

Research on Construction of an Effect Evaluation System of Online English Teaching in Colleges and Universities Based on Kirkpatrick Model

Huazhi Xiao¹ Yuanyou Jiang² Pinglan Yu³

^{1,2,3} College of Public Basic Education, Guangzhou Civil Aviation College, Guangzhou, Guangdong, China

¹Corresponding author. Email: xiaohuazhi@gcac.edu.cn

ABSTRACT

Based on the analysis of relevant literature, an evaluation index system for the effectiveness of online English teaching in colleges and universities is constructed from the four levels of the Kirkpatrick model: reaction, learning, behavior, and results. A questionnaire on the index weight is designed for this system, and SPSSAU online software is used to perform AHP hierarchical analysis on the questionnaire data using the sum product method to calculate the weight of each index, thus constructing the weights of evaluation indexes for the effectiveness of online English teaching in colleges and universities. The construction of this evaluation index system based on the Kirkpatrick model can provide reference for scientifically evaluating the effectiveness of online teaching.

Keywords: Kirkpatrick model, College English, Online teaching, Evaluation system, Index weight.

1. INTRODUCTION

With the development of internet technology and the further implementation of online teaching in colleges and universities, online teaching poses higher requirements for teachers' teaching abilities, students' self-learning abilities, the operation capabilities of online teaching platforms, and hardware conditions. What the effect of online teaching is and how to evaluate the effectiveness have always been the issues explored and focused on in online teaching by colleges and universities. Therefore, establishing an effective evaluation system for the effectiveness of online teaching has become a focus of higher education.

Based on the Kirkpatrick model, in this study an evaluation system for the effectiveness of online English teaching in colleges and universities is established, in order to objectively and scientifically evaluate the effectiveness of online English teaching from multiple levels, promote teaching and learning through evaluation, provide references for monitoring the quality of online English teaching in colleges and universities and improving the effectiveness of online English

teaching, and also provide references for the evaluation of online teaching in other disciplines.

2. LITERATURE REVIEW

In higher education, the effectiveness of online English teaching has received increasing attention. In order to address this issue, many scholars have begun to construct an effectiveness evaluation system based on the Kirkpatrick model.

2.1 Research Status of Online English Teaching Evaluation in Colleges and Universities in China and Foreign Countries

In China, there are not many studies on the evaluation model and index system of online English teaching in universities. In CNKI's academic journal database, a total of 16 relevant articles were retrieved. Zhang[1] studied the quality evaluation index of foreign language online courses from the perspective of learners. Li[2] constructed a formative assessment model for college English writing based on the automatic evaluation system of compositions. Wang[3] proposed the principles and

strategies for building an online teaching evaluation index system from a macro perspective. Xu and Wang[4] used AHP to construct a teaching evaluation system from three different perspectives: MOOC platform, learners, and teachers. Gao[5] studied the evaluation system of college English teaching based on SPOC. Ju[6] proposed a multiple evaluation system for college English under the background of "Internet Plus" from aspects such as evaluation subjects, evaluation methods, and evaluation means. Yin[7] conducted a study on the evaluation mechanism of foreign language teaching effectiveness based on online teaching platforms.

On the core collections of Web of Science, the author conducted a search on research regarding the evaluation system and model for English online teaching effectiveness in Chinese universities, resulting in 23 valid articles authored by researchers from China. For example, Wang[8] proposed a method to evaluate user perception of English distance education based on convolutional neural networks. Han[9] integrated remote monitoring with deep learning algorithms to construct a systematic structure for evaluating English online teaching, evaluating students' behavioral patterns and states during the learning process and teachers' teaching process.

Currently, there is no consensus in the academic community regarding the evaluation model, evaluation indicators, and evaluation system for English online teaching effectiveness in universities. There is a need for further clarification and enhancements in terms of both breadth and depth.

