The Application of Automatic Teaching Evaluation System Based on UTOP and ST Analysis

Qian Zhang, Jian Liao, Geping Liu, & Xiaolang Chen

DOI:10.59668/1269.15634

AECT
THE JOURNAL OF APPLIED INSTRUCTIONAL DESIGN

Teaching evaluation is an effective way to improve the quality of teaching. Nevertheless, traditional teaching evaluation has some problems, such as high labor costs, low efficiency and, based on experience or standards, etc. As well, EAIT has been highly developed in recent years. In this study, the automatic teaching evaluation system is adopted. Through focus group interviews, a comprehensive investigation is conducted on the views and improvement of teachers of various subjects in primary school on automatic teaching evaluation.

Introduction

Research shows the most effective teacher professional development has the following characteristics: expect teachers to examine and reflect on their own practice (Putnam, 2000; Richardson, 1994). The assessment of teaching can support after-class reflection and practice improvement. However, traditional assessment activities have developed some problems that need to be improved, such as the evaluation of lessons based on experience and standards, the lack of evidence-based opinions, etc. (Fang J.2014).

At present, artificial intelligence is gradually integrated into the reform of education evaluation, and a series of Educational Artificial Intelligence Tools (EAIT) have appeared, which can support the automatic teaching evaluation system (AET-system) of elementary education teachers' teaching quality. The quantitative evaluation of classroom teaching behavior reduces the influence of evaluators' subjective experience (Haudek, K.C, 2020). Simultaneously, automated collection and analysis decrease labor costs and can promote the improvement of teaching quality (Kashyap, M.C, 2018).

This paper primarily applies the AET system in experimental schools and analyzes teachers' views through questionnaire surveys and focus group interviews. We will explore the trend of teaching assessment in the future from the advantages and disadvantages of automated teaching evaluation, hoping to promote the development of classroom evaluation in China.

Literature review

Development of assessment tools

The traditional classroom teaching evaluation in China mainly includes two parts. First of all, the evaluator will listen to the teacher's whole teaching process to understand the characteristics of the teacher's teaching. Secondly, teachers should be evaluated, including evaluating teachers' success or failure in classroom teaching and giving reasons. At the same time, evaluators should also analyze the problems in the classroom and propose solutions (Lei Li, 2015). Teaching evaluation can improve teachers' ability to solve practical classroom teaching problems (Guanghua Shao, 2004). Absolutely, the evaluation tools have also been advanced in the development of technology, which has gone through two stages (see Table 1).

The traditional paper and pencil recording stage mainly relies on manual observation of the classroom record after class analysis. Labor costs are high, and the dimensions of records are not comprehensive. There are primarily seating chart observation records (SCORE) (Acheson, 1987), the FIAS interactive analysis system (Flanders, 1970), and the S-T analysis method (Yingjie Shan, 2008).

In the stage of multimedia technology recording, classroom teaching behavior is mainly recorded with the assistance of experts, a camera, or portable equipment. After each dimension is quantified by computer, it is analyzed manually or with the help of big data. For example, an interactive analysis coding system (ITIAS) based on information technology (Xiaoqing Gu, 2004) and wearable sensors record teaching behavior (Prieto, 2018).

It can be seen that China's current teaching evaluation still has the following deficiencies: high labor cost and low efficiency; evaluation data is not easy to keep; class evaluation is based on experience and criteria. Nowadays, teaching evaluation is turning to “subjective experience” and “data.” The AET-system not only provides data but also supports the reasonable selection of expert comments. Meanwhile, it greatly decreases evaluating costs and changes the traditional pattern.

Table 1

Development of evaluation tools

Stage	Method	Record	Content	Disadvantage
Traditional	SCORE	Manual	teacher-student interaction	high labors; not comprehensive
	FIAS	Manual and Computer	teacher-students interaction	
	S-T	Manual and Computer	teacher-students interaction	
Multimedia	ITIAS	Computer	interaction	High costs Poor efficiency
	Wearable device	Computer	Voice and physiological data	Not comprehensive

AI-system for Automated Teaching Evaluation

The AET-system mainly uses voice and text analysis technology based on the classroom observation evaluation framework UTOP (UTeach Observation Protocol) and S - T class teaching method to automatically analyze the classroom.

