Zhang, J.-H., Zhang, Y.-X., Zou, Q., & Huang, S. (2018). What Learning Analytics Tells Us: Group Behavior Analysis and Individual Learning Diagnosis based on Long-Term and Large-Scale Data. *Educational Technology & Society, 21* (2), 245–258.

# What Learning Analytics Tells Us: Group Behavior Analysis and Individual Learning Diagnosis based on Long-Term and Large-Scale Data

**Jia-Hua Zhang[1*], Ye-Xing Zhang[1], Qin Zou[2] and Sen Huang[1]**

[1]Department of Educational Technology, Zhejiang Normal University, China // [2]Dongfang College, Fujian Agriculture Forestry University, China // zjnuzjh@zjnu.edu.cn // zhangyx0216@gmail.com // etzoey@163.com // huangsen0414@gmail.com

[*]Corresponding author

**ABSTRACT**

The practice and application of education data mining and learning analytics has become the focus of educational researchers. However, it is still a difficult task to explore the law of group learning and the characteristics of individual learning. In this study, the online learning logs of 1,088 students from 22 classes were analyzed from the aspects of their login behaviors, resource utilization, quizzes, interactive behaviors, and academic achievement. To address these issues, multiple methods, including statistical analysis, visualization social network analysis and correlation analysis, were used to analyze the process and results of online learning. The results reveal the characteristics of group behavior of online learners and highlight the key factors that influence the learning process and outcomes of individual learners. From the view of students, these factors include the length and allocation of online time, the effective utilization of resources, social interaction, online learning support and services, etc. From the perspective of teachers, the factors include the management of online teaching, the appropriateness of learning resources, the effectiveness of online intervention strategies, the accurate feedback for online learners, etc. Therefore, learning analysis technology can not only standardize the assessment of learning outcomes, but can also focus more attention on the standardization of learning process assessment. It also identifies the main factors that affect the online learning outcomes and the group characteristics of online learners. At the same time, it provides the learners with personalized learning diagnosis reports which can both help learners understand their own learning status and promote instructors' accurate teaching and reasonable evaluation.

**Keywords**

Learning analytics, Learning analysis, Learning diagnostic, Online learning, Educational big data

## Introduction

The rapid development of science technology and the popularity of the Internet have promoted the growth and accumulation of data with amazing speed. The arrival of the era of big data indicates a new trend in the development of information technology. Therefore, the concept of big data has been of great interest to all sectors of society since it was first proposed. Of particular interest is the fact that it brings extraordinary innovation and challenges to the development mode and decision-making methods in the fields such as economics, culture, and politics. Education is no exception, with teaching assessment methods, education management tools, learning behavior analysis, and professional development all affected by big data.

The term "Big data" is similar in meaning to "information explosion," "massive data," and so on. It is difficult to describe the specific meaning of these words. The McKinsey Global Institute defines big data as a database which is beyond the grasp, storage, management, and analysis capabilities of traditional database software tools. It states: "The massive data mining and application may indicate a new wave of productivity growth and the arrival of the wave of consumer surplus, and data has penetrated into every field" (Manyika et al., 2011). Thus, big data are the data which can be effectively used, and contain huge amount of valuable information and a variety of statistical data.

With the popularization of online teaching and learning, more and more learning platforms are being applied in education. Digital learning information and educational data are growing rapidly. It has become a challenging issue to transform these educational data into useful information and knowledge which affects both students' learning effect and instructors' teaching modes as well as the decision-making of educational management and resource allocation. In addition, the new trend of intelligentization of learning environments needs deep research on learning analytics. A smart learning environment can offer instant and adaptive support to learners by immediate analysis of the needs of individual learners from different perspectives (Hwang, 2014). Therefore, it has become a hot research topic for many researchers to explore how to use educational data in scientific and effective learning analysis.

However, there is still a lack of long-term and large-scale teaching practice in the field of learning analysis at present. Most of the case studies in this field have used relatively small samples (Martin & Ndoye, 2016; Kim, Park, Yoon, & Jo, 2016; Sun, Lin, & Chou, 2016; Jin & Kim, 2016; Bainbridge, Melitski, Zahradnik, Lauría, Jayaprakash, & Baron, 2015), not to mention the lack of long-term continuous large-scale learning behavior analysis. Therefore, this study reports the use of multiple learning analytics tools and methods in a large-scale and long-term teaching practice. The online learning data of 1,088 students from 22 classes were analyzed to reveal their behavioral characteristics and to provide personalized diagnostic reports to show the value and significance of learning analytics in education settings.

## Literature review

### Applications of learning analytics

Learning analytics refers to the use of learners' data and analysis models to discover information and social connections, and to predict and advise on learning (Siemens, 2010). It is a multidisciplinary method based on data processing technology, educational data mining, and visual data analysis (Scheffel, Drachsler, Stoyanov, & Specht, 2014). To implement continuous formative evaluation and performance evaluation, learning analytics focuses on the measurement of learning and the assessment of academic risk (such as failure to pass the exam, dropping out of the course). Through presenting facts and problems based on learning data, instructors and academic administrators can take measures to intervene when high risk learners are identified, and hence advice can be provided to reduce their academic risk (Pistilli & Arnold, 2010; Lauría, Baron, Devireddy, Sundararaju, & Jayaprakash, 2012). Some researchers have used learning data to predict students' performance, and have provided appropriate interventions for students who were at risk in their learning (Atif, Richards, Bilgin, & Marrone, 2013; Zhang, Fei, Quddus, & Davis, 2014). In addition, some scholars have done research on the use of big data to analyze students' behaviors in higher education, such as reading materials, submitting homework, interacting with classmates, and taking examinations in online courses. It helped to identify learners with poor performance, and hence to provide suggestions and guidance for improvement (Bienkowski, Feng, & Means, 2012; Lauría et al., 2012; Freitas et al., 2015).

Clow (2013) pointed out that teachers should participate in the process of learning analytics to improve their teaching quality. Using some specific tools, such as learning analytics dashboards, managers, visualization tools, and so on to analyze and monitor students' online learning behavior can provide students with learning support and improve their academic performance. For example, Zapparolli and Stiubiener (2017) presented a FAG Tool integrated with Moodle to help teachers monitor the participation of their students and take corrective measures in the teaching process. Kim, Jo and Park (2016) indicated that a dashboard can encourage learners, and provide support and service for learners with different academic achievement levels. Echeverría, Benitez, Buendia, Cobos, and Morales (2016) used a Learning Analytics Manager to monitor the processes of collaborative activities and the students' motivation. Lai and Hwang (2015) utilized a spreadsheet-based visualized mindtool to improve students' academic performance and their learning attitude. Martin and Ndoye (2016) found that data analysis and the visualization tools were useful in terms of improving teaching and learning. Leeuwen, Janssen, Erkens, and Brekelmans (2015) used two learning analytics tools to increase teachers' confidence and help students benefit more from the teacher's presence.

