the analysis results [1]. The widespread use of distance education technology and the use of logs to dynamically record learning processes allow many parameter data for different learning processes to be preserved and accumulated to the magnitude required for big data analysis. Recently, learning and analysis activities based on big data methods are increasing, and the introduction of big data analysis technology is one of the development trends of higher education in the future [2]. Successful problem solving of knowledge management supposes using accurate measurement of the large number of quantitative and qualitative data comprising adequate means of statistics and analysis.

The scope of data required for applying learning analysis is usually limited to parameters of individual students or a group of students [3]. The research results are often influenced by internal and external factors that influence individual or group learning. These factors include: different size of the study group, different length of the course, insufficiency of selected statistical parameters, the complexity of comparing results of test groups, different length of experimental time, different uniqueness of individual students, the presence of causal data, etc.

Thus, the main objective of this research is to predict the possible factors that influence of students' learning performance improvement by students' collaboration in a blended learning environment. By examining learning management parameters: the number of students in the group, the group composition based on learning performance in the past, the amount of training materials and teacher's influence. The data collected in this experiment ranged from statistical data of 128 students learning in blended learning mode which is further analyzed by using a combination of statistical and factor analysis methods, simulation modeling and table data visualization techniques.

The paper is structured as follows: Sect. 2 presents the background and related papers about effectiveness factor analysis methodology of blended learning environment. Section 3 describes data collection and analysis methods and models. Section 4 shows experimental results and related discussions. Section 5 is the conclusion of this paper. Finally, Sect. 6 presents the limitations and recommendations for future research.

2 Background and Related Work

In recent years, blended learning has become a trend, its principle concept [4] is a combination of learning modes of synchronization (face-to-face) and asynchronization (over the Internet), which achieved better outcomes by effective learning process [5].

Is blended learning efficient? This question can be answered by experience proof and factual results [5]. In general, the assessment performance of blended learning is like that of traditional face-to-face learning. In other words, blended learning is assessed by comparing with traditional face-to-face teaching [6]. Most other researches have realized that blended learning can improve students' learning outcome [7], but few of them gave reasons for the improvement. Some studies reveals that the main reason of such an improvement facilitates to establish collaborations [8] and cooperation [9] within a learning group. Besides, a teacher plays a role as a 'intermediary' who often raises questions. Also confirmed that asynchronous and synchronous e-learning mode has a positive impact on academic performance [10].

Effective teaching strategies are based on the improving of students' educational productivity. Forecasting such productivity under the influence of different factors is the current task when managing blended learning. The model description of the learning environment is laid in the basis of this forecast. Informatization of the learning process creates favorable preconditions of supplying forecasting models with qualitative and quantitative data of learning statistics. Thus, human-computer interaction in a blended learning environment enables a wide range of learning processes to be parameterized and statistically calculated. The collected parameters can be used for various learning content management goals. All the mentioned parameters are stored as log-files of learning management system (LMS), i.e., such as Moodle.

As follows, a large number of parameters and quantitative statistic data of student's educational behavior led to the emergence of some methods of predicting success. The most common is the use of regression forecasting models. However, most of them are based on linear dependencies, take into account a limited number of forecasting factors and describe only individual learning behavior of a student. The results of such forecast are not reliable for building effective collaborative blended learning strategies. To solve such issues A. Elbadrawy and others [11] suggest a class of collaborative multi-regression models. These models have the following parameters: student's past performance, engagement and course characteristics. But these models do not take into account the factor of tutor's active participation and the size of the training group, information overload. Another drawback of these models, together with other regressive ones, is that they give exact forecast only when large student groups are analyzed.

The example of application the another forecasting technology is the work of S. B. Kotsiantis [12]. The researcher suggests the use of machine learning techniques for forecasting students' grades. To construct the model there was used the data of virtual courses, e-learning log file, demographic and academic data of students, admissions/registration info, and so on. The analysis revealed Data Mining technologies that are the most suitable for forecasting. Though, the suggested forecasting model takes into account a limited number of factors, most of which are unmanageable. The main drawback of the model there remains an individual description of a student, not considering the group interaction peculiarities. A similar drawback is peculiar for the forecast model of S. Borkar and K. Rajeswari [13], based on the education Data Mining of the following parameters: graduation, attendance, assignment, unit test, university result. A. Mueen and others [14], C. Romero and others [15] introduce the parameter 'forum participation' to consider the group activity in the Data Mining of the forecasting model. But such model requires a large number of students to form a reliable data sampling and accurate forecast. Also, the factor of the tutor's active participation is not taken into account, etc.

U.R. Saxena and S.P Singh [16] have used Neuro-Fuzzy Systems for forecasting students' performance based on their CPA and GPA. Notwithstanding the fact that researchers confirm a high accuracy of the forecast, it is based only on two individual parameters of a student. Small sample size for building such a model is problematic and made about 20–26 cases of students' results.

