

A kirkpatrick model-based learning evaluation system for blended teaching in military academies—Taking blended English course as an example

Meimei Shi^{1,*}, and Yong Li²

¹Fundamental Department, Engineering University of PAP, Xi'an, China

²Key Laboratory of Counter-Terrorism Command & Information Engineering (Engineering University of PAP), Ministry of Education, Engineering University of PAP, Xi'an, China

Abstract. At present, blended teaching has become an important direction of teaching reform in military academies. Due to the special attributes of military academies as well as its late start of blended teaching, there is currently limited research on blended learning evaluation and a lack of a unified evaluation system. In the light of Kirkpatrick Model, one of the most authoritative evaluation models for training, this paper first reflects on the flaws of current blended learning evaluation in military academies. Then based on the blended English course practice, explores to construct an evaluation system in the four levels of Kirkpatrick Model: level of reaction, level of learning, level of behavior and level of result. In the end, Delphi Method is used for index weight distribution for the whole system. The system could serve as a template and reference for establishing a practicable evaluation system, enhancing blended teaching reform in military academies.

1 Introduction

For the reason of confidentiality, although blended teaching has been an emerging form of education and there are a lot of researches around the construction of online+offline courses, it has not started in military academies until the Covid-19. In the following four years, blended mode of teaching are being adopted in College English, College Math and some other non-confidential public courses, but how to evaluate student learning has become a challenge. Blended teaching mode usually refers to the combination of traditional teaching mode and information technology[1]. Changes in learning environment, contents and methods require more advanced assessment methods and more advanced techniques of obtaining the relevant data. Due to the special attributes of military academies as well as its late start of blended teaching, there is currently limited research on blended learning evaluation and a lack of a unified evaluation system. Thus, how to construct the appropriate

* Corresponding author: 19591579361@163.com

evaluation system becomes one of the most important problems demanding prompt solution in order that blended learning can be carried out effectively in military academies.

2 Kirkpatrick model and its implications

2.1 Kirkpatrick model

Kirkpatrick Model was proposed by American scholar Donald Kirkpatrick in 1959. He identified four levels for evaluating training or summarized as four “Rs”: (1) reaction, or how well learners were satisfied with the program;(2)learning, or the principles, facts, and techniques that were learned; (3) behavior, or the changes in job behavior that resulted from the training; and (4) results, or the benefits to organization as a result of the training. As a systematic and efficient means to account for outcomes among employees and in organizational systems, Kirkpatrick Model has been the dominant method of evaluating training programs for more than 30 years[2]. The application of the model has reached the different higher education fields and aspects as well (Sahin, 2006 on teacher training program[3]; Chang & Chen, 2014 on online information literacy course[4]; Embi et al., 2017 on blended learning environment[5];Liao & Hsu, 2019 on medical education program[6]), but no research has been made on the adoption of Kirkpatrick Model for blended learning evaluation in military environment, which leaves much space for exploration.

2.2 Implications to current evaluations

Assessment can be generally classified into formative assessment and summative assessment. At present, the teachers in military academies mainly adopt 2 ways for blended course evaluation: (1) teacher-dominated summative assessment. A final exam or standardized test is organized at the end of a learning unit or course, which includes part of knowledge from the online content, usually 5%-10% of the overall score. In College English, it is mainly in the form of vocabulary multiple choices, C-E or E-C translation or alike. (2) teacher-dominated summative(60%-70%)+formative(40%-30%) assessment. In this approach, the teacher gives each student a rough score on their online performance during the learning process based on online quizzes, participation in online discussions, pre-class assignments, and other similar activities, which is usually 20%-30% of the formative score.

Under the light of Kirkpatrick model, it is obvious to conclude at least 3 flaws of the current evaluations in military academies: (1) the subjects of valuation should be diversified. As can be seen from above, in current evaluations, the teacher is playing a dominant role as the only judge. However,blended learning is a complex process with student role increasingly highlighted in different links, thus the subject of evaluation should be enlarged to include platform data, teachers, learners, and peers to enhance student initiative, enthusiasm, and creativity. (2) the evaluation content should be more comprehensive. The blended learning consists of face-to-face learning activities, self-paced network learning, collaboration and assessment, learning supporting materials of transfer and so on. Due to the lack of relevant online evaluation experience and accumulation, teachers often simply rated based on explicit records, such as visit frequency, video viewing time, online quiz as online learning scores, without considering the implicit factors which are also important objectives of learning, such as critical thinking, group cooperation, and creativity capability. Also, they have also not measured the effective connection between online and offline teaching data from an overall perspective. Therefore, a comprehensive and systematic

