

# **A Teaching Quality Evaluation System of Massive Open Online Courses Based on Big Data Analysis**

<https://doi.org/10.3991/ijet.v14i14.10818>

Zhifang Wang (✉), Jia Liu

Hebei University of Science and Technology, Shijiazhuang, China

48150335@qq.com

**Abstract**—Massive open online courses (MOOC) transcends the time and space limits of traditional classroom teaching, and promotes the sharing of teaching resources. However, the effect of this emerging teaching mode is yet to be determined. In this paper, the big data analysis is introduced to evaluate the MOOC teaching quality. Taking several online courses as an example, a video player was designed to compute the learning time using the Hadoop platform. On this basis, the author constructed a teaching quality evaluation platform. In addition, the learning cost coefficient was calculated by the naive Bayesian model, and the evaluation results were analysed in details. The research findings shed practical new light on the evaluation of MOOC teaching quality.

**Keywords**—Big data, cloud computing, massive open online courses (MOOC), teaching quality evaluation system

## **1 Introduction**

Along with the rapid development of computer technology and Internet technology, online education has become a very popular teaching method in today's society. It breaks the limitation of traditional classroom teaching by time and space, and realizes the sharing of teaching resources and balance of education, which provides a platform for lifelong learning. This is also in line with development concept for building a learning society in China [1]. In this context, universities and related enterprises have increased the research and development of online teaching platforms, to continuously improving the construction of teaching resources and curriculum resources, and the related technologies have been more mature. But the teaching quality and learning effects of online education are widely questioned by the society. Therefore, it has become a bottleneck for the further development of online education on how to test the course learning, and evaluate the teaching quality and the learning effect [2].

Teaching quality evaluation is one of the important contents for teaching management [3]. At present, there are two evaluation methods of teaching quality: qualitative evaluation and quantitative evaluation, and four main evaluation systems such as system of attending a lecture, supervision system, teaching inspection system, and student evaluation system [4]. In most universities, a teaching quality evaluation form is formulated by the teaching administration staff for the students to conduct online

evaluation of teachers' teaching quality in the mid-term or end-term of the semester, and then the administration staff combines the test scores of students to comprehensively assess and grade the teaching quality [5]. However, this student-oriented evaluation method has strong subjectivity, lacking analysis for relevant teaching data, and failing to play a guiding role in teaching [6]. In the students' learning process through the MOOC platform, a large number of browsing, communication, testing and other learning-related data shall be generated. If these data can be deeply explored and integrated, the results of the teaching quality evaluation will be more objective, and closely aligned with reality [7]. Big data [8] is generated along with the increasingly popular network behaviour. It is a large data group that is collected through multiple channels and exists in multiple forms. Its main features include huge data volume, various types, low value density, high processing speed, and authenticity. The analysis and storage of massive data is the core value of big data [9]. The purpose of data mining is to "purify" useful information from massive data [10], but a single computer can't handle such huge data. For this, cloud storage, distributed database, distributed processing technology based on cloud computing can meet the needs of big data processing [11].

Based on the above analysis, this paper briefly introduces the cloud computing-based big data processing architecture and Hadoop platform. Taking the online courses of MOOC platform as an example, Hadoop was used to connect with the MOOC platform, and design a video player capable of the duration calculation. Then, a teaching quality evaluation algorithm model was established. The naive Bayesian model was used to solve the learning cost coefficient, and analyse the data screening and specific results.

## **2 Platform Construction**

### **2.1 Cloud computing-based big data processing architecture**

For easy understanding, this paper maps the application technology to the seven-layer network protocol of open system interconnect (OSI) reference model using the Hadoop technology, and then divides the cloud computing-based big data processing from bottom to top into six layers: data integration layer, file storage layer, data storage layer, programming model layer, data analysis layer, and platform management layer [12]. Figure 1 shows the related architecture.