K.D. Kolo and others [17] suggest a decision tree approach for predicting students' academic performance. To build the decision tree structure there was used the method of Chi-Square automatic interaction detection of factors: student's grade, student's status, student's gender, financial strength, attitude to learning. Nurafifah Mohammad Suhaimi and others [18] also used such classifier methods as Decision Tree compared with Support Vector Machine, Neural Network and Naïve Bayes for similar goals. Md Rifatul Islam Rifat and others [19] used six state-of-the-art classification algorithms for the prediction task of academic performance. However, such models do not describe the peculiarities of group interaction and cannot be widely applied when elaborating effective teaching strategies.

Therefore, the elaborating of adequate forecast models is complicated by that the process of blended learning is complex and dynamic. However, many of the measured data are stochastic, which requires a deep and representative statistical sample to reveal deterministic dependencies. Data sets in small scale will greatly affect the results of the analysis [20].

The reason for complexity increasing of the sample is objective. First of all, it is a small group scale. The number of groups ranges from 3 to 30, and the number is not enough for accurate factor analysis. Second, the process of collecting statistical information is long. Typically, information is collected during a learning cycle of six months or longer. During this time, the effects of factors may be unequal. There are possibilities occurring influence of factors that are not considered, i.e. the impact of the semester. Third, there are measurement errors associated with assessment for student learning effects.

Even adopting the most accurate model (with an accuracy of more than 86% [21]) in the above range (small scale 3–30 people), significant inconsistencies with the analysis results might be possibly occurred, which complicates learning management. On the one hand, assessment methods given by most hypothesis test methods (p-values, β-values, t-tests, confidence intervals, etc.) are difficult to assess students' performance in a small group. On the other hand, quantitative research can qualitatively explain the main learning mechanisms, student interaction, and information exchange. It is common practice to interpret these results in the form of structural equation modeling (SEM). However, thorough implementation of SEM in learning management requires further research especially for the aims of forecasting. Establishing a simulation model based on SEM is a helpful suggestion because it enables to verify models of the teaching strategy in the form of scenario collection of management parameters. Multiple simulation experiments can greatly improve the reliability of the predicting and therefore expand the research samples as endogenous and exogenous parameters of models.

3 Data, Methods and the Models

The data sample in this paper derived from 128 third-year students of economic cybernetics specialty, who had been studying for 2013–2018 academic years in Chernivtsi Institute of Trade and Economics of KNUTE (ChITE KNUTE). Blended learning mode was applied for those students in their learning process. The same subject in a semester is considered as an optional unit, all subjects were taught in Ukrainian language. Students participating in the study group with different performances. This study involved a total of 12 learning groups, there were 6–21 members in each group. Teaching work consists of two parts: classroom and online courses. Online courses were conducted by Moodle software over the Internet which can be found on ChITE KNUTE's website (http://www.dist.chtei-knteu.cv.ua:8080/).

The number of textbooks was determined by ECTS (European Credit Transfer System) and was divided into several grades ranging from 1.5, 3 to 4.5 points. The textbook was divided into 6–18 independent topics. The amount of work was the same for every week. Each topic was designed for students to learn theoretical knowledge to complete practical tasks. During the course learning, students attended lectures as required, students learned strategies for competing learning tasks, teachers evaluated students' achievements on previous learned subjects. Implementing learning tasks was through a forum, and an open asynchronous online discussion was conducted under teachers' guidance. Discussions were conducted in an asynchronous way, including posting, reading, and replying to posts written by group members or teachers. Learning outcome for each subjects is rated by 100 scores. The final scores for all subjects are calculated by the arithmetic mean value.

The measurement parameters of a blended mixed learning process include: student ID, group ID, achievement based on past course performance, final course achievement, subject number, ECTS academic credits, group size, theoretical material amount, amount of each student reading post, amount of each student witting post, amount of a teacher writing post. The log files generated by Moodle were analyzed in this study, the problems of accurately and comparatively evaluating parameters were occurred, it is not easy to identify some parameters including amount of theoretical textbooks, amount of readings posts for a student, amount of writing each post for a student, amount of writing each post for a teacher. Using the quantity indicator for writing or reading posts cannot clarify the actual amount of information exchange in a blended learning environment. In order to overcome this problem, it is necessary to calculate the number of characters written and read by each student in the learning process. The obtained characters can be converted to bytes using the method of [22]. Therefore, the information unit containing the characters of the 33-letter Ukrainian alphabet is shown as follows:

$$H_0 = \log 233 = 5.044\,\text{bits} \quad (1)$$

The goal of data analysis is to rely on other parameters valid in a blended learning environment and to maximize performance out of students. Reading and writing activity are treated as a direct and important factor, which represents degree of student's collaboration in a group as confirmed by K. Bielaczyc and A. Collins [23]. Proactive reading and posting are in turn related to other management and non-management factors in blended learning environment. The structure of the relationship model of these factors is shown in Fig. 1.