2.2 Research Status of Online Teaching Evaluation Based on Kirkpatrick Model at Home and Abroad

The author conducted a high-level search in the CNKI database for research on teaching effectiveness evaluation based on Kirkpatrick model with topic "Kirkpatrick model * teaching effectiveness * (evaluation + assessment)". The search was not limited by year and the matching type was set as "accurate". A total of 17 relevant articles were retrieved, and most of the studies only discussed the evaluation model or system of teaching effectiveness based on Kirkpatrick model from a macro perspective. For example, Yang[10] proposed a preliminary evaluation system for the ideological and political teaching quality of higher vocational courses based on the four levels of Kirkpatrick model. Xie[11] designed a teaching effectiveness evaluation system based on the

framework of "three categories-four levels-five dimensions" using Kirkpatrick model, and evaluated the blended teaching effectiveness of the "Video Post-production" course using diversified evaluation. Cai[12] constructed an evaluation index system for undergraduate higher education quality based on Kirkpatrick model, and used Delphi method and analytic hierarchy process to scientifically assign weights to each level of indexes. However, among the retrieved literature, there is only one study on the evaluation of online teaching effectiveness, which is that Su, Zhao and Hu[13] proposed a teaching effectiveness index and its weight based on Kirkpatrick model, including reaction, learning, behavior, and result layers, to evaluate the online teaching effectiveness in universities. However, their research did not specifically focus on any particular discipline.

Research on evaluation based on Kirkpatrick model in foreign literature has broadened from initially evaluating enterprise training effectiveness to fields such as education and medicine. Among numerous literature, there are still not many studies on evaluating the online teaching effectiveness of university disciplines. Chrysafiadiadi[14] developed a new evaluation method called PeRSIVA based on Kirkpatrick model and hierarchical evaluation mode, which evaluated students' online learning from satisfaction, performance, progress, behavior, and status aspects. Liang[15] compared Kirkpatrick model with the Analytic Hierarchy Process (AHP) to evaluate the teaching effectiveness of blended teaching mode. Chang[16] used Kirkpatrick model to evaluate the learning effectiveness of the online general education information literacy material "Library and Information Usage" launched by the Ministry of Education (MoE) in Taiwan.

Throughout the current research situation at home and abroad, there are currently few studies on the evaluation of online English teaching effectiveness, as well as the application of Kirkpatrick model to teaching evaluation. Research on the evaluation of online English teaching effectiveness based on Kirkpatrick model is even rarer.

3. DETERMINATION OF THE EVALUATION INDEX SYSTEM FOR THE ONLINE TEACHING EFFECTIVENESS OF FOREIGN LANGUAGES IN COLLEGES AND UNIVERSITIES BASED ON KIRKPATRICK MODEL

In this article, an evaluation system for online English teaching effectiveness in colleges and universities is constructed based on the four-level evaluation model of Donald L. Kirkpatrick (hereinafter referred to as the “Kirkpatrick Model”). The Kirkpatrick model, proposed by the internationally renowned scholar Donald L. Kirkpatrick from Wisconsin University in 1959, is the most widely used training evaluation tool in the world. It mainly includes four levels: reaction evaluation, learning evaluation, behavior evaluation and results evaluation. Reaction evaluation is to evaluate the satisfaction of trainees; learning evaluation is to measure the learning achievement of trainees; behavior evaluation is to examine the extent of knowledge application and changes in behavior of trainees; results evaluation is to evaluate the benefits brought to the organization by training.

Based on the research on relevant literature on online teaching evaluation systems at home and abroad, combined with the content of the English Curriculum Standards for Higher Vocational Education (2021 Edition), and on the basis of the Cognitive Model, some representative indexes that meet the objectives and principles of constructing an evaluation system for online English teaching effectiveness in colleges and universities are pre-selected from the four aspects of reaction, learning, behavior, and results, forming a pre-selected index set. Then the experts anonymously expressed their opinions on the pre-selected indexes, and after several rounds of modifications, the experts' opinions were basically agreed upon, thus determining the evaluation index system for online English teaching effectiveness in colleges and universities. Under the four levels, six primary indexes have been established, namely, “satisfaction with online teaching platforms”, “satisfaction with online teaching quality”, “knowledge and skills”, “cognitive effects”, “learning behavior improvement”, and “learning results”. Under the primary indexes, 34 secondary indexes have been established.