After the automatic evaluation of a lesson, the AET-system will automatically generate a report. The evaluation report can be seen in Figure 1. In addition, various dimension indexes and explanations based on the UTOP and the ratio of S-T behavior will be presented. The system will also present the visualized graphs of teacher intervention times, student participation activities, high-frequency words, immediate evaluation, exploratory questions, resource application, the proportion of content asked, the proportion of teacher-student interaction, the proportion of teacher-student behavior, etc.

Figure 1

AI system evaluation report

二、课堂教学行为数据概览

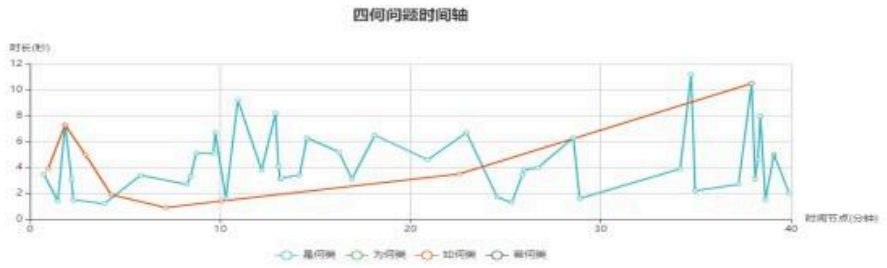
1.课堂环境 Classroom Environment

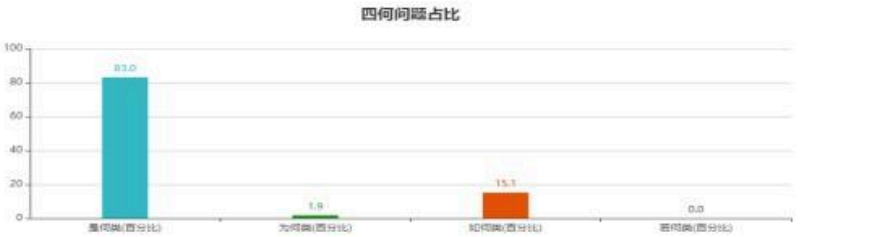
依据	证据或关键点	数据或信息	指数	说明
课堂互动	小组活动或展示(百分比)	88.9	6.1	学生之间通过合作解决问题。
	实验实践(百分比)	11.1		
课堂干预	教师课堂管控(百分比)	10.0	9.6	教师的管理指令清楚有效，学生能保持良好的课堂纪律。
	教师观察或巡视(百分比)	90.0		
课堂参与	学生参与度（百分比）	93.1	9.3	创造良好的课堂氛围，让学生提出想法、问题等。
课堂对话	学生机械性回答(百分比)	28.7	8.5	课堂检查、活动关注。
	学生认知记忆性回答(百分比)	49.5		
	学生推理性回答(百分比)	19.8		
	学生创造性回答(百分比)	2.0		