Learning process of 128 students in the statistical sample was analyzed in order to describe and confirm the importance of the relationship. In case of considering the number of students in a group, amount of reading and posting are accumulated. In order to effectively study other factors, at this stage, the total amount posted by the group reflects the influence of the group, group size and the total number of posts are not differentiated and are treated as the same factor. Post reading indicator is the proportion to all posts read by a student. The results of SEM analysis of the coefficients and the multiple linear regression indicator forms are given in Table 1.

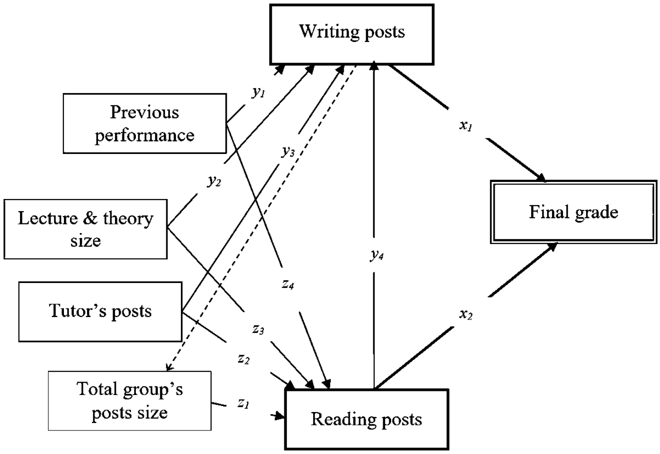

Fig. 1. SEM of student's performance

Table 1. SEM parameters of student's performance in a blended learning group

	Reading posts*					Writing posts**					Final grade***		
R^2	0.81					0.60					0.55		
Standard error	0.10					11.69					7.18		
F	0.007					0.02					0.013		
α	0.00					0.00					0.00		
Coefficients	Y-intercept	z_1	z_2	z_3	z_4	Y-intercept	y_1	y_2	y_3	y_4	Y-intercept	x_1	x_2
Value	−0.155	−0.001	−0.001	0.001	0.01	−83.19	1.28	0.091	−0.158	−15.44	61.67	0.295	19.96
t-value	−2.144	−10.35	−1.528	8.923	12.75	−9.810	10.13	5.47	−1.864	−2.011	34.19	8.13	6.70
p-value	0.034	0.00	0.129	0.00	0.00	0.00	0.00	0.00	0.065	0.046	0.00	0.00	0.00

Units of measurement: * from 0 to 1, ** Kbyte, *** from 0 to 100.

For the formation of SEM there were made the hypotheses about the existence of substantial connections between the studied input and output parameters. Hypothesis testing was performed by the means of MS Excel multiple regression of observed statistical parameters.

The analysis result of the significance of the SEM reveals that the value of index R^2 was higher than its critical value 0,349 (when $\alpha = 0.01$). The value of index F is always much smaller than the critical value, and the value of calculated α (significance of F) is close to zero. The t-value of absolute index exceeds 1.97 despite coefficients z_2 and y_3. The p-value is smaller than 0.05, despite the coefficients z_2 and y_3.

The results of the statistical analysis confirmed the relationship between input and output parameters of SEM. Factors affecting parameters of reading and postings include: the student's previous performance and the number of theoretical textbooks on the subject. Teacher guidance has a negative impact on students' reading and witting posts, but the impact is not big. The similar influence of teacher guidance on student learning is described in the work of Mazzolini and Maddison [24]. It has been confirmed that after completing the first two tasks, students use more suggestions from other group members than teacher's guidance [25]. In this study, a teacher can create conditions for conduct discussion and collaboration among students and supervise the process of ongoing discussion. In case of occurring passive discussions and off-topic situations, the teacher provides timely correction. The effects mentioned above can explain why multiple linear regression analysis shows a negative impact on the practice of writing discussion contents. Analysis confirms the existence of this impact. When discussing the actual online task, the output parameters of scores are generated by the positive influence of reading and writing posts.

Although dependency exists, the output SEM parameters contain an evident standard error index. In order to establish an effective knowledge management strategy, it is necessary to accurately predict the influence of available management factors. These calculations enable to generate a combination of certain factors in blended learning environment in order to identify the area of promoting students' learning performance. Obtained SEM parameters revealed the nature of output performance relying on input performance. To accurately describe a blended learning process, it needs to use proper methods to analyze students' behaviors upon collected parameters.

The nature of students' reaction to external stimuli is behaviors and a kind of clustering. This article has introduced some quantitative and qualitative methods to describe certain human behaviors. Therefore, in recent years, the methods of human-centered systems (HCS: Human Centered Systems) have often been used to describe the behavior of staff in industry, military, and stock markets, as well as in other environments [26]. When behavioral simulation model is used to study human behaviors, it is common to construct a single person's model (for example, a worker) at first that specific environmental model is established based on it, and the model is applied to a group of people in social networks afterwards [27].