4. ESTABLISHING THE WEIGHT OF INDEXES FOR THE EVALUATION SYSTEM OF ONLINE ENGLISH TEACHING EFFECTIVENESS IN COLLEGES AND UNIVERSITIES BASED ON KIRKPATRICK MODEL

In order to comprehensively understand and evaluate the overall effectiveness of online English teaching, this study conducted a questionnaire survey on indicator weights, comprehensively considering each level from response, learning, behavior to outcome, and calculating the weights of each indicator in the evaluation system to ensure the comprehensiveness and accuracy of the evaluation.

4.1 Designing the Questionnaire for Index Weight

Based on the six primary indexes established in the previous section and the 34 secondary indexes, a questionnaire with 34 questions has been designed to determine the importance of each index. The questionnaire adopts the five-level Likert scale, with five levels ranging from “very unimportant” to “very important”, each level being assigned a score of “1”, “2”, “3”, “4”, and “5”. The survey is conducted anonymously, and respondents are asked to make a true choice based on their actual situation.

4.2 Survey Participants and Tools

The participants in this questionnaire survey were English teachers from several colleges and universities from Guangzhou, Shaoguan, Guizhou, Zunyi with 52 valid questionnaires received. The survey data were statistically analyzed with SPSSAU online software.

4.3 Test of Questionnaire Reliability and Validity

To ensure the reliability and validity of the questionnaire, the online tool SPSSAU was used to verify the reliability and validity of the survey results. The reliability test is generally based on Cronbach α , when Cronbach α is lower than 0.6, it indicates that the reliability is low and unacceptable. At the same time, the corrected item total correlation (CITC) analysis method was used to eliminate the items that did not meet the requirements. When the α coefficient of the deleted

item is higher than Cronbach α , the item should be deleted. The Cronbach α value of the index weight questionnaire is 0.983, which is much higher than 0.6, indicating that the reliability of the research data is very high. Moreover, the CITC values of the analysis items are all greater than 0.4, indicating that there is a good correlation between the analysis items. Overall, the data reliability is high and can be used for further analysis.

The validity of the survey data was verified using online software SPSSAU for KMO and Bartlett tests. The KMO value of the questionnaire was 0.808, which was greater than 0.7, and the P value of the Bartlett test was 0.000, which was much smaller than 0.05, indicating that the selected variables were suitable for factor analysis. The detailed data is shown in "Table 1".

Table 1. KMO & Bartlett's Test

KMO	0.808
Bartlett Test of Sphericity	approximate Chi-Square 2680.175
df	561
p	0.000

4.4 Confirmatory Factor Analysis

As the index system of the effectiveness evaluation of online English teaching in colleges and universities based on Kirkpatrick model constructed in this article is not mature, factor analysis needs to be conducted. In this study, a confirmatory factor analysis was conducted with SPSSAU for 6 factors and 34 analysis items. As shown in "Table 2", all the AVE values corresponding to the 6 factors are larger than 0.5, and all the CRs are higher than 0.7, indicating that the data in this study have a good convergent validity.