evaluation system needs to be constructed.(3) the evaluation methods should be enriched. Since blended learning makes use of various learning theories, technologies and patterns, diverse assessment methods should also be used in blended learning assessment. For example, platform data statistics, online class discussions and quiz, group work and mutual appraisal and etc. can be adopted to enhance both teaching and student learning, and promote personalized development of students.

3 Design of the four-level evaluation system

Assessment in blended learning environments can present unique challenges compared to traditional face-to-face instruction. Based on the blended English teaching practice, this paper proposes a four-level (reaction level, learning level, behavior level, and result level) six-dimension (evaluation level, evaluation objective, evaluation index, evaluation content, evaluation method, and evaluation subject) evaluation framework as below(Table 1).

Table 1. 4-level evaluation system for blended English learning.

Evaluation Level	Evaluation Objective	Evaluation Index	Evaluation Content	Evaluation Method	Evaluation Subject
Reaction level	Student subjective feelings about blended learning	Online learning attitude	1.Platform visit frequency 2.Platform learning duration 3.Online assignment submission 4.Idea posting	Platform data	Platform
		Offline learning attitude	1.Class attendance 2.Class mentality 3.Question answering	Behavior observation & record	Teacher
Learning level	Knowledge and skill acquisition through blended learning	Knowledge mastery ability	1.Pre-reading 2.Question answering 3.Platform unit test	Behavior observation & record +Platform data	Teacher +Platform
		Independent learning ability	1.Resource loading 2.Personal assignment	Platform data	Platform
		Collaborative learning ability"	1.exchange of ideas and mutual appraisal 2.Division of tasks within the group 3.Group work assignment	Questionnaire	Leaner +Peer+ Teacher
Behavior g level	Attitude and behavior change after blended learning	Communication and expression ability	1.Reporting demonstration 2.Group discussion	Questionnaire	Leaner +Peer+ Teacher
		Change in learning attitude	1.Extension beyond the classroom 2.Problem solving	Questionnaire	Learner
Result level	Learning outcomes and subsequent development	Change in creative thinking	1.Reflection and critical thinking 2. Innovation in assignment	Questionnaire+ Assignment analysis	Leaner +Teacher+ Peer
		Learning outcome	1.End-of-term test 2.CET 4/CET 6	Standard test	Teacher
		Subsequent developments	1.Competition participation 2.Awards	Questionnaire+ E-portfolio	Leaner +Teacher

3.1 Evaluation design of the reaction level

The reaction level mainly evaluates students' reactions, feelings and attitudes towards blended learning. Positive and enthusiastic attitude will be conducive for active participation and course devotion. As shown in table 1, the indexes are divided into two dimensions: online and offline with each dimension assigned multiple observation points. The online data can be obtained through the statistics of the learning platform while the offline data relies on teacher's observation and record of the students' classroom performance. The data lasts throughout the teaching process and is a type of formative assessment. As blended learning in military academies needs to be carried out in unique environment, with strict rules on organization and internet access, the data in this level could also serve as important references for improving the overall cooperation and coordination of different departments involved.

3.2 Evaluation design of the learning level

The learning level measures the level of knowledge, skills, and values acquired by the students from blended environment. This level is the most important part of the whole evaluation system, which can most intuitively and effectively reflect the blended learning

implementation details. Based on the cognitive, emotional and skill objectives of College English in military academies, the evaluation indexes are divided into four dimensions, namely, “knowledge mastery ability”, “independent learning ability”, “collaborative learning ability” and “communication and expression ability”. “Knowledge mastery ability” involves the teachers’ observation record and the data statistics of the platform. For “independent learning ability”, “collaborative learning ability” and “communication and expression ability”, learners, peers and teachers are invited to complete the evaluation. The self-evaluation is conducive to the formation of a sense of achievement, the determination of goals and the cultivation of personalization; peer assessment is actually a process of communication, learning and collision of sparks of thought among all students, which helps for the mutual learning of the whole class and mutual improvement; at the same time, the teacher’s evaluation serves as the correct guidance of the direction of inquiry, helping and ensuring its relative fairness. A combination of teacher-valuation, peer-evaluation and self-evaluation makes the whole process more scientific and rational. The evaluation at this level can reflect students’ knowledge and skills acquisition in real time, and at the same time, it can enable learners to obtain feedback information instantly to improve their learning.