In order to study students' learning behavior, a simulation model was established in this study, and the dependence of SEM parameters was calculated based on the model. A simulation model for each student was built using MATLAB's Simulink tool. The model uses some simulation programming elements such as 'Add', 'Subtraction', 'Division', 'Discrete Time Integrator', 'Uniform Random Number', 'Pulse Generator', 'Oscilloscope' and so on. The model contains a data input module (red ellipse 1 in Fig. 2) and a data output module (red ellipse 4 in Fig. 2). The input data is submitted for five parameters: Previous performance, Lecture & Theory size, Tutor's posts, Total student's write, Subjects. The input parameters are then multiplied by the SEM coefficients according to its structure (Gain blocks) and added (Subtract blocks). Multiple linear regression coefficients were described using SEM (indicated by Ellipse 2 in Fig. 2). Both the effects of standard error and the randomness of student behaviors are marked with Ellipse 3 in Fig. 2. Uniform Random Number generators are used to take into account random variables. The result of this processing is three discrete outputs indicated by Ellipse 4 in Fig. 2: Grade, Read, Write.

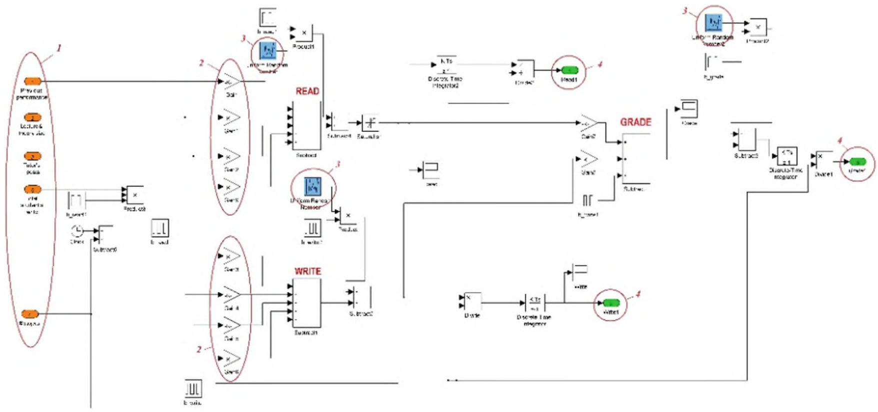

Fig. 2. Simulink/MATLAB model of a student's learning activity in a blended learning environment

A student's simulation model is in the status of discrete. The number of weeks is equivalent to the number of learning subjects. Each learning subject is an independent learning unit that highlights the characteristics of discreteness. Each learning subject contains a task, an online discussion module and a grade. This approach differs from a similar simulation model of continuous behaviors described by S. Elkosantini and D. Gien [27].