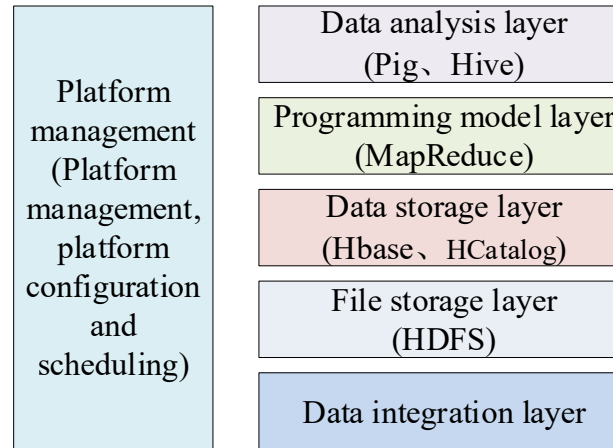


Fig. 1. Cloud computing-based big data processing architecture

**Data integration layer:** Various structures and types of data that the system needs to process are concentrated in the data integration layer, which can be called and processed by MapReduce, or directly stored in Hadoop distributed file system (HDFS).

**File storage layer:** It plays a role of linking up and down, which can provide the access service of bottom layer to the upper layer through the unified interface, and access the data of the storage layer downwards for the efficient access to massive files by using distributed parallel technology.

**Data storage layer:** The data storage layer requires fast reading and writing of massive data under low conditions, to realize the management capability of big data tables. HCatalog and Hbase are two Hadoop-based technical foundations to support data sharing operations of upper layers such as MapReduce and Pig etc.

**Programming model layer:** It is the core part of the whole processing architecture, providing a programming and running environment for large-scale data processing. MapReduce is dominant in cloud computing-based big data processing because of its efficient and concise algorithm.

**Data analysis layer:** The data analysis layer can provide advanced tools for improving the reading speed of data results, such as Hive and Pig in Hadoop.

**Platform management layer:** Security management, operation monitoring, configuration management, etc. are the main components of the platform management layer, with the purpose to ensure the safe and stable operation of data processing platform.

## 2.2 Hadoop platform

**Hadoop composition:** Hadoop is a distributed system infrastructure. Figure 2 shows the Hadoop architecture. HDFS and MapReduce are the core design of the Hadoop architecture [13].

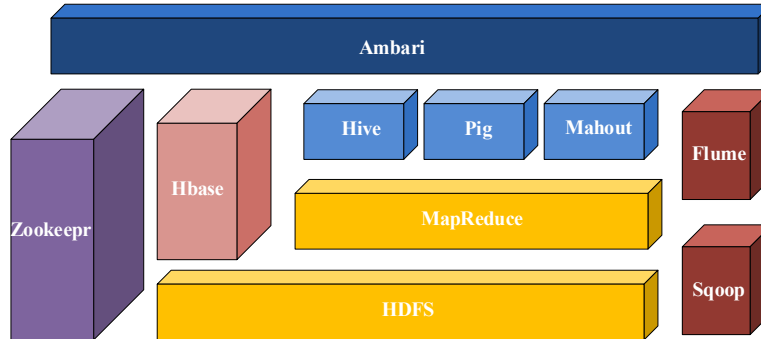


Fig. 2. Hadoop architecture

**HDFS:** The Hadoop distributed file system (HDFS) [14] includes components such as Name Node, Secondary NameNode, DataNode, and Client. The Name Node is responsible for the namespace of the file system, playing the role of a manager in the system; the Secondary NameNode is a system standby node for periodically back up data and avoid data loss due to system failure; the DataNode is the node that stores data in the system; Client is the user of the system and can obtain the corresponding data by directly accessing the DataNode.

**MapReduce:** MapReduce [15] is a software framework that can process large data sets in parallel, including at least three parts: Map, Reduce, and Main functions. The Map function converts the received data into a list of key/value pairs; Reduce processes and outputs the data to obtain the final result; the main function combines file input/output with job control.

**Hadoop platform construction:** In this paper, the Hadoop architecture was regarded as a big data processing environment to analyse the big data generated in MOOCs. In order to set up a Hadoop platform, Cygwin software for virtual Linux environment was first downloaded and installed, followed by Hadoop software. Figure 3 shows Hadoop installation interface.