The application of different methods in learning analysis can present different data analysis results for teachers and students. Clow (2013) suggested that learning analytics should link to the management approaches that focus on quantitative metrics. These methods include data statistics and visualization analysis, correlation analysis, social network analysis, clustering and outlier analysis, and so on (Macfadyen & Dawson, 2010; Shiri, 2016). For example, correlation analysis can help the instructor to analyze the relevance between students' learning behavior and performance, so as to make more scientific decision-making and improve the process of teaching and learning (Sun et al., 2016; Martin & Whitmer, 2016; Auvinen, Hakulinen, & Malmi, 2015; Firat, 2016). Content and sequence analysis can be used to verify the effectiveness of specific learning methods for improving academic performance (Hwang & Chen, 2017; Hwang, Hsu, Lai, & Hsueh, 2017). Teachers and researchers can take specific methods to obtain the corresponding learning analysis results, then provide appropriate services for their students.

Applying some specific tools and methods in learning analysis to exploit data on platforms (such as Blackboard, Moodle) has been a key issue that many researchers have focused on. Whether these tools and methods are appropriately used directly affects the results and value of the data analysis. Moreover, using a predictive model in LMS may help to provide effective intervention and support for at-risk students in a timely manner. For example, Purdue University used a prediction algorithm to determine the risk of students and to provide intervention based on the data collected from the curriculum management system (Arnold & Pistilli, 2012).

Some researchers respectively designed and applied learning analysis and forecasting models to identify those students at risk (Piech, Sahami, Koller, Cooper, & Blikstein, 2012; Kim, Park, Yoon, & Jo, 2016). These studies also indicated that the high precision prediction model provided the possibility of early detection of high risk students and timely intervention.

### The challenges faced by learning analytics research

The research of learning analytics has become deeper, partly because of the expansion of data storage capacity and the development of research ability. Due to the convenience of data acquisition and the availability of data, learning analytics can provide a better model for online learning and interaction. At the same time, it also faces many challenges. Avella, Kebritchi, Nunn, and Kanai (2016) pointed out that learning analytics emphasizes the need for more in-depth understanding of how to analyze data and use information from the learning process to optimize results in the education process. In addition, better standardized assessment can improve student participation, while also helping to improve the level of the learner's ability. Moreover, few studies have focused on how to obtain the data of standard learning activities, which are independent of the teaching methods and learning behaviors in the e-learning environment (Jin & Kim, 2016).

In general, the main challenges faced by learning analytics research are the following: (1) Most teaching activities are still face-to-face in traditional classrooms, and few technologies are used in the learning process. It makes the data of the students' learning processes difficult to collect. (2) The functions of the tools used for learning analysis are limited, and most data statistics and analysis work is carried out manually. This not only increases the workload and brings inconvenience for the instructors and researchers, but also means that some of the value hidden in the data cannot be fully excavated. (3) The tools and methods used in the learning analysis lack standardization, causing some limitations and inconformity in the research. It is therefore a potential but important issue to propose effective data integration, cleansing methods and management tools for processing educational data (Hwang, Chu, & Yin, 2017). (4) Because of the complexity of educational data collection and the sparsity of educational data structures, the depth of learning data analysis is currently insufficient, meaning that much of the value hidden within the education data cannot be fully excavated. Baker (2016) suggested that learning analytics had considerable potential for achieving high quality adaptive personalization of learning, rather than simply evaluating. (5) The ethical problem in learning analytics also faces challenges. Use of the data collected can infringe on the learners' personal privacy. In order to solve the ethical dilemma, researchers need to improve the data transparency, and must obtain the consent from learners in the research (Lawson, Beer, Rossi, Moore, & Fleming, 2016).

### Research questions

This research used a variety of tools and methods to mine learning data. The data generated by a massive blended course were analyzed from two aspects, namely group level and individual level. The concrete method, process and results of the education data analysis are displayed. It provides personalized learning feedback and diagnostic reports for learners, as well as accurate teaching feedback and decision-making reference for teachers and administrators. In so doing, the study aimed to address the following research questions:

- What are the characteristics of students' online learning behavior?
- Which data variables in Moodle correlate significantly with students' achievement?
- How can teachers provide personalized learning feedback for students?

## Method

This study is based on an online course on the Moodle platform. The course combined online and offline ways of teaching, and lasted from September 2015 to January 2016. A total of 22 classes and 1,088 learners taking this course participated in the study. Students who take the course need to participate in traditional face-to-face instruction in the classroom, and also log into the Moodle Platform for online learning. Teachers and teaching assistants provide the necessary support services throughout the whole semester.

In this study, multiple methods were adopted, including statistical analysis, visualization, social network analysis and correlation analysis. Some third-party plugins for Moodle (including Gismo, Course Dedication, and Forum Graphs and so on) were used to analyze students' behavior with the idea of Big Data and educational data mining technology. Through analyzing the behavior of learning groups and individuals, some interesting principles and characteristics of online learning were revealed. In addition, it provides students, teachers, and administrators with useful information for learning diagnosis and decision-making purposes.

## Results

### Group behavior analysis of online learning

Moodle can record detailed information of students' access to each task module. In this paper, multiple plugins for Moodle were used to collect data according to the research needs. Nine variables were identified to be tracked in this course, which may relate to students' academic achievement in the course. These variables include number of access times, online time, browsing resource, access frequency, forum posted, forum replied, assignment submission, access period of time, and quiz attempts. The various types of operational behavior of the students involved in this course learning were collected, processed, and analyzed. Finally, the visual analysis results are presented.

*Analysis of access behavior*

Through data collection and behavior statistics, the basic information of students' access to the online course is shown as Table 1.

*Table 1.* The basic information of students' access behavior

| Total students | Average number of accesses | Maximum number of accesses | Average number of access days | Maximum number of access days |
|---|---|---|---|---|
| 1088 | 522 | 52632 | 12 | 49 |

To analyze students' information of accessing the online learning platform, we take the "week" as a statistical unit to calculate the access rate, which is equal to the quotient of the total access number in the semester divided by the access number in one week. The result is shown as the left part of Figure 1, which indicates that students accessed the online course with a relatively stable frequency in the first 15 weeks, except for the 1st week and 4th week, which both had extremely low frequencies. From the 15th week, the rate began to rise significantly and reached its peak in the 19th week. Thus, the students' learning behavior was relatively concentrated near the end of the last month of the semester. The access frequency was significantly lower at the beginning of the semester compared with the highest rate at the end of the semester. Therefore, the teacher may intervene with those students who access the online course with very low frequency during the semester. For example, teachers need to urge students to participate in the activities and tasks in the online course in the 5th week.