Table 2. AVE Model and CR Index Results

Factors	Average Extraction (AVE)	Variance Composite Reliability (CR)
Satisfaction Towards Online Teaching Platform	0.761	0.941
Satisfaction Towards Online Teaching Quality	0.584	0.907
Knowledge & Skills	0.831	0.936
Cognitive Effects	0.797	0.940
Improvement of Learning Behavior	0.778	0.961
Output of Learning Results	0.800	0.970

4.5 Calculation of Evaluation Index Weight

In this study, with SPSSAU software, the weight was calculated by using the method of analytic hierarchy process (AHP) with the sum method for the analysis results shown in "Table 3". In the AHP level analysis to calculate the weight, it is necessary to conduct a conformance test to analyze the consistency of the weight calculation results. The *composite reliability* (CR) ($CR=CI/RI$) was calculated for the 34-level judgment matrix. The composite index (CI) calculated for this study was 0.000, and the random index (RI) was 1.683 according to the table. Therefore, the CR was calculated to be $0.000 < 0.1$, indicating that the judgment matrix met the conformance test and the calculated weight had consistency.

Table 3. Results of conformance test

Maximal Characteristic Root	CI	RI	CR	Results of Conformance Test
34.000	0.000	1.683	0.000	Pass

4.6 Determination of Evaluation System

Through analytic hierarchy process (AHP) by sum-product method, the weights of the four levels, primary indexes, and secondary indexes were calculated statistically to form a specific evaluation index system for online English teaching effectiveness in colleges and universities based on Kirkpatrick's model (see "Table 4"). The weights of the four levels are as follows: Reaction (35.91%), Learning (20.54%), Behavior (20.29%), and Results (23.26%). The weights of the primary indexes are as follows: Satisfaction Towards Online Teaching Platform (15.22%), Satisfaction Towards Online Teaching Quality (20.68%), Knowledge and Skills (8.94%), Cognitive Effect (11.60%), Improvement of Learning Behaviors (20.29%), and Output of Learning Results (23.26%). The weights of the secondary indexes are as follows: Usability of Online Teaching Platform (3.06%), Stability of Online Teaching Platform (3.10%), Security of Online Teaching Platform (2.99%), Interactivity of Online Teaching Platform (3.06%), Online Teaching Resource Platform (3.02%), Teaching Models (2.81%), Teaching Methods (2.93%), Teaching Implementation (3.01%), Guidance Feedback (2.85%), Textbooks (3.02%), Teaching contents (3.02%), Student Participation (3.05%), Basic Theoretical Knowledge (3.02%), Basic

English Skills (2.99%), Cross-cultural Communication Knowledge (2.93%), English Learning Attitude (2.97%), Cooperative Learning Awareness (2.93%), English Learning Confidence (2.86%), English Learning Commitment (2.84%), Language Communication (2.86%), Cooperative Learning (2.93%), Information Literacy (2.94%), Cross-cultural Communication (2.85%), Language Thinking (2.83%), Self-regulated

Learning (3.01%), Learning Evaluation (2.88%), English Skills (2.95%), Language Communication Ability (2.98%), Cooperative Learning Ability (2.93%), Information Technology Literacy (2.88%), Cross-cultural Communication Ability (2.83%), Language Thinking Ability (2.85%), Course Certification (2.92%), Self-regulated Learning ability (2.93%).

Table 4. Evaluation indexes and their weight values for online English teaching effectiveness in colleges and universities based on Kirkpatrick Model

