Predicaments and Countermeasures of Multimodal Application for English Teaching in Higher Vocational Colleges

XI Yayun[a],[b],*; WANG Qian[a]

[a] School of General Education, Hefei Technology College, China.
[b] English Language Studies, Lyceum of Philippines, Batangas, Philippines.
*Corresponding author.

Abstract
In the intelligent learning era, multimodal data is an important carrier to accurately depict learners learning situations. Studying the teaching of multimodal learning analytics is helpful to restore original teaching process, reveal teaching rules, and help learners grow. This paper reviews the current situation of multimodal learning analytics at home and abroad, and points out that multimodal application for English teaching in higher vocational colleges currently faces difficulties such as the weighting and proportion of multiple data sources, the protection of ethical privacy, and the acquisition of platform data. With the learners needs based on a questionnaire survey, the research believes that a multi-modal and data-driven English teaching system should be built, a reasonable learning ability evaluation system should be created, and the characteristics of the development of English teaching in higher vocational education should also be integrated, so as to promote the empowerment of education big data technology, and high-quality development of intelligent education project and higher vocational education.

1. INTRODUCTION
The National Vocational Education Reform Implementation Plan of China clearly states that vocational education should adapt to the development needs of Internet plus vocational education and use modern information technology to improve teaching method. From traditional teaching mode to modern information-based teaching and then to intelligent teaching, the concept of modern information technology has been constantly innovated. In such intelligent teaching environment, it has become an inevitable trend in the new era to use multimodal data to build a digital intelligence campus and promote the empowerment of education big data technology, which also raise higher requirements for foreign language teaching reform in higher vocational colleges. This article will analyze the predicament of multimodal application for General English teaching in higher vocational colleges and propose corresponding strategies.

2. MULTIMODAL LEARNING ANALYTICS
Since Tim Johns put forward the theory and method of data driven learning (DDL) in 1991 (Johns, 1991), great progress has been made in the research of applying multimodal data to foreign language learning. The concept of multimodal learning analytics (MMLA) was

first proposed by Stefan Scherer at the 14th International Conference on Multimodal Interaction, who believed that multimodal learning analytics was consisted of multimodal data, multimodal teaching and learning, and computer-supported analysis (Scherer, Worsley, & Morency, 2012). This concept promotes the development of learning analysis towards data, science and intelligence.

2.1 The Connotation of Multimodal Learning Analytics

Modality refers to "the meaning potential of material media shaped by society for a long time, which is a social and cultural resource used to represent and exchange meanings" (Kress & Van Leeuwen, 1996). The ways human beings communicate with the external environment through their senses, such as sound, image and text, are all called modes. In recent years, with the advent of the era of artificial intelligence, the trend of deep integration of intelligent technology and information is irresistible, and the concept of modality has also been extended. In statistics, when there are multiple modes in a data set, then the set is called multimodal. In particular, relevant scholars define "multimodality" as a combination of explicit data representation modes such as text, voice and video and implicit data representation modes such as physiological information.

2.2 Current Study of Multimodal Learning Analytics in China

Research on Foreign language learning based on multimodal learning analytics in China is still at the early stage. Many Chinese scholars proposed to apply multimodal data such as text, video and audio to the field of multimodal interaction research. For example, Zhang Zhenhong (2014) studied the application of the multimodal corpus of college English (Zhang, He, & Han, 2014), and Ma Yunfei (2020) suggested that multimodal data should be used for in-depth learning in vocational education (Ma, Yue, & Di, 2020). However, the potential of multimodal application for English teaching in vocational colleges with promotional value remains underutilized.

2.3 Types of Multimodal Learning Analytics

The current research on multimodal data-driven learning can be divided into four categories. One is the analysis of students' classroom performance based on computer vision technology. Through the intelligent collection of students' expression, action and other image information in smart classrooms, students' concentration and emotional state can be analyzed. Secondly, potential views and learners' cognitive development can be extracted according to semantic network and knowledge map. The last is to analyze learners' learning preferences and thus recommend learning resources with the help of learners' retrieval, browse, view and test data based on online learning flow data of learners on the intelligent teaching platform, which will be the focus of this article.

3. PREDICAMENTS OF MULTIMODAL ENGLISH TEACHING IN HIGHER VOCATIONAL COLLEGES

At present, China has built the world's largest vocational education system. In 2022, the "National Vocational Education Intelligent Education Platform" has been launched, providing more possibilities for multimodal application for higher vocational education. Additionally, the Implementation Opinions on Promoting the High-Quality Development of Modern Vocational Education issued in December, 2021 proposes to deepen education and teaching reform of such four aspects as teaching models, teaching methods, teaching content and teaching materials. Therefore, the application of MMLA to vocational education has great potential. However, due to the practical difficulties such as the imperfect theoretical framework, the implementation of MMLA in higher vocational foreign language teaching is still facing challenges.