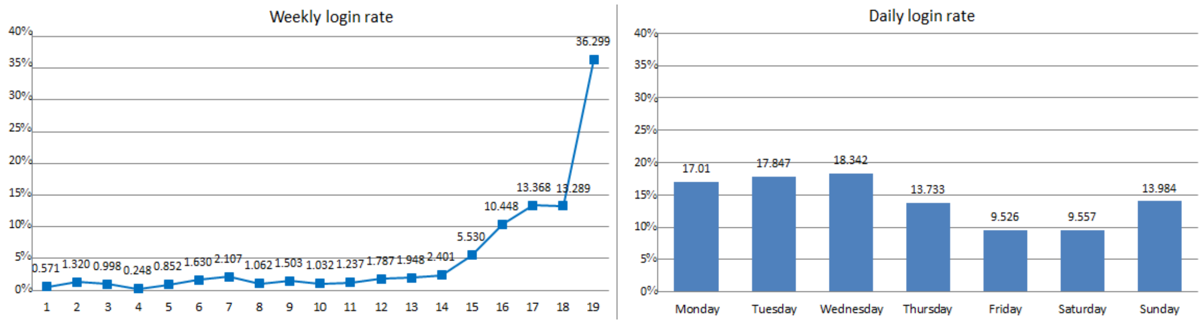


*Figure 1.* Weekly login rate and daily login rate

The analysis result reveals the concentration and dispersion law of the students' learning behavior in each week. This feature is also reflected in a smaller unit of time, such as each day of a week. We calculated the students' access rate for a 7-day week as statistical items, as shown in the right part of Figure 1. Students' participation rates were relatively lower on Thursday, Friday, and Saturday than on the other days. This reflects the characteristics of students' learning behavior on campus to some extent. It seems that they have more enthusiasm in the first three days of the week, and they may be used to completely reviewing and previewing tasks on weekends.

In addition, the regularity of the students' behavior varies in the different periods of a whole day. We divided one day into four time periods, which include 0:00-6:00, 6:00-12:00, 12:00-18:00 and 18:00-24:00. The login rates of these periods are 2.85%, 19.11%, 37.56%, and 40.48%. It indicates that students' online learning mainly occurs in the afternoon and evening. The peak hours focus on the six hours from 18:00 to 24:00. However, there are still a few students who learn online in the period of 00:00--06:00. They may be a relatively small group of hard-working students. The statistics result shows that 64 learners learned in the early morning, of whom 48 got high scores in the comprehensive performance evaluation. Their scores exceeded more than three-quarters of all learners in the course. Most of these students also stayed up late before the final examination. This may manifest that they paid much attention to the tasks of the online course or just dealt with the final examination. Whatever the reason, this special group of students needs attention.

In order to analyze the relationship between online time input and gender, an independent sample *t* test was performed. As shown in Table 2, the mean difference of online input time (minutes) between boys and girls was significant. The mean time of girls was 464.45, and that of boys was 252.98, where $p = .000 < .001$, with statistically significant contributors. The results showed that there was a significant difference in online input time between boys and girls, with girls spending more time online than boys. Further analysis revealed that the average overall score for girls is 9.18 points higher than that for boys.

*Table 2.* The independent sample *t* test on online time input of students of different genders

| Online time input | *N* | Mean | *SD* | *t* |
|---|---|---|---|---|
| Girls | 878 | 463.45 | 244.953 | $11.867^{***}$ |
| Boys | 210 | 252.98 | 158.855 | |

*Note.* $^{***}p < .001$.

The blended course requires students to finish different types of online tasks and offline tasks. The correlation analysis was made to verify the correlation between online input time and the scores of different tasks. The result shows that the five parts of the overall scores are significantly correlated with online input time. Table 3 lists the common scores ($r = .205$, $p < .01$), experiment scores ($r =.206$, $p < .01$), courseware scores ($r =.214$, $p < .01$), micro video scores ($r =.204$, $p < .01$), and online examination scores ($r = .479$, $p < .01$). It indicates that more online input time will result in higher scores.

*Table 3.* The relationship between online time input and students' scores

| | Common scores | Experiment scores | Courseware scores | Micro Video scores | Online Exam scores | Overall scores |
|---|---|---|---|---|---|---|
| Online time input | $.205^{**}$ | $.206^{**}$ | $.214^{**}$ | $.204^{**}$ | $.479^{**}$ | $.446^{**}$ |

*Note.* $^{**}p < .01$.

**Analysis of resource utilization**

The online course is composed of different modules, as shown in Table 4. The page view of each module differs from high to low as follows: quiz, thematic discussion, assignment submission, expanded resource, daily communication, courseware download, and course notification. This order is totally different from that of the layout of each module in the online course. It indicates that the position of resources does not affect the extent of their use.

*Table 4.* Course description of each module

| Module | Pageview | Students | Participation rate |
|---|---|---|---|
| Course notification | 969 | 498 | 45.77% |
| Daily interaction | 3455 | 579 | 53.22% |
| Thematic discussion | 42024 | 1045 | 96.05% |
| Assignment submission | 35568 | 934 | 85.85% |
| Courseware download | 3296 | 670 | 61.58% |
| Expanded resources | 5299 | 719 | 66.08% |
| Quizzes | 303941 | 1011 | 92.92% |

The curriculum teaching content of "Modern Educational technology" contains 11 themes (chapters). The visit of each chapter is shown in Figure 2. As an online learning resource, the degree of utilization of each resource and the importance of the resources to the different students differ greatly. Students' participation is the highest for Topic0 and Topic1, while the participation for the following subjects is significantly lower. The main reason for this phenomenon is that Topic0 and Topic1 are directly related to the classroom learning tasks or experiment tasks, whereas the other resources are not related.

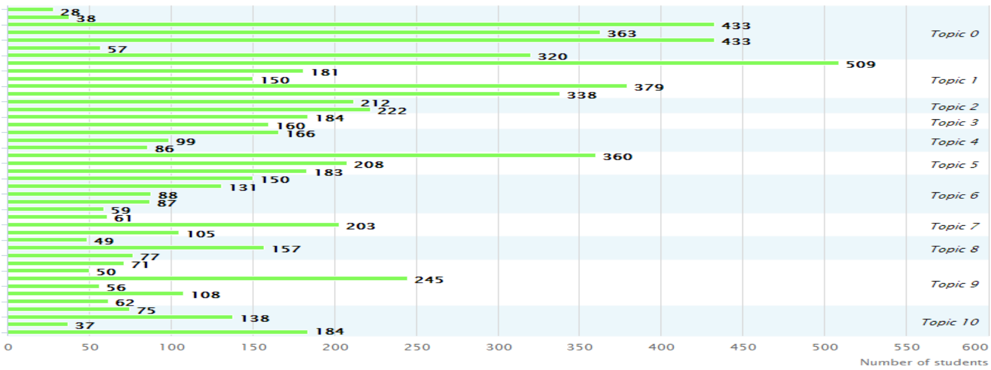


*Figure 2.* Statistics of the number of visits to learning resources with different topics

From the learner's access to the resources of individual cases, the individual behavior of different learners is also very obvious. For analysis purposes, they were divided into a positive group (Access ≥ 25) and a negative group (Access ≤ 5). The average access of the positive group was about 5 times that of the negative group, as shown in Figure 3. Combined with the final scores of the analysis, the resource access is proportional to academic performance.