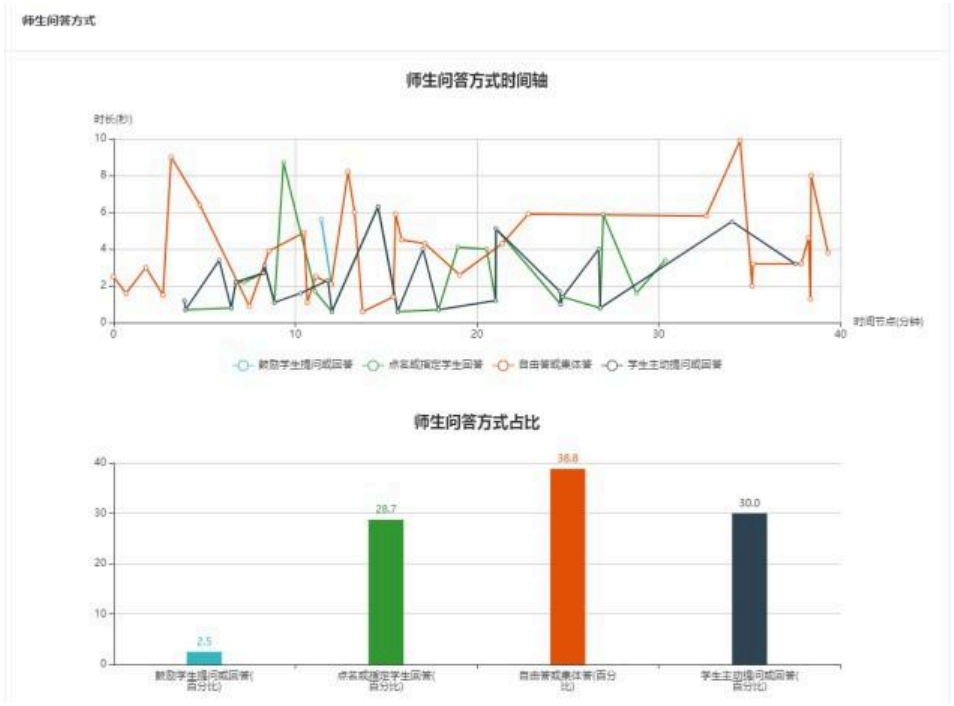
四、4MAT分析

1.四何问题

四何问题







Research Questions

The purpose of this study is to empirically examine teachers' perceptions of accepting an automated lesson evaluation system. Specifically, three research questions are set for achieving the research goal:

RQ1: What are the advantages of using an automated teaching evaluation system in teaching?

RQ2: What are the disadvantages of using an automated teaching evaluation system in teaching?

RQ3: What should the automated teaching evaluation system improve in the future?

Method

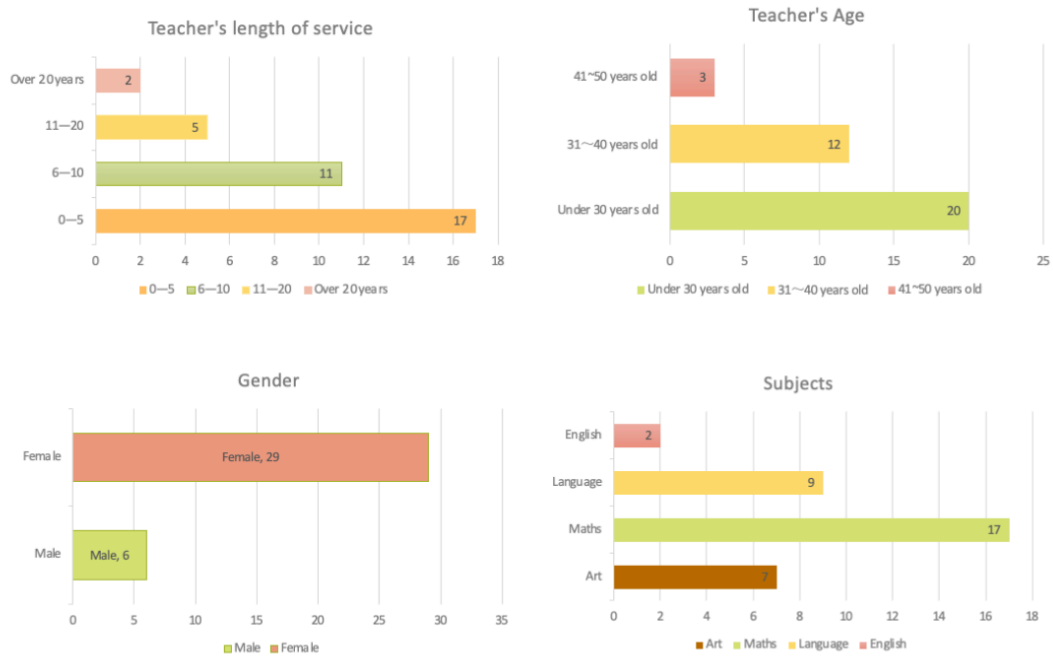
Participants

Prior to the experiment, all participants had signed informed consent forms, agreeing that their information would be used for research purposes. The current study involved 35 teachers, including 83% female (n = 29) and 17% male (n = 6). The English teacher, Language teacher, Math teacher, and art teacher were 2, 9, 17, and 7, respectively. Detailed demographic characteristics of the participants are presented in Figure 2.

Eight people were randomly selected to participate in interviews. When the eight participants entered the classroom, they were told the purpose and topic of the interview. The survey took about 1.5 hours.