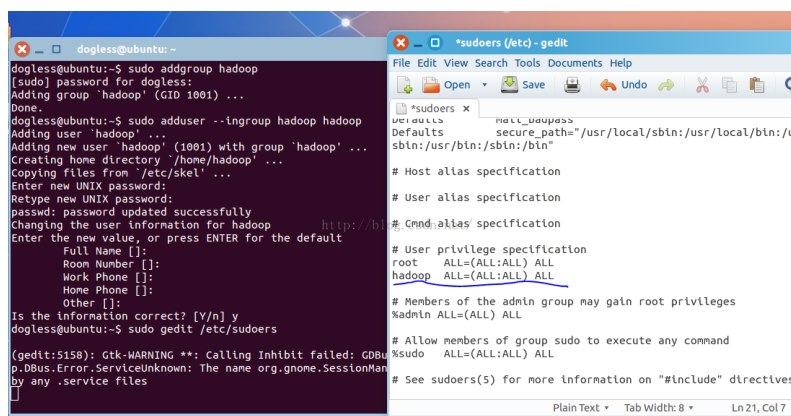


Fig. 3. Hadoop installation interface

### 3 MOOC Teaching Quality Platform Evaluation System based on Big Data Analysis

#### 3.1 Big data analysis process for MOOC teaching quality evaluation

Massive data related to learning content, exercises, tests, questions and answers are generally generated in the MOOC teaching and learning process. Therefore, when selecting teaching evaluation data, firstly, it is necessary to identify and determine the useful data in video learning, exercises or homework accuracy related to knowledge points; then the relevant data is mined, and converted into operable data through the relevant software, attempting to establish the corresponding relationship; afterwards, the information re-obtained through data mining is verified; finally the teaching quality evaluation results are obtained; the process of big data analysis for MOOC teaching quality evaluation is shown in Figure 4.

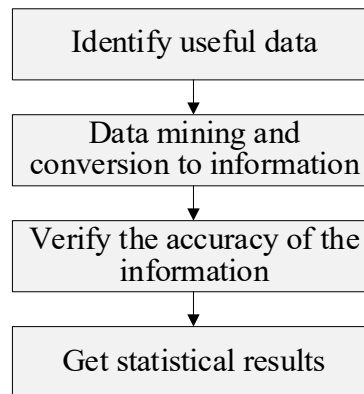


Fig. 4. Open classroom teaching quality evaluation big data analysis process

#### 3.2 Data acquisition and reading

In order to evaluate the MOOC teaching quality more objectively and accurately, this paper takes the learning time of students watching video as the main evaluation indicator of teaching quality, and divides the behaviour of students' online learning video into video playback, fast forward times and total learning time. Besides, a simple questionnaire was conducted at the end of the course to reflect the teaching quality through data statistics and mining. In this study, using the ActionScript language, the author developed a FLASH video play software that can accurately record the learning time, applied it to the open MOOC video playback, and established a link with the database.

### 3.3 Algorithm model of teaching quality evaluation

**Assumption of teaching quality evaluation:** It's assumed that A and B are two groups of students who learn the same course content, and spend the same learning time, but the correctness rates of answering the after-class questions are different. All these were taken as important evaluation indicator for teaching quality, for which "1" indicates correct answer and 0 is wrong answer, while the learning time is recorded by the video software. Table 1 lists the database design for the students' video learning during the learning process, where ID is the serial number of students generated after they logged in to watch the video.

**Table 1.** Database structure design

ID	Fast forward times	Start time	End Time	Number of playbacks	Start time	End Time	Knowledge point 1	Knowledge point 1
24	1	85	114	1	258	204	1	—
	2	194	237	2	356	299		—
	3	288	364	—	—	—	0	—
	4	427	450	—	—	—		—

**Algorithm model design of teaching quality evaluation:** In order to further mine the useful data from massive data for teaching quality evaluation, this paper takes learning time (correct, wrong learning time), problem correctness rate and learning cost coefficient as analytical indicators.

The learning cost coefficient is the ratio of the time taken by the students to learn and master a certain knowledge point (based on correctly answering the corresponding exercises) and the duration required to explain the knowledge points. This paper uses the naive Bayesian model to solve the learning cost coefficient.

Bayes algorithm is a simple, fast and accurate classification one that can be applied to large databases. Naive Bayes is comparable to neural network classification algorithms and decision trees to some extent. The Bayes theorem is shown as:

$$P(H|X) = \frac{P(X|H)}{P(X)} \tag{1}$$

Based on the actuality in this study, the variables in formula (1) are defined as:

$$P(H) = \frac{\text{The exercise answers the correct number of times}}{\text{The number of answers to this exercise}},$$

$$P(X) = \frac{\text{Average length of video watched}}{\text{Total length of the knowledge point}},$$

$$P(X|H) = \frac{\text{Answer the correct viewing time}}{\text{Total length of the knowledge point}},$$

$$P(H|X) = \text{Predicted time cost for mastering knowledge points.}$$

Figure 5-7 shows the data analysis of the learning cost coefficient, the correctness rate of the exercises and the learning time of the four knowledge points included in the classroom teaching video 1.