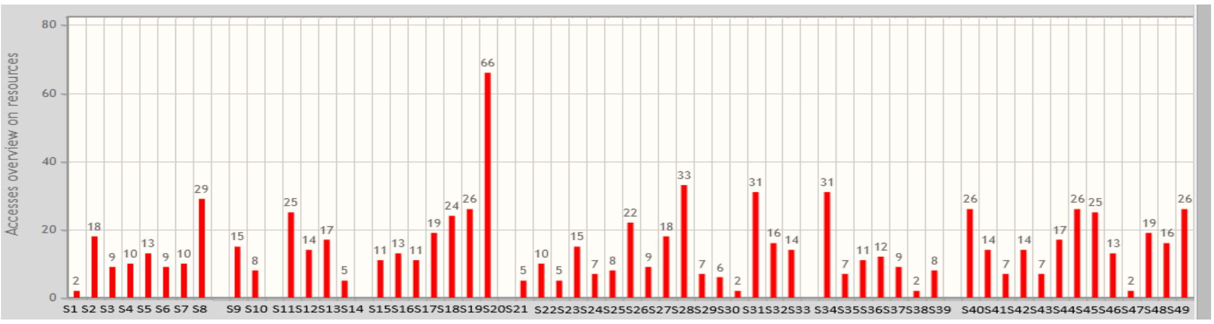


*Figure 3.* Comparison of the frequency of individual access to learning resources
*Note.* S1 means student 1, S2 means student 2, etc.

In addition, there are nine teachers who taught 22 classes in this research. Each teacher taught a different number of classes. Utilization of learning resources in different classes is shown in Figure 4. It was found that there were 11 classes with the highest average score, in which the students were mainly majoring in Chinese language and literature, primary education, humanities education, and English. These classes were taught by six teachers. The average resource utilization rate was close to 60% in these 11 classes. Learning resources are therefore an important influencing factor of online learning.

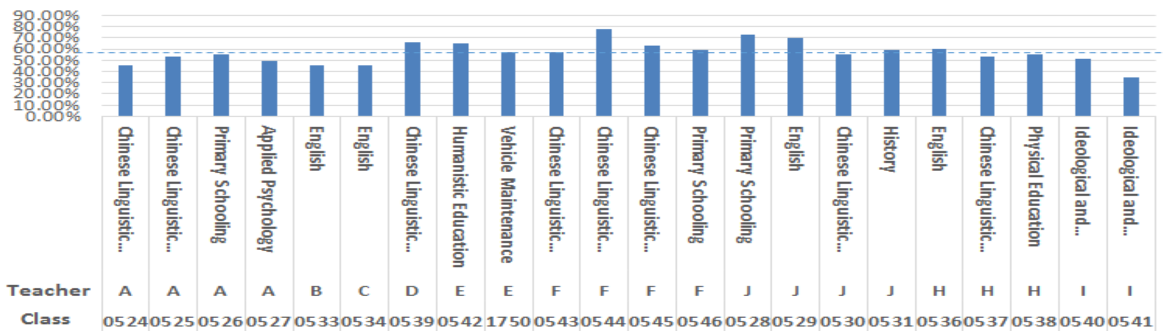


*Figure 4.* Comparison of resource utilization for 22 classes
*Note.* The blue dotted line is the average resource utilization rate of all classes (56.75%).

### Analysis of the quizzes

In order to allow students to strengthen their understanding of the knowledge and to consolidate their knowledge in time, the online course provides several quizzes according to the content of each chapter. Students can receive timely feedback and make up the shortfall through quizzes and repeated exercises. We can see the completion rate of all quizzes from Figure 5. Because this module is closely related to the final examination, its overall completion rate is higher and better than that of other modules.

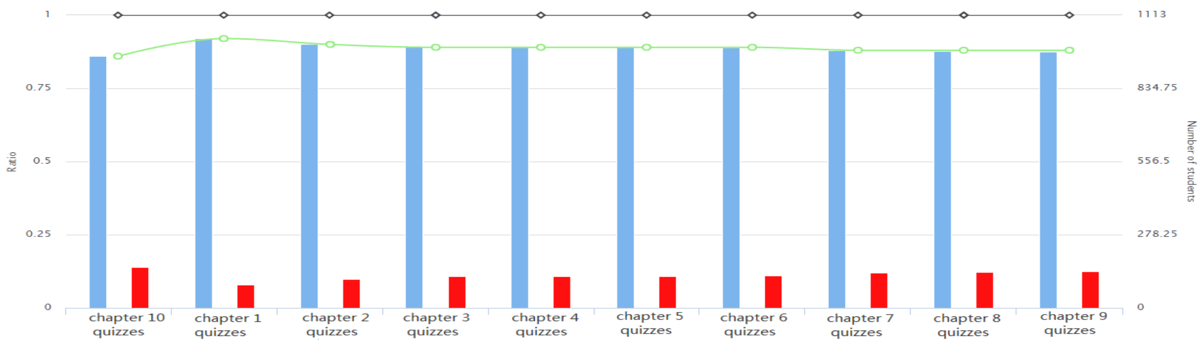


*Figure 5*. Completion of 10 quizzes

**Interactive behavior analysis**

From the time distribution of the discussion and daily communication of the students, the same characteristic of the learning time is presented. As shown in Figure 6, it is mainly concentrated in the last month, when the frequency of discussion significantly increases. Through the analysis of the interactive relationship between teachers and students with social networks, we found diverse and interesting modes of interaction among students and teachers: (1) Single-center mode. It is usually found in the required discussion area in this course. All the members are in the same social network. There are no isolated nodes. A poster (teacher) is located in the center of the discussion. Other members post around the central node, but interaction between different members is rare, as shown in Figure 7-A. (2) Multi-center mode. Different classes of teachers and students carry out free discussions based on different topics in daily communication. They maintain a high degree of enthusiasm and participation that shows frequent interaction. The more active members have a greater number of nodes in the graph and a closer connection, as shown in Figure 7-B. (3) Small-group mode. Some students participate in an optional discussion area, and active individuals become the opinion leaders. Around these leaders, several small groups are set up, as shown in Figure 7-C. (4) Isolated-individual mode. In some optional topics, few students participated. They only expressed their opinion, but did not care about others' opinions. In this mode, each node is isolated from others, as shown in Figure 7-D.