Figure 2

Experimenter related information



Procedure

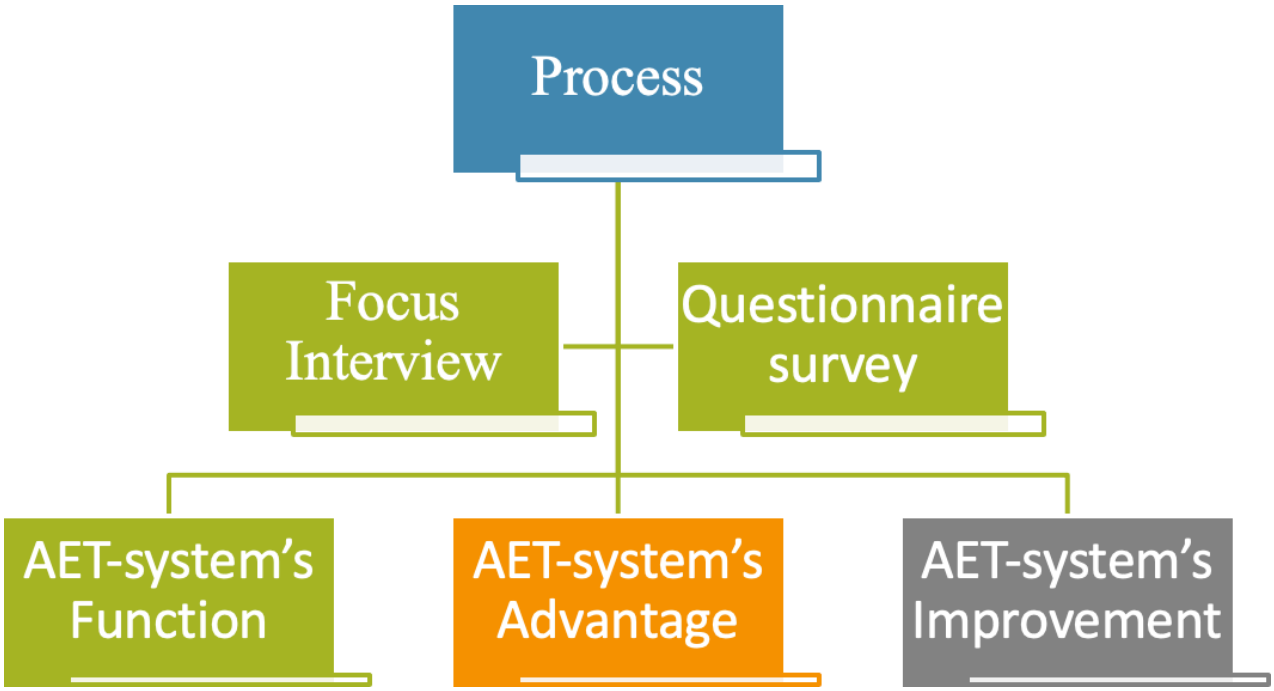
This study will introduce participants to the automated teaching evaluation system, including how the system operates, its functions, and the results of system analysis. Use the system to evaluate the teacher's classroom teaching and present the analysis results to the teacher. It mainly includes classroom environment, lesson structure, implementation, and content, as shown in Figure 1. After the teachers had checked the evaluation results, questionnaires were issued to investigate the teachers who had participated in the evaluation. And randomly selected teachers to conduct interviews. The main experimental process is shown in Figure 3.

The questionnaire was designed to be completed independently by participants using their personal electronic devices. Participants were estimated to spend approximately 15 minutes in total providing responses to the demographic items (1 minute), 17 items on the result's validity and accuracy (6 minutes), as well as 15 items on the system's usefulness and ease of use questionnaire (6 minutes).

We mainly adopt Focus Group Interview, which is a method by which researchers formulate specific topics and collect materials in group discussions. Using this method, we can carry out a "conversational" exploration of the research object, deeply understand the interview object, and obtain original materials in the actual investigation. Thus, the topic of the interview is whether the results of automatic evaluation based on the system contribute to the improvement of teaching quality.

Figure 3

Experimental process



In the interview process, the researcher will gradually draw out the teachers' opinions and suggestions on the application of automatic evaluation of educational products, including their own requirements. After sorting out the interview content, the advantages and disadvantages of applying AET-system in primary schools and the improvement direction of course evaluation in the future are preliminarily determined.

Results

Reliability analysis

To ensure that the study maintained good quality data, data with missing answers were discarded, leaving valid answers. 35 valid questionnaires were collected. The reliability analysis conducted on the questionnaire demonstrated good internal consistency, with a Cronbach's alpha coefficientof 0.963.

Table 2

The questionnaire Cronbach α coefficient

Subject		Cronbach α
	Classroom Environment	0.942
AET-system's validity and	Lesson Structure	0.914
accuracy of the results	Implementation	0.866
	Content	0.928
AET-system's the usefulness and ease of use 0.943		

Advantages and Disadvantages of AET-system

Most teachers agree that the application of automatic evaluation systems in the teaching and research activities of primary schools can effectively solve the limitations of time and space. It can support teachers to conduct timely evaluation in their spare time after class. At the same time, the method of automatic record analysis provides a certain basis for evaluation, and it is highly efficient so that the classroom can be analyzed quickly and the teaching strategy can be adjusted in a timely manner.