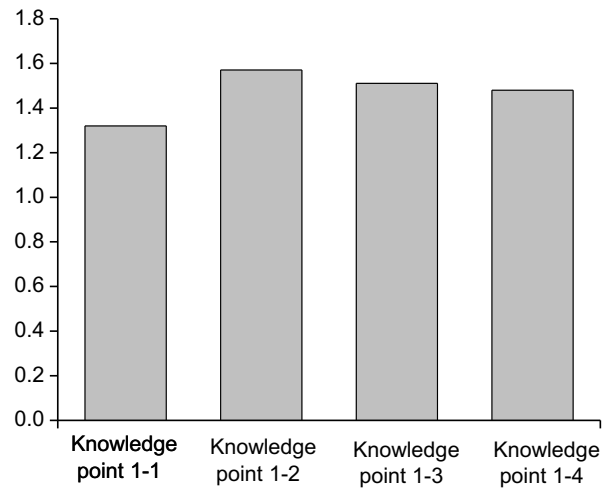


Fig. 5. Learning cost coefficient data analysis summary

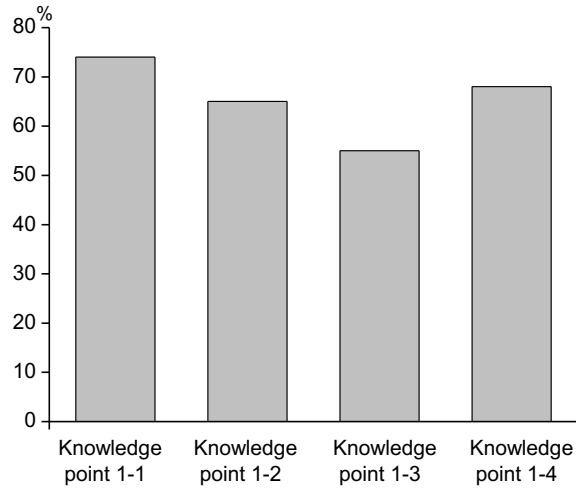


Fig. 6. Summary of problem correctness data

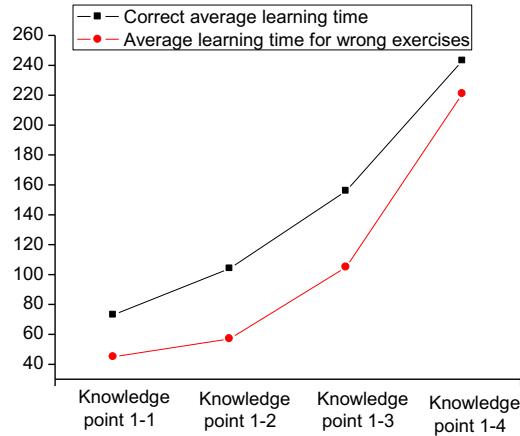


Fig. 7. Data analysis summary of learning time

**Data screening:** Among the massive data generated in the students' learning process, some are invalid false data. For instance, teachers deliberately reduce the problem difficulty in order to obtain a good evaluation result, or the student's learning attitude is not serious, etc., which will affect the student's learning cost coefficient. Thus, the evaluation results cannot reflect the true teaching quality. To obtain more realistic learning data, this paper performs data screening by excluding teachers' invalid exercises and students' invalid learning behaviour records.

*Excluding teachers' invalid exercises:* In the evaluation of teaching quality, the data with the over-high average learning cost coefficient, over-low problem correctness rate are deleted, since these invalid data are deviated from the data centre point with high dispersion degree.

*Excluding the student's invalid behaviour record:* Generally, the statistic value of the learning time required for students to correctly answer the exercises should be consistent with the normal distribution function. As shown in Figure 8, when the students quickly complete the study by fast-forward and answer the questions indiscriminately, or after interruption for a certain time, the students spend longer learning time in re-playing the video and complete the learning, all these data collected are invalid, and should be excluded. The confidence intervals of different knowledge points may not be the same, so it's necessary to select the confidence intervals of valid data by combining with the different knowledge points.