3.3 Evaluation design of the behavior level

The behavior level mainly evaluates the attitudes and behavioral changes that learners experience after participating in blended learning. It is divided into two dimensions: changes in learning attitude and changes in innovative thinking. For example, whether there is extracurricular extension of open and associative thinking, whether the ability to analyze and solve problems has been improved, and whether the innovation of translation and writing assignments has been improved. Psychological research has shown that changes in a person’s ideology and behavior require three stages: obedience, identification, and internalization. Repetition for more than three months can form relatively stable habits. Therefore, within three months after the end of the course, students can be observed, homework analyzed, and questionnaire evaluated, with a focus on comparing their behavior before and after the blended course. The data collection at this stage is relatively lengthy but significant. Only when students truly undergo behavioral changes in their learning attitude, thinking, and other aspects can the purpose of teaching reform be truly achieved, and necessary support be provided for the next round or different subject teaching practices.

3.4 Evaluation design of the result level

The result level evaluation focuses on whether the blended learning can bring concrete and direct performance results to individuals and organizations, and it is the highest level and most profound assessment that extends from the classroom to the outside of the classroom and from the individual to the organization. The evaluation indexes include two parts: students’ learning outcome (academic performance) and subsequent developments. The most commonly used test for “learning outcome” is the standard test, which can be online or offline, and the subject of evaluation corresponds to the data of the platform and the teacher, which is a kind of summative evaluation at the end of the course. The “follow-up developments” is to measure and evaluate the learning effect of students through their participation in competitions and winning awards for their works, requiring lengthy time and support from relative departments, which is also the reason why it is rarely involved in traditional evaluation. For the evaluation of these indexes, it is recommended the use of remote survey questionnaires, targeted question interviews, and the establishment of electronic archives.

3.5 Weight determining of indexes

Based on the four-level framework, this paper uses Delphi Method to determine the weight of the indexes, that is, the relationship between the index and the proportion in the results, and the flowchart is shown in Figure 1.

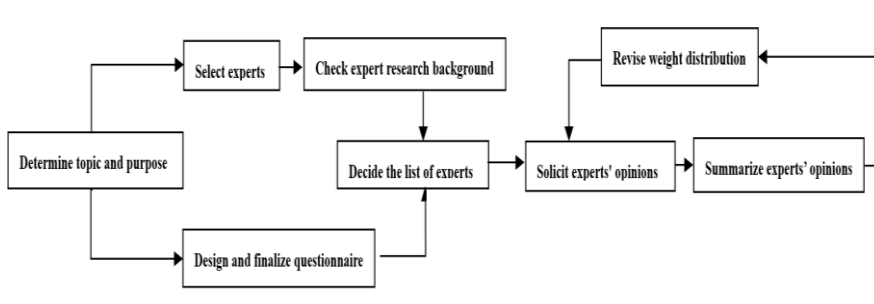


Fig. 1. Flowchart of index weight determining.

The key steps are as the followings:

Select experts. 10 experts from 5 military academies are selected based on research need. Selection conditions: one is being acting as a member of the Teaching Supervision Committee/professor or associate professor engaging in teaching at least 15-year ; the other is to work with a serious attitude and high enthusiasm.

Design questionnaire and solicit experts’ opinions. Design a weight determining questionnaire based on the evaluation levels and indexes in the constructed evaluation system and invite the 10 experts to rank the 4 evaluation levels and all indexes in terms of importance.

Weight the four evaluation levels. Firstly, determine the ranking weight value of each rank for the four levels from all experts. The quantitative calculation method is as follows: assuming that there are M indexes participating in the ranking, the value of the top ranked index is M points and the second top ranked index is M-1 points. By analogy, the last ranked index gets 1 point. Secondly, calculate the average weight value of each level. The calculation formula is: average weight value of each level= \sum ranking frequency \times ranking weight value/total number of ranking. Finally, calculate the percentage weight of each level. The calculation formula is: percentage weight=(average weight value of a certain level \times 100)/total score of the four levels. (See Table 2)

Table 2. Weight determining of the 4 evaluation levels.