Levels	weight value	Primary indexes	weight value	Secondary indexes	Description of indexes	weight value				
Reaction	35.91%			Usability of Online Teaching Platform	Whether the online teaching platform is easy to operate, fully functional, applied in multiple terminals, and can meet the needs of learners at different levels	3.058%				
				Satisfaction Towards Online Teaching Platform	15.22%	Stability of Online Teaching Platform	Whether the online teaching platform operates stably and smoothly	3.096%		
						Security of Online Teaching Platform	Whether the online teaching data information is safe and reliable, and updated in a timely manner	2.993%		
						Interactivity of Online Teaching Platform	Whether the links of online teaching platform are clear and easy to switch	3.058%		
						Online Teaching Resource Platform	Whether the online teaching resources are rich and presented in various ways	3.019%		
						Teaching Models	Whether teachers adopt diversified online teaching models	2.813%		
						Teaching Methods	During the online teaching process, whether the teacher applies a variety of effective teaching methods	2.929%		
						Satisfaction Towards Online Teaching Quality	20.68%	Teaching Implementation	Whether the teacher organizes classroom teaching in an orderly manner, interacts fully with students, and objectively evaluates students	3.006%
								Instructor Feedback	Whether the teacher answers questions in a timely manner, supervises and reminds students to study	2.851%
								Textbooks	Whether the difficulty level of textbooks is suitable for students' current level	3.019%
Learning	20.54%	Knowledge & Skills	8.94%	Teaching Contents	Whether the teaching contents cover all learning objectives and whether the key points and difficulties are highlighted	3.019%				
				Student Participation	Whether students actively participate in classroom activities, ask and answer questions	3.045%				
				Basic Theoretical Knowledge	Whether students basically grasp the English knowledge taught in classroom, including grammar, vocabulary, etc.	3.019%				
				Basic English Skills	Whether students basically grasp such skills as listening, speaking, reading, writing, translation, etc. taught in the classroom	2.993%				
				Cross-cultural Communication Knowledge	Whether students grasp necessary cross-cultural knowledge	2.929%				

Levels	weight value	Primary indexes	weight value	Secondary indexes	Description of indexes	weight value
Learning	20.54%	Cognitive Effects	11.60%	English Attitude	Learning Whether students are willing to actively participate in classroom activities and complete coursework on time	2.967%
				Cooperative Learning Awareness	Whether students are willing to participate in group cooperative learning	2.929%
				English Confidence	Learning Whether online teaching can enhance students' confidence in learning English	2.864%
				English Commitment	Learning Whether students are willing to spend more time and energy on online English learning	2.838%
Behavior	20.29%	Improvement of Learning Behavior	20.29%	Language Communication	Whether students use English for communication and exchanges in study and daily life	2.864%
				Cooperative Learning	Whether students actively participate in group cooperative learning during the learning process	2.929%
				Information Literacy	Whether students use the internet to search and organize English learning resources	2.942%
				Cross-cultural Communication	In cross-cultural communication, whether students hold an equal, inclusive and open attitude	2.851%
				Language Thinking	Whether students can distinguish the similarities and differences between Chinese and English thinking modes	2.825%
				Autonomous Learning	Whether students can effectively manage time and independently carry out learning activities without supervision and guidance of teachers	3.006%
				Learning Evaluation	Whether students often conduct self-testing of their learning effectiveness	2.877%
Results	23.26%	Output of Learning Results	23.26%	English Skills	Whether students' English skills meet post requirements	2.954%
				Language Communication Ability	Whether students can clearly express their own views to others in English	2.980%
				Collaborative Learning Ability	Whether students' willingness and ability to collaborate in learning have been improved	2.929%
				Information Technology Literacy	Whether the students' abilities to collect and organize learning resources using information technology have been improved	2.877%
				Cross-cultural Communication Ability	Whether students possess cross-cultural communication skills in communication and collaboration	2.825%
				Language Thinking Ability	Whether students' logical, critical and innovative thinking have been improved	2.851%
				Q33 Certification	Whether the pass rate of students' English proficiency tests meets expected targets	2.916%
				Autonomous Learning Ability	Whether students develop good habits of autonomous learning, acquire learning resources through multiple channels, and independently and effectively carry out learning	2.929%

5. CONCLUSION

Despite its immature nature and areas for improvement, the evaluation system for online English teaching effectiveness in colleges and universities based on Kirkpatrick model has opened up a new direction for evaluating the effectiveness of online teaching of foreign languages and expanded its research field, and also provided it with a quantifiable and operational evaluation system.