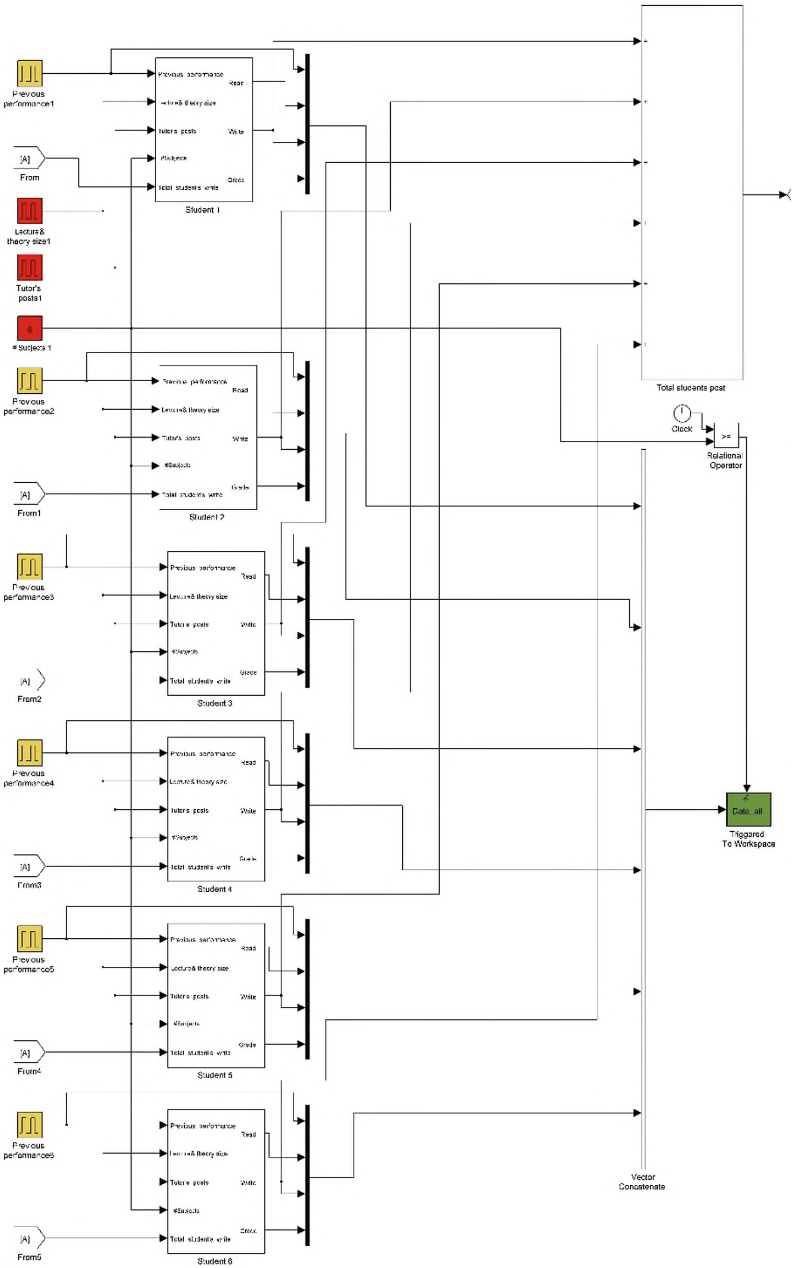

Fig. 3. Example of a learning group model (Simulink/MATLAB) with 6 students

Reading-writing practices and subject ratings are calculated in one cycle of the model. The final score is comprehensively calculated based on the results of each subject, and the students' scores are given by the real assessment system in a blended learning environment. The individual student models are combined into a learning collaboration network and model of a group is formed on this basis. Figure 3 presents a blended learning environment for a learning group with 6 students. The input parameters of the model include: the number of the textbook, the number of posts written by a teacher, the number of subjects (the course lasted for several weeks), and previous performance of each student. The input parameters were formed using a pulse generator with the period of signaling - one training week. The inputs for each student model (Student blocks) were formed individually, taking into account their training peculiarities, number of subjects, tutor's activity and others (Fig. 3, right side). Thus, there were formed the initial conditions of the experimental forecast. The outputs of each student sub-model were commuted to form a repository of student-produced information in the form of posts (Total students' posts block). The results of the model calculations were exported to an external file for later study using MATLAB software for further analysis.

4 Results Analysis and Discussion

By calculating the experimental results obtained from the experiment with different input parameter sets, seven groups of models were established, in which the numbers of students were 3, 6, 8, 10, 12, 15, 20 students. Students in these groups are completely different in their previous performance. The average scores of the group model ranges from 60 to 90. Some students got 60–95 scores before joining the group. In a blended learning environment, the performance indicator is dynamically changed accordingly as a student's actual performance changes.

There are three kinds of textbooks in the model: six subjects (capacity: 160 Kbyte), 12 subjects (capacity: 340 Kbyte), and 18 subjects (capacity: 480 Kbyte). The number of selected subjects depends on the ECTS credits required by the courses. The influence after a teacher posting 30–110 Kbytes was analyzed and examined.

Modeling 105 combinations of scenario input parameters received results in form of table are converted in form of matrix by the XYZ-Gridding tool of the OriginPro application, which is further converted to a visualization matrix by the tool – 'Color Map Surface (OriginPro)' (see Fig. 4 and 5). Due to the restriction of the number of input factors, it makes more difficult to accurately estimate achieved results, such an inferiority may result in misinterpreting other factors. However, the simulation model fully inherits the accuracy of the SEM indicators (Table 1: R^2 for reading activity is equal 0.81, R^2 for writing activity is equal 0.6, R^2 for grading score is equal 0.55) that gives a reason to consider the data of Fig. 4 and 5 reliable. In addition to the SEM accuracy indicators, there was held the comparing of academic performance for two academic groups of students who completed the course. The first group consisted of 6 students who studied 18 subjects. When doing the tasks, the tutor's activity was 35.6 Kbyte posts per student. The results of the course showed that the average performance has increased by 2.71 points. Simulation modeling predicts a 1.9 point performance increase (downward deviation 30%). The second group consisted of 7 students who studied 6 subjects. In this case, the tutor's activity was 10.4 Kbyte posts per student. After completing the course, the average performance has increased by 2.17 points. Simulation modeling predicts a 2.7 point performance increase (upward deviation 25%). The performance rates obtained correspond to the calculated R^2.

By the influences of a teacher's guidance and the factor of student's previous performance, the experiment in this research studied students' writing practice in blended learning environment (Fig. 4).

Figure 4 presents the absolute index of the writing practice (vertical axis) for three different course sizes: 6, 12, 18 subjects. The absolute index of writing practice was predicted in Kbyte of written posts by the entire modeling group. The horizontal axes show the initial conditions of prediction: tutor's posts per student (Kbyte) and previous performance (in a 100 point scale).