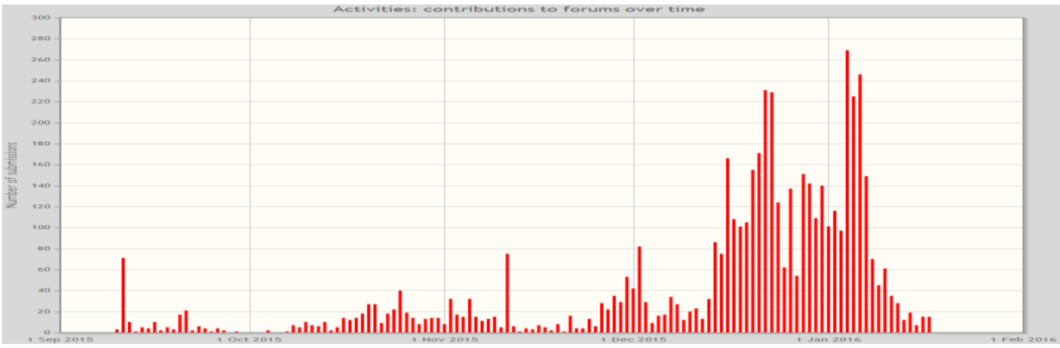


*Figure 6*. Forum access time distribution

**Academic achievement analysis**

Numerous factors influence academic achievement. However, many students are often only concerned about the modules which are directly related to their final grades, and easily ignore some implicit modules. From the learning process evaluation and comprehensive evaluation, the final grade mainly includes normal performance, assignment performance (such as experiment, micro-video, courseware) and the final examination. According to the level of student achievement, the students are divided into three groups: excellent, good and general. More than half of the students got good scores in this course. It is worth noting that the teachers did not set a mandatory requirement for access to the learning resources. However, 56.75% of the students persisted in online learning throughout the whole semester. The extent of students paying attention in different modules is closely related to the requirement of learning tasks and assessment. In addition, most of the students who got the highest page views scored 90% or more. On the contrary, the students with the lowest page views often scored much lower. In some classes, we can find a significant positive correlation between the resource utilization rate and average score. This means that learning resources are an important factor affecting student achievement, but not the absolute and only factor.

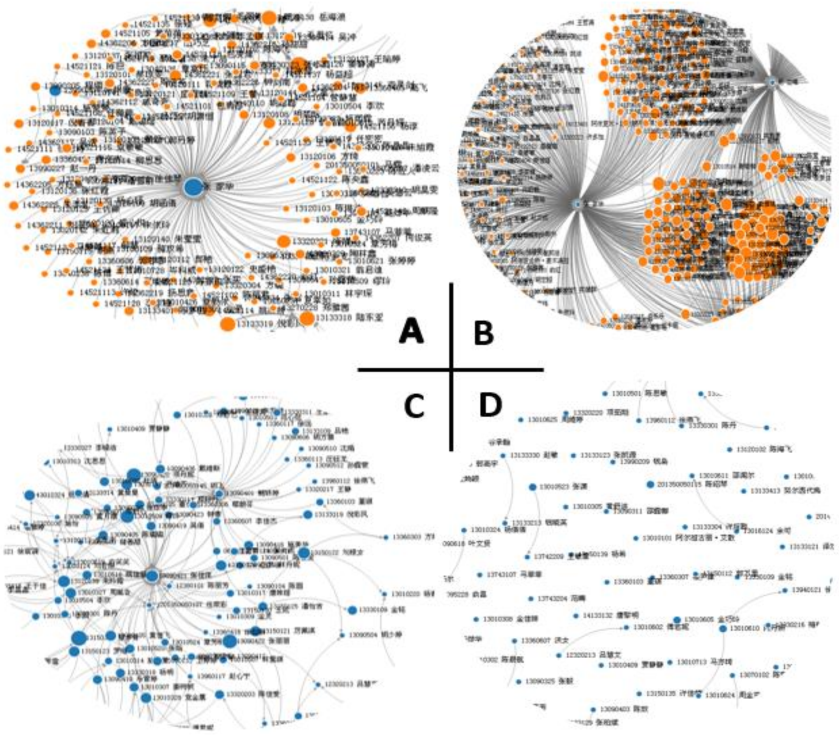


*Figure 7.* Forum interactive modes

In addition, the average grade differed by teacher and class. Even different classes with the same teacher showed significant differences, which is shown in Figure 8. For example, four classes were taught by teacher F, and the total academic achievement of these classes ranked in the top 10 of all classes. They achieved better results because the average resource utilization rate of the four classes was 64.18%, which is higher than the average resource utilization rate of 56.75%. In addition, Teacher F emphasized interaction with the students on the online platform. Teachers and students initiated and replied to more than 400 posts, and each student posted and replied at least once in the discussion area. The teacher set up a good learning atmosphere and promoted students' enthusiasm. Outstanding comprehensive performance is also attributed to the average of more than 85 points of normal performance, assignment performance, experimental performance, and the final grade. Most notably, the normal performance and experiment performance are both more than 90 points. In other words, online learning attendance is affected by many factors, such as teaching methods, learning resources, learning activities, basics of learners, learning atmosphere, etc. However, the most important factor is students' enthusiasm and initiative for online learning tasks and activities.

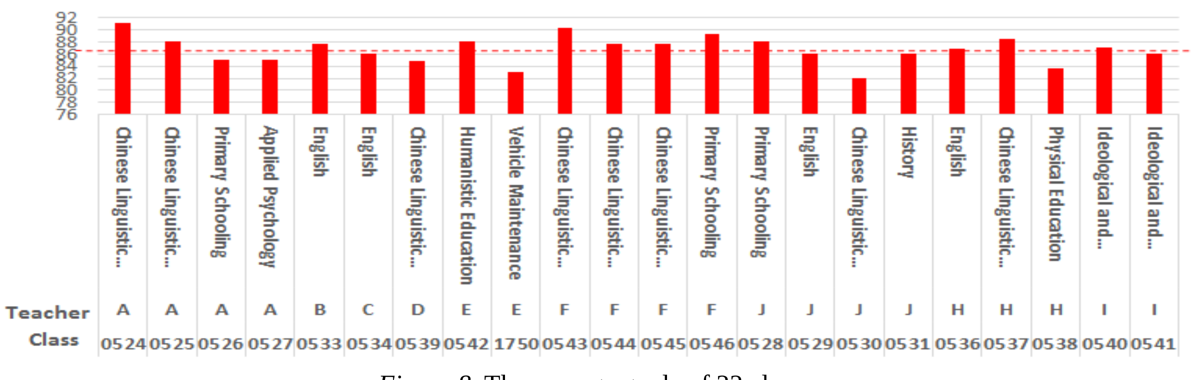


*Figure 8.* The average grade of 22 classes
*Note.* The red dotted line is the grade average score (86.70).