*Data analysis results:* Substituting the learning cost coefficient obtained by data screening into the expectation formula, the expected learning cost coefficient was derived. The closer this value is to 1, the higher the teaching quality of the course. In this study, the expected value of the selected experimental course is 0.88, indicating a higher teaching quality. The expected value of the same course taught by different teachers may be the same, but the variance is different. The variance can reflect the average difficulty of the course; the higher variance indicates a great difference in the difficulty level of the knowledge points, and there exists the polarization. Figure 9 shows the expected variance of the two different teachers in the same course.



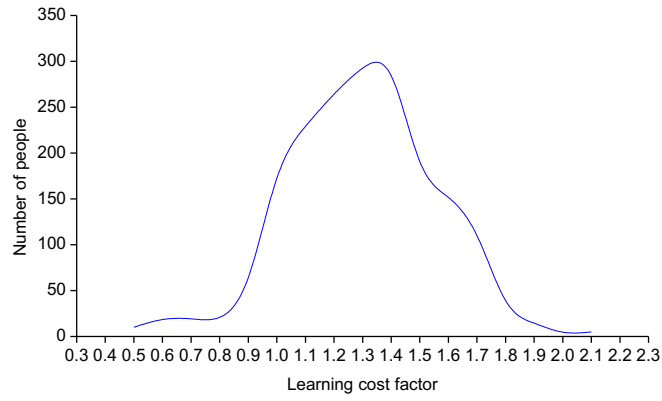
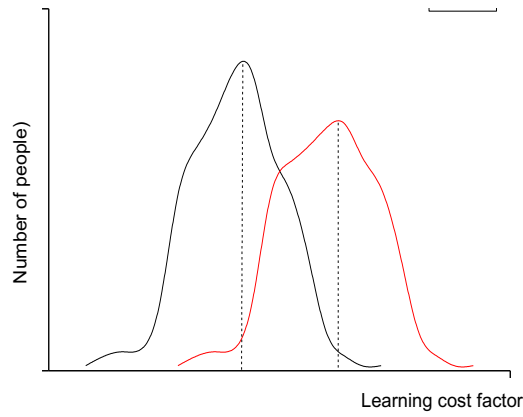
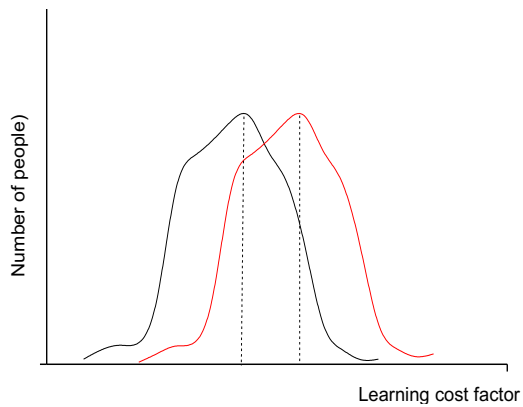


Fig. 8. Data analysis field summary



(a)



(b)

Fig. 9. Expected variance of two instructors in the same course

## 4 Conclusion

In order to promote the continuous development of online education, this paper uses big data analysis technique to conduct research on the MOOC teaching quality evaluation system. The specific conclusions are as follows:

- A cloud computing-based big data processing architecture and a Hadoop platform were constructed, which is connected with the MOOC platform as a user interface for data collection.
- The students' learning time for watching the teaching video and the correctness rate of the problems were taken as the main indicators of the teaching quality evaluation. For this, the video player that can record and calculate the student's learning time was developed.
- The teaching quality evaluation algorithm model was designed; the naive Bayesian model was applied to solve the learning cost coefficient, and analyse the data screening and specific results.