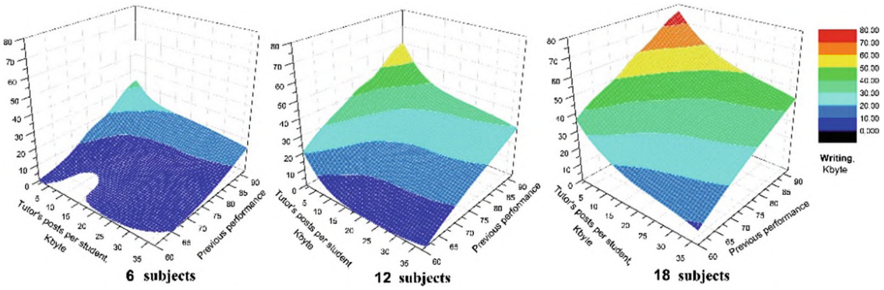

Fig. 4. Predicting results of writing practice in blended learning environment

As we can see from Fig. 4, for courses with more lectures and theoretical subjects, writing amount is larger. However, in the case of different number of subjects, the study did not find the differences in the form of exercises. With the decreasing of a teacher's guidance, the amount of exercises for students with different previous scores increases. For students with good previous scores, the growth speed is faster. As a result, students are tending to establish online collaboration more actively when solving specific tasks, finding out key points of the task, and solving difficult theoretical problems.

For students who had more than 75 scores previously, the study found that the role of a teacher is only assistance and the teacher's influence in promoting learning performance is not evident, the teacher can not optimize the performance greatly. For students with worse previous performance, more teacher's guidance can lead to reduce the amount of posts. Through verification, the proper explanation for the above phenomena is that students tend to read the posts by teachers and by the students who had better previous learning performance than to write their own posts, which suggests that these students become knowledge receiver according to Gillies [28].

A small peak appearing in Fig. 4 reflects that active posters are made up of a small number of students who usually have better performance. AbuSeileek [29] confirmed this characteristic. Although the results obtained by this experiment are different than

Qiu's study [30], there exists a direct proportion between the total number of posts posted by the students and the total number of posts posted by a teacher. Operational modeling shows that more teacher's guidance provides more information and might result in information overload. Students write less posts due to the influence of information overload, which was confirmed by Qiu [30].

To perform objective analysis, additional average performance of a group (the difference between the final grade of the course and the previous grade) was observed. The experimental group mixed the previous scores. The homogenous case was not considered because it is impossible to set up multiple groups with 6 students with the same level of previous performance.

Figure 5 represents the added average academic performance of the training group (vertical axis) for three different course sizes: 6, 12, 18 subjects. The added average academic performance was predicted as the difference between the received and previous course performance of the group. The horizontal axes show the initial prediction conditions: tutor's posts per student (Kbyte) and the size of the training group.

Figure 5 shows that the amount of theoretical material greatly affects critical performance indicator. In case of 6 subjects, even if teacher's guidance is reduced, the scores of groups with different numbers are promoted. For those groups with more than 10 student, with the increase in the number of posts posted by a teacher, the average scores of the groups are declined. As shown below, students in larger groups rely more on teacher's guidance than on collaborations of students.

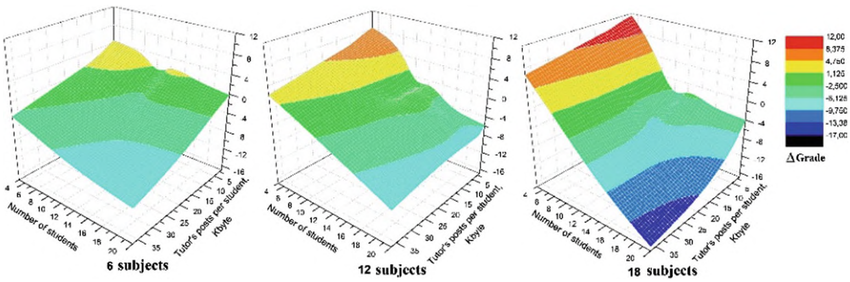

Fig. 5. Predicting results of the students' performance (course grade) in blended learning

In case of 18 subjects, teams with up to 8 students achieved very high additional scores at any level of teacher's guidance. In case of large groups, the bottleneck of optimizing learning performance still exists, but only when a group contains 10 students. A further increase in the number of a group can suddenly reduce the ratio of performance increasing. Students can not deal with the large number of posts sent by group members. More teacher's guidance may decline the average scores of these groups. This result indicates that a teacher should pay more attention on groups with lower initiative and less problems. However, teacher's guidance can not significantly affect the collaborations among students. The study results revealed the significant influence of group size in information overload that also was described by Hewitt and Brett [31] and Pfister and Oehl [32].