The correlation analysis was carried out to further verify the relationship between the overall score and resource access, quiz attempts, and forum browsing, as shown in Table 5. When the confidence (double sides) between the overall scores and the resource browsing is 0.01, the correlation is obvious. That is, based on the whole tendency, the higher the students' overall scores, the more the amount of resource browsing. The correlation between overall scores and quiz attempts is that when the confidence (double sides) is 0.01, the correlation is obvious, which means that the higher the students' overall scores, the higher the quiz completion rate. The correlation between students' overall scores and forum browsing is obvious when the confidence (double sides) between the overall scores and the forum browsing is 0.05. That is, the higher the students' overall scores, the more the amount of forum browsing, but the correlation is a little weaker.

*Table 5.* Correlation analysis between students' overall scores and the key online behaviors ($N$ = 1088)

| | Resource access | Quiz attempts | Forum browsing |
|---|---|---|---|
| Students' overall scores | $.470^{**}$ | $.362^{**}$ | $.074^{*}$ |

*Note.* $^{*}p < .05$; $^{**}p < .01$.

## Individual diagnosis of online learning

By analyzing the overall learning situation, we can provide personalized diagnostic reports for each student on the basis of the existing data. These reports include the learning process, outcomes, problems, and recommendation through visual presentation for the whole semester. Visual comparison of the results also includes personal and class average grades, so that students can clearly understand their own performance. What's more, they are provided with evidence of learning reflection on the learning platform. The teachers therefore have a comprehensive and detailed understanding of the students' learning status in a data-driven assessment. It can also provide valuable reference for improving teaching methods, curriculum construction, learning resource development, learning activities design, supervision, and intervention of the learning process. At the same time, it provides useful information to optimize the allocation of resources for academic administrators. As shown in Figure 9, a student was selected as a sample to show his personal learning diagnosis report. It provides a clear description of the student online learning process and outcome, which can help the student to know the problems in his/her online learning behavior.

As shown in Figure 9, the student's learning diagnosis report is composed of six aspects:

- **Access Frequency Analysis**. This module is mainly to compare the access frequency of individual students and the average access frequency of all students in the learning platform. As shown in the figure, the student's total access number is 382 times, which is lower than the average access number of 522 times. Therefore, the student was not so active in the learning process during the semester.

- **Resource utilization rate**. This module is mainly to compare the resource utilization rate of individual students and the average utilization rate of all the students in the learning platform. As shown in the figure, the students' utilization rate is only 1.27%, which is far below the average utilization rate of 61.38%. This indicates that the student did not make full use of the learning resources provided in the online course. The student's attention to the learning resources needs to be improved.

- **Completion of quizzes**. This module is mainly to compare the students' completion rate of the quizzes and the average rate of all the students in the learning platform. As shown in the figure, this student's completion rate for the quizzes is relatively poor. In some chapters, he failed to pass the quizzes, and his average score is about 20 points lower than the grade and class average. The student therefore needs to practice more.

- **Forum interaction**. This module is mainly to compare the interaction between the students and the grade in the forum. As shown in the figure, the student's interactions are as follows in the forum: total read times of 39 and reply times of 3. The average read times is 42 and the reply times is 3. The student is thus active in the forum.

- **Score of all kinds of achievements**. This module is mainly to compare all kinds of individual scores and average scores of the class and grade. As shown in Figure 9, the final examination score is about 30 points lower than the average for the grade and the class in all kinds of achievement. However, his usual performance exceeds the grade average.

- **The distribution of comprehensive performance**. This module is mainly to compare a student's score with the comprehensive results of the whole grade and class. For example, the student's comprehensive score is 75, which indicates a general level, but it is lower than the average score of the whole grade and class.