Interview data：

Teacher A: "Automated evaluation can greatly reduce the time we spend writing notes and give more time on advising teachers."

Teacher B: "The quantitative method provides evidence to support our suggestions, making them more accurate and appropriate."

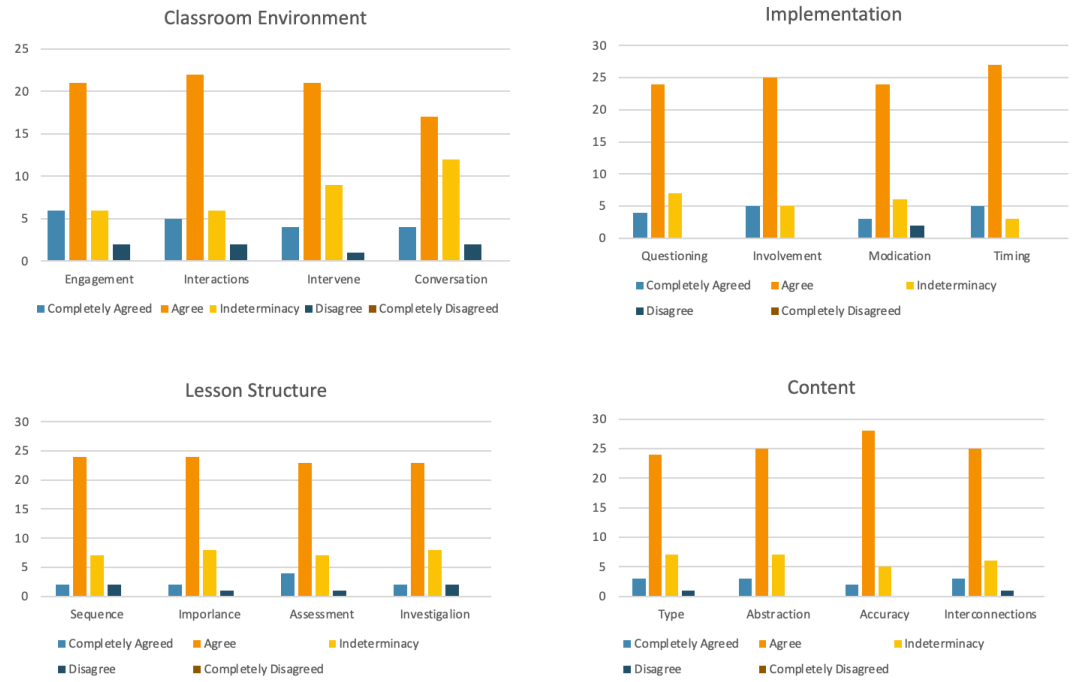
Teacher C: "The generation time of evaluation results is fast, and teaching can be improved according to the evaluation report in time."

Teacher D: "But the evaluation results were presented only in charts without corresponding improvement strategies."

Teacher E: "The evaluation of the class was fragmentary and incomplete, and the interpretation of visual charts was few, and no constructive suggestions were put forward. It may lead to formalization of teaching and deviation from art."

Figure 4

Data analysis result



The Direction of Improvement

In addition to the graphic presentation, the automatic evaluation report also needs to put forward a diagnostic evaluation. According to the improvement points of the teachers, the corresponding improvement suggestions can be given from multiple perspectives, so as to enable the teachers to make improvements based on their own teaching.

Interview data:

Teacher B: "The evaluation criteria should be divided according to subject knowledge field, teacher development stage, subject research situation, curriculum type, etc."

Meanwhile, the evaluation criteria should absorb excellent scales and combine the subject characteristics to develop the evaluation criteria of each subject, so that the evaluation results are more accurate and appropriate.

Conclusion

Nowadays, teaching evaluation is turning to "subjective experience" and "data." The AET-system not only provides data, but also supports the reasonable selection of expert comment. Meanwhile, it greatly decreases the cost of evaluating and changes the traditional pattern.

This study finds that the automatic evaluation system can be accepted and applied by primary school, and it has benefits in the evaluation subjects, manners and results, etc. However, it also has the following shortcomings: lack of in-depth exploration of the evaluation subject needs; evaluation results are simple, lack of diagnostic evaluation of teachers, difficult to support teachers personalized reflection.