The proven accuracy of the predictions and the consistency of the results with the work of other scientists give reason to claim that the research objectives were achieved. The simulated forecast conditions have shown the utmost ability to optimize the management of LMS in ensuring active collaboration and increasing the success of training groups under the conditions of blended learning.

5 Conclusion

This paper uses the predicting results of the effects obtained in a blended learning environment as the outcome of implementing different knowledge management strategies. Input management parameters include learning group size, grouping based on previous learning performance, the amount of learning materials and the cycle of learning a subject, the teacher's guidance, appropriate learning strategies are chosen by considering the combination the parameters in blended learning environment. By analyzing the whole process of students learning in a blended environment, correlated simulation modeling experiment and the student collaborations within a group are evaluated objectively. The simulation model considers random dependence of the learning behavior set of 128 students in blended learning environment described by SEM.

Our study revealed that the success of achieving group's additional scores depends greatly on group size and the amount of learning materials. According to the maximum number of learning subjects and the maximum course cycle, small groups of up to 8 students can achieve 2–12 additional scores. This study suggests that the role of teachers is only 'assistance' in the case of different course scale and group size, teachers is not able to solve the bottleneck of restricting performance improvement in blended learning environment. Teacher's guidance is helpful to avoid the occurrence of information overload.

6 Limitations and Future Work

The limitations of this study are related to the fact that the data contained in constructing SEM and simulation model is only relevant to the exchange of textual information among students and the learning environment. The estimation of the theoretical textbook does not consider the included graphic images. This study did not consider the time delay between learning the textbook and starting the discussion. The quality index of the discussion posts was also not analyzed.

Due to the restriction of the number of input factors, it makes more difficult to accurately estimate achieved results, such an inferiority may result in misinterpreting other factors. However, the simulation model fully inherits the accuracy of the SEM indicators.

Therefore, future studies could explore the possibility of improving the SEM and the simulation model. In particular, it is assumed taking into account graphic data interchange in a blended learning environment. Specification of the SEM indicators will be held by considering new statistic data of the observation.

Acknowledgments. This scientific work was partially supported by RAMECS and self-determined research funds of CCNU from the colleges' primary research and operation of MOE (CCNU19TS022).

References

1. Dyckhoff, A.L., Zielke, D., Bültmann, M., Chatti, M.A., Schroeder, U.: Design and implementation of a learning analytics toolkit for teachers. Educ. Technol. Soc. **15**(3), 58–76 (2012)
2. Siemens, G., Long, P.: Penetrating the fog: analytics in learning and education. Educause **46** (5), 30–40 (2011)
3. Daniel, B.: Big data and analytics in higher education: opportunities and challenges. Br. J. Educ. Tech. **46**(5), 904–920 (2015). https://doi.org/10.1111/bjet.12230
4. Garrison, D.R., Kanuka, H.: Blended learning: uncovering its transformative potential in higher education. Internet High. Educ. **7**(2), 95–105 (2004). https://doi.org/10.1016/j.iheduc.2004.02.001
5. Means, B.: The effectiveness of online and blended learning: a meta-analysis of the empirical literature. Teach. Coll. Rec. **115**(3) (2013). http://www.tcrecord.org/library/content.asp?contentid=16882/. Accessed 23 May 2019
6. Means, B., Toyama, Y., Murphy, R., Bakia, M., Jones, K.: Evaluation of evidence-based practices in online learning: a meta-analysis and review of online learning studies. US Department of Education, Washington, DC (2009)
7. Zhao, Y., Lei, J., Yan, B., Lai, C., Tan, H.S.: What makes the difference? A practical analysis of research on the effectiveness of distance education. Teach. Coll. Rec. (2005). https://doi.org/10.1111/j.1467-9620.2005.00544.x
8. Kolloffel, B., Eysink, T.H.S., de Jong, T.: Comparing the effects of representational tools in collaborative and individual inquiry learning. Int. J. Comput.-Supported Collab. Learn. **6**(2), 223–251 (2011). https://doi.org/10.1007/s11412-011-9110-3
9. Johnson, D.W., Johnson, R.T.: Cooperation and the use of technology. In: Johanssen, D.H. (ed.) Handbook of Research on Educational Communications and Technology, 2nd edn, pp. 785–811. Lawrence Erlbaum Associates, Mahwah (2004)
10. Dada, E.G., Alkali, A.H., Oyewola, D.O.: An investigation into the effectiveness of asynchronous and synchronous e-learning mode on students' academic performance in National Open University (NOUN), Maiduguri Centre. Int. J. Mod. Educ. Comput. Sci. (IJMECS) **11**(5), 54–64 (2019). https://doi.org/10.5815/ijmecs.2019.05.06
11. Elbadrawy, A., Studham, R.S., Karypis, G.: Collaborative multi-regression models for predicting students' performance in course activities. In: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge - LAK 2015 (2015). https://doi.org/10.1145/2723576.2723590
12. Kotsiantis, S.B.: Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. Artif. Intell. Rev. **37**(4), 331–344 (2011). https://doi.org/10.1007/s10462-011-9234-x
13. Borkar, S., Rajeswari, K.: Predicting students academic performance using education data mining. Int. J. Comput. Sci. Mob. Comput. **2**(7), 273–279 (2013)
14. Mueen, A., Zafar, B., Manzoor, U.: Modeling and predicting students' academic performance using data mining techniques. Int. J. Mod. Educ. Comput. Sci. (IJMECS) **8** (11), 36–42 (2016). https://doi.org/10.5815/ijmecs.2016.11.05