## 5 References

- [1] Iezzi, D.F. (2005). A method to measure the quality on teaching evaluation of the university system: the Italian case. *Social Indicators Research*, 73(3): 459-477. <https://doi.org/10.1007/s11205-005-2923-6>
- [2] Lalla, M., Facchinetti, G., Mastroleo, G. (2005). Ordinal scales and fuzzy set systems to measure agreement: an application to the evaluation of teaching activity. *Quality and Quantity*, 38(5): 577-601. <https://doi.org/10.1007/s11135-005-8103-6>
- [3] Looney, J. (2011). Developing high-quality teachers: teacher evaluation for improvement. *European Journal of Education*, 46(4): 440-455. <https://doi.org/10.1111/j.1465-3435.2011.01492.x>
- [4] Palmer, S. (2012). The performance of a student evaluation of teaching system. *Assessment & Evaluation in Higher Education*, 37(8): 975-985. <https://doi.org/10.1080/02602938.2011.592935>
- [5] Kember, D., Leung, D.Y.P., Kwan, K.P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education*, 27(5): 411-425. <https://doi.org/10.1080/0260293022000009294>
- [6] Webbstock, D. (1999). An evaluative look at the model used in the assessment of teaching quality at the university of natal, south Africa: reflections, rewards and reconsiderations. *Assessment & Evaluation in Higher Education*, 24(2): 157-179. <https://doi.org/10.1080/0260293990240206>
- [7] Pounder, J.S., Hung Lam, E.H., Groves, J. (2016). Faculty-student engagement in teaching observation and assessment: a Hong Kong initiative. *Assessment & Evaluation in Higher Education*, 41, 1-13. <https://doi.org/10.1080/02602938.2015.1071779>
- [8] Liu, Z., Yang, C., Rho, S., Liu, S., Jiang, F. (2017). Structured entropy of primitive: big data-based stereoscopic image quality assessment. *IET Image Processing*, 11(10): 854-860. <https://doi.org/10.1049/iet-ipr.2016.1053>
- [9] Gu, K., Tao, D., Qiao, J.F., Lin, W. (2018). Learning a no-reference quality assessment model of enhanced images with big data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(4): 1301-1313. <https://doi.org/10.1109/tnnls.2017.2649101>
- [10] Broughman, J.R., Chen, R.C. (2016). Using big data for quality assessment in oncology. *Journal of Comparative Effectiveness Research*, 5(3): 309-319. <https://doi.org/10.2217/cer-2015-0021>

- [11] Barrie, S., Ginns, P., Prosser, M. (2005). Early impact and outcomes of an institutionally aligned, student focused learning perspective on teaching quality assurance1. *Assessment & Evaluation in Higher Education*, 30(6): 641-656. <https://doi.org/10.1080/02602930500260761>
- [12] Grammatikopoulos, V., Linardakis, M., Gregoriadis, A., Oikonomidis, V. (2015). Assessing the students' evaluations of educational quality (seeq) questionnaire in greek higher education. *Higher Education*, 70(3): 395-408. <https://doi.org/10.1007/s10734-014-9837-7>
- [13] Perry, A., Prichard, E.A., Penn, H.E. (2010). Indicators of quality teaching in intensive behavioral intervention: a survey of parents and professionals. *Behavioral Interventions*, 21(2): 85-96. <https://doi.org/10.1002/bin.212>
- [14] Watts, B., Sarah, L., Renée H. (2009). Teaching quality improvement in the midst of performance measurement pressures: mixed messages? *Quality Management in Health Care*, 18(3): 209. <https://doi.org/10.1097/qmh.0b013e3181aea266>
- [15] Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, 36(1), 51-62. <https://doi.org/10.1080/02602930903197883>

## 6 Authors

**Zhifang Wang**, female, born in April 1982, from Xingtai, Hebei, lecturer (specialized in computer science), graduated from Hebei University of Technology, majored in pattern recognition and intelligent systems, and now works in the teaching management department of Hebei University of Science and Technology. I specialize in big data analysis and research, and have been engaged in education and teaching management for many years. I have served as a lecturer in "Java Programming", "C Language Programming", "Computer Basics", "Web Design" and other courses, and have published 3 papers in related professional fields.

**Jia Liu**, female, born in June 1984, from Shijiazhuang, Hebei, lecturer (specialized in ideological and political education), graduated from Hebei University of Science and Technology, majored in medicinal chemistry. I had served as the director of the international exchange center of institute of technology for Hebei University of Science and Technology, and now I am the secretary for the second league branch of engineering of the institute of technology of Hebei University of Science and Technology. I specialize in international talents training, international curriculum construction, and college students ideological and political education management. I have been engaged in education and teaching management for many years and have taught courses such as "Ideological and Moral Cultivation and Legal Basis", "College Students Career Planning and Employment Guidance", etc. I have led and participated in 10 provincial and departmental projects and published more than 20 papers and a monograph.

Article submitted 2019-05-17. Resubmitted 2019-06-24. Final acceptance 2019-06-28. Final version published as submitted by the authors.