	Ranking weight value 4	Ranking weight value 3	Ranking weight value 2	Ranking weight value 1	Average weight value	Percentage weight
Reaction level	0	1	5	4	1.7	17
Learning level	7	2	1	0	3.6	36
Behavior level	1	6	1	2	2.6	26
Result level	2	1	3	4	2.1	21
Total	10	10	10	10	10	100

Weight the indexes of the whole system. After determining the weight of the four evaluation levels, the weight of the evaluation indexes under each level can be determined through the same process, and finally a complete weight distribution to support blended English learning can be obtained(see Table 3). The whole process provides a weighting template for learning in blended environment, as for different courses, the weight could be adjusted and determined according to different needs and actual situations.

Table 3. Weight determining of the whole evaluation system.

Evaluation Level	Weight	Evaluation Index	Evaluation Content	Weight
Reaction level	0.17	Online learning attitude	1. Platform visit frequency 2. Platform learning duration 3. Online assignment submission 4. Idea posting	0.4
		Offline learning attitude	1. Class attendance 2. Class mentality 3. Question answering	0.6
Learning level	0.36	Knowledge mastery ability	1. Course preparation 2. Question answering 3. Platform unit test	0.35
		Independent learning ability	1. Resource loading 2. Personal assignment	0.2
		Collaborative learning ability	1. Exchange of ideas and mutual appraisal 2. Task division within the group 3. Group work assignment	0.2
		Communication and expression ability	1. Reporting demonstration 2. Group discussion	0.25
Behavior level	0.26	Change in learning attitude	1. Extension beyond the classroom 2. Problem solving	0.3
		Change in creative thinking	1. Reflection and critical thinking 2. Innovation in assignment	0.7
Result level	0.21	Learning outcome	3. End-of-term test 4. CET 4/CET 6	0.6
		Subsequent developments	3. Competition participation 4. Awards	0.4

4 Conclusion

Blended teaching is a complex project, and evaluation system is the foundation for this project. The core of evaluation of blended learning is the value judgments on learning process by synthetically employing different assessment methods, tools and patterns so as to promote and improve blended learning ultimately. This paper constructs the basic framework of evaluation design from the 4 levels of Kirkpatrick Model and provides the corresponding weighting templates as well, which can reflect the different aspects of blended learning in a diversified, multi-level and multi-dimensional way. It could serve as a reference and basis for establishing a practicable evaluation system, enhancing blended teaching reform in military academies.

References

1. Wang R. Research on effectiveness of college english blended teaching mode under small private online course based on machine learning[J]. *SN Applied Sciences*, 2023, 5(2):1-13. DOI:10.1007/s42452-023-05278-y.
2. Galloway D L. Evaluating distance delivery and e-learning is kirkpatrick’s model relevant?[J]. *Performance Improvement*, 2010, 44(4):21-27. DOI:10.1002/pfi.4140440407.
3. Sahin, V. (2006). Evaluation of the in-service teacher training program “The Certificate for Teachers of English” at the Middle East Technical University School of Foreign Languages. (Dissertation, Middle East Technical University, Ankara, Turkey) Retrieved from <https://etd.lib.metu.edu.tr/upload/12607752/index.pdf>
4. Chang, N. & Chen, L. (2014). Evaluating the learning effectiveness of an online information literacy class based on the Kirkpatrick framework. *Libri*, 64(3), 211-223. <https://doi.org/10.1515/libri-2014-0016>
5. Embi, Z.C., Neo, T.K., & Neo, M. (2017). Using Kirkpatrick’s evaluation model in a multimedia-based blended learning environment. *Journal of Multimedia Information System*, 4(3), 115-122, 2383-7632. <http://dx.doi.org/10.9717/JMIS.2017.4.3.115>
6. Liao, S.C., & Hsu, S.Y. (2019). Evaluating a continuing medical education program: New World Kirkpatrick Model Approach. *International Journal of Management, Economics and Social Sciences*, 8(4), 266-279. <http://dx.doi.org/10.32327/IJMESS/8.4.2019.1>