15. Romero, C., López, M.-I., Luna, J.-M., Ventura, S.: Predicting students' final performance from participation in on-line discussion forums. Comput. Educ. **68**, 458–472 (2013). https://doi.org/10.1016/j.compedu.2013.06.009
16. Saxena, U.R., Singh, S.P.: Integration of neuro-fuzzy systems to develop intelligent planning systems for predicting students' performance. IEEE Technol. Eng. Educ. (ITEE) **7**(4), 1–7 (2012)
17. Kolo, K.D., Adepoju, S.A., Alhassan, J.K.: A decision tree approach for predicting students academic performance. Int. J. Educ. Manag. Eng. (IJEME) **5**, 12–19 (2015). https://doi.org/10.5815/ijeme.2015.05.02
18. Nurafifah, M.S., Abdul-Rahman, S., Mutalib, S., Hamid, N.H., Ab Malik, A.M.: Review on predicting students' graduation time using machine learning algorithms. Int. J. Mod. Educ. Comput. Sci. (IJMECS) **11**(7), 1–13 (2019). https://doi.org/10.5815/ijmecs.2019.07.01
19. Rifat, M.R.I., Al Imran, A., Badrudduza, A.S.M.: Educational performance analytics of undergraduate business students. Int. J. Mod. Educ. Comput. Sci. (IJMECS) **11**(7), 44–45 (2019). https://doi.org/10.5815/ijmecs.2019.07.05
20. Xie, K., Di Tosto, G., Lu, L., Cho, Y.S.: Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis. Internet High. Educ. **38**, 9–17 (2018). https://doi.org/10.1016/j.iheduc.2018.04.002
21. Marbouti, F., Diefes-Dux, H.A., Madhavan, K.: Models for early prediction of at-risk students in a course using standards-based grading. Comput. Educ. **103**, 1–15 (2016). https://doi.org/10.1016/j.compedu.2016.09.005
22. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
23. Bielaczyc, K., Collins, A.: Fostering knowledge-creating communities. In: O'Donnell, A.M., Hmelo-Silver, C.E., Erkens, G. (eds.) Collaborative Learning, Reasoning, and Technology, pp. 61–98. Lawrence Erlbaum Associates, Mahwah (2006)
24. Mazzolini, M., Maddison, S.: Sage, guide or ghost? The effect of instructor intervention on student participation in online discussion forums. Comput. Educ. **40**(3), 237–253 (2003). https://doi.org/10.1016/S0360-1315(02)00129-X
25. Moallem, M.: An interactive online course: a collaborative design model. Educ. Tech. Res. Devel. **51**(4), 85–103 (2003). https://doi.org/10.1007/BF02504545
26. Elkosantini, S.: Toward a new generic behavior model for human centered system simulation. Simul. Model. Pract. Theory **52**, 108–122 (2015). https://doi.org/10.1016/j.simpat.2014.12.007
27. Elkosantini, S., Gien, D.: Integration of human behavioural aspects in a dynamic model for a manufacturing system. Int. J. Prod. Res. **47**(10), 2601–2623 (2009). https://doi.org/10.1080/00207540701663490
28. Gillies, R.M.: Teachers' and students' verbal behaviours during cooperative and small-group learning. Br. J. Educ. Psychol. **76**(2), 271–287 (2006). https://doi.org/10.1348/000709905X52337
29. AbuSeileek, A.F.: The effect of computer-assisted cooperative learning methods and group size on the EFL learners' achievement in communication skills. Comput. Educ. **58**(1), 231–239 (2012). https://doi.org/10.1016/j.compedu.2011.07.011
30. Qiu, M.: A mixed methods study of class size and group configuration in online graduate course discussions. ProQuest Dissertations and Theses (2010)
31. Hewitt, J., Brett, C.: The relationship between class size and online activity patterns in asynchronous computer conferencing environments. Comput. Educ. **49**(4), 1258–1271 (2007). https://doi.org/10.1016/j.compedu.2006.02.001
32. Pfister, H.R., Oehl, M.: The impact of goal focus, task type and group size on synchronous net-based collaborative learning discourses. J. Comput. Assist. Learn. **25**(2), 161–176 (2009)