ACKNOWLEDGEMENTS

This research is supported by grants from “2022 Educational Science Planning Project in Guangdong Province (Special Project for Higher Education) ” (Grant No. 2022GXJK439: Research on the Evaluation of the Teaching Effect of Ideological and Political Education in College English Courses Based on Kirkpatrick Model), “The Fourteenth Five-Year Plan 2021 Higher Education Research Project of Higher Education Association in Guangdong Province” (Grant No. 21GYB133: Research on the Evaluation of Online Foreign Language Teaching Effectiveness Based on Kirkpatrick Model) , and “2022 Special Project of English Curriculum Standards and Teaching Standards of Foreign Language Majors for Vocational Education” (Grant No. WYJZW-2022-19-0217: Research on the Evaluation of English Teaching Effectiveness in Higher Vocational Education from the Perspective of the New Curriculum Standard).

REFERENCES

- [1] Zhang, W., Dong, Y., Sun, L. and Cao H. “Learner-based evaluation indexes of foreign language online course,” *Foreign Languages and Their Teaching*, 2022. (04): 111-121+149.
- [2] Li, S. “Applied research on formative assessment model of college English writing” *Overseas English*, 2022. 471(11):87-88.
- [3] Wang, W. “Research on the construction of an evaluation index system for the quality of online classroom teaching in college English,” *Journal of Jiangxi Vocational and Technical College of Electricity*, 2021. 34(05): 77-78.
- [4] Xu, D. and Wang, Y. “Evaluation and improvement of college English teaching based on MOOC,” *Vocational and Technical Education*, 2020. 41(17): 67-71.
- [5] Gao, X., Yan, L. and Kang, N. “Analysis of online college teaching quality evaluation indexes based on SPOC,” *Journal of Jilin TV & Radio University*, 2020. 221(05): 22-23.
- [6] Ju, C. “Construction of college English evaluation system under the background of ‘internet +’,” *Journal of Hebei Open University*, 2022. 27 (02): 62-65.
- [7] Yin, A. “Research on the evaluation mechanism of foreign language teaching effect based on online teaching platform,” *Crazy English (Theoretical Edition)*, 2018. 1073 (02): 52-53+58+207-208.
- [8] Wang, J. and Fu, J. “Convolutional neural network based network distance English teaching effect evaluation method,” *Mathematical Problems in Engineering*, 2022. 1-10.
- [9] Han, Y. “Evaluation of English online teaching based on remote supervision algorithms and deep learning,” *Journal of Intelligent & Fuzzy Systems*, 2021. 40 (4): 7097-7108.
- [10] Yang, H. “Exploring the teaching quality evaluation model of ideological and political education in higher vocational colleges based on Kirkpatrick model,” *Journal of Kaifeng Vocational College of Culture*, 2022. 42(12): 75-77.
- [11] Xie, S. “The effect evaluation of blended teaching based on Kirkpatrick model---A case study ‘Video Post-Production’ course in higher vocational colleges,” *Vocational Technology*, 2023. 22(05): 1-7.
- [12] Cai, Y., Cao, H. and Feng, Y. “Result-orientation research on evaluation system of higher education quality in China,” *China University Science & Technology*, 2022. (05): 10-15.
- [13] Su, B., Zhao, S. and Hu, Q. “Research on evaluation of online teaching effect in colleges and universities based on Korotkoff model,” *Journal of Shangrao Normal University*, 2021. 41(03): 98-105.
- [14] Chrysafiadi, K. and Virvou M. “PeRSIVA: an empirical evaluation method of a student model of an intelligent e-learning environment for computer programming,” *Computers & Education*, 2013. (68): 322-333.

- [15] Liang, F., Shen, W., Shi, H., et al. "Evaluation method of mixed teaching efficiency of college teachers based on Kirkpatrick model," *Computational Intelligence and Neuroscience*, 2022. 8609387. <https://doi.org/10.1155/2022/8609387>.
- [16] Chang, N. and Chen, L. "Evaluating the learning effectiveness of an online information literacy class based on Kirkpatrick framework," *LIBRI-International Journal of Libraries and Information Studies*, 2014. 64(3): 211-223.