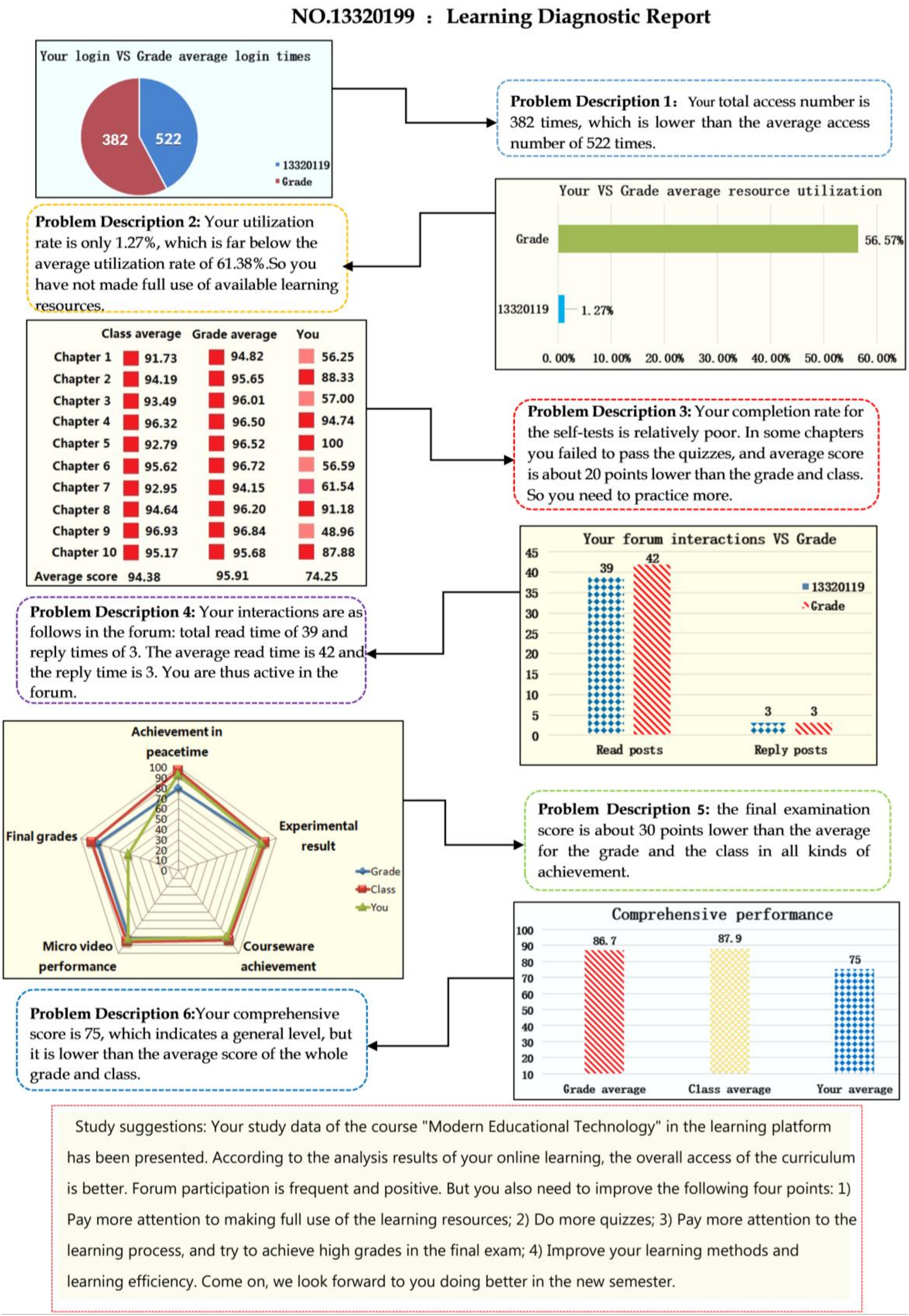


*Figure 9.* A sample individual student learning diagnosis report

Finally, the individual learning diagnosis report shows a detailed understanding of the student's learning situation. Some suggestions are provided for the students from all aspects, which is shown in Figure 9. The individual learning diagnosis report is a visual representation of the state and outcomes of learning. It transforms data into valuable information and helps learners to understand their own learning participation in all dimensions and multiple perspectives, which forms a digital self-analysis and summary.

## Discussion and conclusions

### Characteristics of online learning

Learning analysis can help teachers understand the characteristics of students' online learning behaviors and provide the basis for the adjustment and optimization of follow-up teaching. In this study, in terms of learning time, students' participation frequency grew from slow to fast. At the beginning and mid-term, the attendance was relatively low, but there was a rapid rise at the end of the semester. This indicates that the students' online learning time allocation was unbalanced and unreasonable. In this course, the teachers had no rigid demands regarding students' study time, so the students did not spend much time on the learning platform, especially at the beginning and in the middle of the course. However, they paid more attention to the learning resources and activities which were closely related to the final examination, which led to a sharp increase in student participation at the end of the semester. Therefore, the teachers should actively advocate that the students participate in online learning. If necessary, teachers can also forcibly set the minimum online learning time for students to increase their participation in the course.

From the aspect of learning resources, the resource utilization rate is relatively low. The main reason is that no corresponding intervention has been implemented on students. Therefore, the frequency of learners accessing online resources varies irregularly. For example, the frequency of accessing PowerPoint slides increased dramatically before the end of term. This may indicate that students need to review the content before they finish the quizzes of each chapter. The teachers therefore need to optimize the presentation of resources and enhance the attractiveness of these resources. Also, teachers can establish appropriate monitoring and intervention mechanisms to urge students to effectively use the resources of the online course.

In addition, the Forum Graph plug-in presented the visual results of the online interaction. It indicates that teachers usually act as leaders and serve an important intermediary role in most thematic forums. Most learners can communicate with their classmates and teachers. However, only a few students initiated discussion and exchanged ideas with others. For most students, their interaction in optional tasks were particularly weak. The possible reason is that the participation of teachers in the forum was not high. They were only the initiators of topics and did not pay enough attention to students' performance in the forum. Hernández-García, González-González, Jiménez-Zarco, and Chaparro-Peláe (2015) found that a lack of instructor's activity couldn't ensure better group performance, and the students who got more replies from teachers tended to get higher grades. Therefore, teachers should actively participate in the discussion and pay more attention to students' discussion contents, then give them the appropriate answer in time. In addition, the teachers may select more interesting topics and try active and reasonable strategies to stimulate students' participation.

### Significant indicators of learning achievement

Learning analysis can identify important indicators that affect academic performance, then help teachers make more scientific decisions. In this study, the students' resource utilization, quiz attempts, forum browsing, and overall scores were correlation analyzed by Pearson analysis using the SPSS software. The conclusions are as follows. Students' resource access, quiz attempts and forum browsing showed positive correlations with their overall scores. Among them, the correlation between resource access, quiz attempts, and overall scores was strong, but the correlation between forum browsing and overall scores was relatively weak. This conclusion is somewhat different from that of Bainbridge et al. (2015) who showed that relative number of forum posts and amount of content read were the most important predictors of success in a course. In this study, the reason why the relevance of the forum browsing and overall scores was weak has two aspects. On one hand, it may be due to the fact that the proportion of the number of forum posts in the overall score was small, making students not pay enough attention. On the other hand, it may be due to the lack of interaction between teachers and students, also resulting in weak interaction among students. The four models of forum interaction presented above also predicted this result to a certain extent.

Therefore, optimization of the assessment method and increase in teacher-student interaction are both important measures to improve online learning behavior. Teachers need to establish a comprehensive assessment mechanism, such as uploading and sharing resources. The topics and the influences of forum interaction should also be evaluated, and more attention should be paid to the interaction between teachers and students. Teachers should also promptly reply and appropriately guide students to discuss the topic, urging them with timely feedback. If teachers increase their interaction with the students, it might not only improve the group performance, but could also improve the students' academic achievement. These discoveries above are worthy of teachers' reflection.

### Learning diagnostic reports for individual students

The statistical analysis of student group behavior by using learning analysis is aimed at showing the overall situation of students' online learning. It is helpful for teachers to make the necessary adjustments to their teaching according to the statistical results and to make corresponding demands on students' online learning to improve their learning process. The analysis of group behavior mainly provides the basis for teachers to improve their teaching. The individual learning diagnosis report can provide teachers with a better understanding of individual students' learning situations, then they can provide personalized learning intervention for those students who are at risk. It also provides students with personalized learning feedback, so as to help them obtain personal learning experience. For example, students can know the gap between the whole class and themselves on the learning situation through the diagnostic report, so as to adjust learning and improve academic performance according to the learning suggestions. Some researchers also put forward similar visual online behavior models for learners to improve their awareness of the learning activities, which help to improve their academic achievement (Charleer, Klerkx, Duval, Laet, & Verbert, 2016; Ramos-Soto, Vazquez-Barreiros, Bugarín, Gewerc, & Barro, 2017). Therefore, a personalized learning diagnostic report for individual students is necessary.

### Suggestions to teachers

Teachers should provide diverse and appropriate learning resources. Because of the autonomous characteristic of online learning, the important influencing factors of the academic success of the learners are the learning resources. In order to improve the utilization of learning resources, teachers need to enhance the attractiveness of learning.

Teachers should design appropriate learning and discussion topics. Because online learning time is not fixed and non-realistic, most of the interaction is asynchronous. Teachers need to design discussion topics or encourage learners to issue a discussion topic carefully. In order to attract learners to actively participate in the discussion, good interaction between teachers and students is needed to enhance mutual understanding.