Therefore, the future automated evaluation system will be human-AI coordination and mutual promotion, multi-subject division of labor cooperation, multi-subject evaluation mode, and integration of multi-subject knowledge and wisdom forming a perfect evaluation system to provide more diverse and comprehensive evaluation results.

The limitation of this study is that only 8 teachers from one school were selected to participate in the study, with a small sample size and no combination of quantitative research. In the future, the number of samples can be expanded, and qualitative and quantitative research methods can

be adopted, which is more conducive to the promotion of the conclusion.

References

Jie, F. (2014). An Overview and Reflection on Classroom Observation and Evaluation in the Past Twenty Years in China. *Journal of Northwest Normal University* (Social Sciences Edition) (03),104-108.

Putnam, R.T., & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have to Say About Research on Teacher Learning? *Educational Researcher, 29*, 15 - 4.

Zhai, X., Haudek, K. C. , Stuhlsatz, M. A. M. , & Wilson, C. . (2020). Evaluation of construct-irrelevant variance yielded by machine and human scoring of a science teacher pck constructed response assessment. *Studies In Educational Evaluation, 67*. 100916

James, A., Kashyap, M. , Chua, Y. H. V. , Maszczyk, T. , & Dauwels, J. . (2018). Inferring the Climate in Classrooms from Audio and Video Recordings: A Machine Learning Approach. *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. IEEE.

Richardson-Koehler, V. (1994). *Teacher change and the staff development process: a case in reading instruction*. Teachers College Press.

Li, L. (2015). The Research on the Techniques for Classroom Observation and Evaluation. *China Educational Technology* (06), 74-79.

Liang, J. (2015). Live video classroom observation: An effective approach to reducing reactivity in collecting observational information for teacher professional development. *Journal of Education for Teaching, 41*(3), 235–253. https://doi.org/10.1080/02607476.2015.1045314

Acheson, K. A., & Gall, M. D. . (1987). *Techniques in the clinical supervision of teachers: preservice and in-service applications* (2nd ed). Longman.

Cohen, E. G. (1971). *Analyzing teaching behavior.* American Educational Research Association. https://doi.org/10.2307/1161944

Shan, Y. - J. (2008). Using the S-T Analytical Method to Analysis the Classroom Instruction in Educational Technology Undergraduate Courses. *Modern Educational Technology* (10), 29-31.

Gu, X. -Q, Wang, W. (2004). New exploration of classroom analysis techniques to support teacher professional development. *China Educational Technology* (07),18-21.

Prieto, L.P., Sharma, K., Kidzinski, L., Rodríguez-Triana, M.J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. J*ournal of computer-assisted learning, 34*(2), 193-203 .

Walkington, C.A., & Marder, M.P. (2018). Using the UTeach Observation Protocol (UTOP) to understand the quality of mathematics instruction. *ZDM,* 50, 507-519.

Ling, C., Dan, Y. , & Xinran, H. . (2018). Does mobile classroom observation improve equal communication and deep collaboration among teachers ——an empirical study based on mobile classroom observation tool. *China Educational Technology* (00), 107-114.

Lei L.. (2016). Design and Research of Technology System of Evaluating the Effect of Real-time Network “Observing and Discussing Lesson”—— Aims at the Professional Development of Visiting Teachers. *China Educational Technology* (12), 146-149.

Wen-Yan, L., Tao, L. I. , Education, F. O. , & University, B. N. . (2018). Evaluating teaching quality based on classroom observation: Framework, practice, and implication. *Teacher Education Research* (01), 64-71.

Wei-Fu W., Mei-Juan, M., & Han-Bing, Y. (2022). Study on an analysis model for teacher evaluation feedback from the perspective of precision teaching and research. *e-Education Research* (01), 122-128.

Xiao, M.-M. (2019). From the perspective of three orientations of teacher professional development. *Journal of Curriculum and Instruction, 02*, 16-20.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (NSFC) under the project "Synergistically Automated Recognition and Interpretation of Student Classroom Attention Based on Graph Neural Networks" (Project No. 62177039), the Research Project of Continuing Education in Southwest University under the grant number (Project No. SWU2108023), the Innovation Research 2035 Pilot Plan of Southwest University under the grant number (Project No. SWUPilotPlan002) and The key project of the National Education Science Planning and the Ministry of Education, "Strategic Research on Human-Computer Collaborative Evaluation of Lessons to Empower Primary School Mathematics Teachers' Professional Development" (Project No. DHA230394).