Teachers should publish timely and accurate online learning feedback. Learning feedback enables learners to make clear the current learning progress and learning situation. This is an important basis for adjusting learning methods and learning time in a timely manner. Based on learning analytics, teachers can provide visual feedback of different stages of the learning process to learners, which can facilitate improvement in their learning.

Teachers should adopt effective online learning intervention strategies. Learning analytics can visualize the learning process and the results of the data. According to the results, teachers can be provided with quantitative, objective, and timely feedback. It is possible to predict when students are at risk in their academic studies and to take the necessary learning interventions to reduce their academic risk.

## Acknowledgements

This study is supported by the MOE (Ministry of Education in China) Youth Fund Project of Humanities and Social Sciences (Project No. 17YJC880125).

## References

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270). New York, NY: ACM.

Atif, A., Richards, D., Bilgin, A., & Marrone, M. (2013). Learning analytics in higher education: A Summary of tools and approaches. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conferenc*e (pp. 68-72). Sydney, Australia: Australasian Society for Computers in Learning in Tertiary Education.

Auvinen, T., Hakulinen, L., & Malmi, L. (2015). Increasing students' awareness of their behavior in online learning environments with visualizations and achievement badges. *IEEE Transactions on Learning Technologies*, *8*(3), 261-273.

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, *20*(2), 13-29.

Bainbridge, J., Melitski, J., Zahradnik, A., Lauría, E. J., Jayaprakash, S., & Baron, J. (2015). Using learning analytics to predict at-risk students in online graduate public affairs and administration education. *Journal of Public Affairs Education*, *21*(2), 247-262.

Baker, R. (2016). Using learning analytics in personalized learning. In M. Murphy, S. Redding, & J. S. Twyman (Eds.), *Handbook on personalized learning for states, districts, and schools* (pp. 165-174). Philadelphia, PA: Center on Innovations in Learning Temple University.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An Issue brief. *US Department of Education, Office of Educational Technology*, *1*, 1-57.

Charleer, S., Klerkx, J., Duval, E., Laet, T. D., & Verbert, K. (2016). Creating effective learning analytics dashboards: Lessons learnt. In *Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning* (pp. 42-56). Berlin, German: Springer International Publishing.

Clow, D. (2013). An Overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683-695.

Echeverría, L., Benitez, A., Buendia, S., Cobos, R., & Morales, M. (2016). Using a learning analytics manager for monitoring of the collaborative learning activities and students' motivation into the Moodle system. In *2016 IEEE 11th Colombian Computing Conference (CCC)* (pp. 1-8). Piscataway, NJ: IEEE.

Firat, M. (2016). Determining the effects of LMS learning behaviors on academic achievement in a learning analytic perspective. *Journal of Information Technology Education: Research*, *15*, 75-87.

Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., Dunwell, I., & Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175-1188.

Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A Case study. *Computers in Human Behavior*, *47*, 68-80.

Hwang, G. J. (2014). Definition, framework and research issues of smart learning environments-a context-aware ubiquitous learning perspective. *Smart Learning Environments*, *1*(1), 4. doi:10.1186/s40561-014-0004-5

Hwang, G. J., & Chen, C. H. (2017). Influences of an inquiry-based ubiquitous gaming design on students' learning achievements, motivation, behavioral patterns, and tendency towards critical thinking and problem solving. *British Journal of Educational Technology*, *48*(4), 950-971.

Hwang, G. J., Chu, H. C., & Yin, C. J. (2017). Objectives, methodologies and research issues of learning analytics. *Interactive Learning Environments*, *25*(2), 143-146.

Hwang, G. J., Hsu, T. C., Lai, C. L., & Hsueh, C. J. (2017). Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers & Education*, *106*, 26-42.

Jin, S. H., & Kim, Y. (2016). Learning activities and learning behaviors for learning analytics in e-learning environments. *Educational Technology International*, *17*(2), 175-202.

Kim, D., Park, Y., Yoon, M., & Jo, I. H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education*, *30*, 30-43.

Kim, J., Jo, I. H., & Park, Y. (2016). Effects of learning analytics dashboard: Analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*, *17*(1), 13-24.

Lai, C. L., & Hwang, G. J. (2015). A spreadsheet-based visualized Mindtool for improving students' learning performance in identifying relationships between numerical variables. *Interactive Learning Environments*, *23*(2), 230-249.

Lauría, E. J., Baron, J. D., Devireddy, M., Sundararaju, V., & Jayaprakash, S. M. (2012). Mining academic data to improve college student retention: An Open source perspective. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 139-142). New York, NY: ACM.

Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of "at risk" students using learning analytics: The ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, *64*(5), 957-968.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A Proof of concept. *Computers & Education*, *54*(2), 588-599.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute. Retrieved from http://www.citeulike.org/group/18242/article/9341321

Martin, F., & Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, *13*(3), Article 7.

Martin, F., & Whitmer, J. C. (2016). Applying learning analytics to investigate timed release in online learning. *Technology, Knowledge and Learning*, *21*(1), 59-74.

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160). New York, NY: ACM.

Pistilli, M. D., & Arnold, K. E. (2010). In practice: Purdue Signals: Mining real-time academic data to enhance student success. *About Campus*, *15*(3), 22-24.

Ramos-Soto, A., Vazquez-Barreiros, B., Bugarín, A., Gewerc, A., & Barro, S. (2017). Evaluation of a data-to-text system for verbalizing a learning analytics dashboard. *International Journal of Intelligent Systems*, *32*(2), 177-193.

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Journal of Educational Technology & Society*, *17*(4), 117-132.

Shiri, A. (2016). Learning analytics: supporting teaching and learning through learner's data analytics and visualization. In *Data Analytics 2016: The Fifth International Conference on Data Analytics*, 126-130.

Siemens, G. (2010). *What is learning analytics*. Retrieved from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Sun, J. C. Y., Lin, C. T., & Chou, C. (2016). Applying learning analytics to explore the influence of online learners' motivation on their online learning behavioral patterns. In *2016 5th IIAI International Congress on Advanced Applied Informatics* (pp. 377-380). Piscataway, NJ: IEEE.

Leeuwen, A. V., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education*, *90*, 80-94.

Zapparolli, L. S., & Stiubiener, I. (2017). FAG-a management support tool with BI techniques to assist teachers in the virtual learning environment Moodle. *Advances in Science, Technology and Engineering Systems Journal*, *2*(3), 587-597.

Zhang, Y., Fei, Q., Quddus, M., & Davis, C. (2014). An Examination of the impact of early intervention on learning outcomes of at-risk students. *Research in Higher Education Journal*, *26*, 1-12.