3.1 Weighting and Interpretation of Multiple Data Sources

Among the four types of multimodal learning research data, the data of learners' online learning flow in the intelligent teaching platform is easier to collect. With the reduction of the cost of wearable devices, portable wearable devices have entered the field of education. The acquisition of psychological data such as Electroencephalogram (EEG), heart rate and skin electricity can capture learners' learning behavior and facilitate the portrayal of their learning image. The data of multiple platforms and multiple modes are also complementary. Some data show that the error rate of predicting the learning effect by simply using learning behavior data (such as keys, touch screens, mouse clicks, etc.) is as high as 39%, while the error rate can reduce to 15% after adding survey data (such as self-report), and the error rate can be as low as 6% after combining physiological data (such as EEG, heart rate, skin electricity, etc.) (Peng & Jiang, 2022). However, the data obtained from different platforms and channels need to be sorted out and further integrated. For example, for the same learner, multiple data may cause information inconsistency or even conflict, so the weighting and interpretation of multiple data should be considered. Therefore, when there is a conflict between the learning flow data on the intelligent learning platform and the data generated by wearable devices, multiple data should be integrated with appropriate integration standards. Teachers should also make comprehensive judgments based on their understanding of students and the higher vocational colleges-based reality, so that multiple data can make collaborative decisions.

3.2 The Dilemma of Privacy Invasion

Multimodal learning analytics is to analyze a large amount of multimodal data obtained by computer software, which

involves the problem between data and privacy protection. Multimodal data driven teaching is to accurately depict the image of learners with big data, so as to grasp the learning rules and progress of learners. However, whether the detective devices that may be used in multimodal data collection must inform students of their special purposes in advance, whether the collected image data such as students' expressions and actions in the intelligent education scene can be protected without invasion, and how can the collected physiological data such as EEG, heart rate and skin electricity not be disclosed to illegal businesses are all factors that need to be considered in multimodal learning analytics. In addition, whether the monitoring equipment, contact or non-contact, will cause pressure on the monitored learners and lead to information deviation, and whether the monitored learners can faithfully and normally display their learning behavior without the interference of the monitoring equipment are all the premise to ensure the authenticity and accuracy of the data.

3.3 Difficulties in Obtaining Platform Data

Multimodal data collection should be long-term, procedural and multi-platformed. The click-stream of intelligent platform should last for a long time to obtain a visual atlas with analytical significance. This requires students not only to participate in intelligent teaching according to specific teaching steps in class, but also to log in to various platforms to search, browse and view recommended learning resources after-school. However, although higher vocational colleges students have a strong subjectivity in learning, they are weak in self-regulation and self-management compared with undergraduate students in universities. They lack corresponding planning and regulation in their learning (Wang, 2021), which brings difficulties to multimodal data collection. As for their purposes for General English learning in colleges, a questionnaire survey was conducted to test the learners needs. The survey targeted 579 first-year students of different majors from a higher vocational college as the subjects. The group of participants consisted of 267 females and 312 males. Their ages ranged from 18 to 21 years. The average age was 19.4 years (mean = 19.40). As for the participants' purposes of learning English, table 1 below shows the results.

Table 1
Percentage of students' purposes for learning English

Item	Percentage
Communicating with foreign friends	21.6
Travelling abroad	41.8
Applying for jobs	74.6
Furthering education in the future	11.5
Understanding western culture	63.2

According to learners' needs analysis, higher vocational students mainly learn English for future job seeking and foreign cultural acquisition. Besides, they have strong thirst for knowledge, but weak learning initiative. They are strong in perception, but poor in cognitive foundation. They have strong desire for performance, but weak willingness to cooperate. Under such circumstances, it is difficult to obtain the data of group activities on platforms. According to the requirements of 2035 China Education Modernization for education modernization, the talents cultivated by higher vocational colleges must be social-oriented with high-quality technical skills required for economic production, service and management. Therefore, when recommending learning resources on learning platforms, it should not only meet the learners' learning preferences, but also meet the requirements of all-around education of higher vocational education. All these put forward higher requirements for the collection and integration of resources.

4. STRATEGIES TO ENHANCE ENGLISH TEACHING IN HIGHER VOCATIONAL COLLEGES BASED ON MULTIMODAL ANALYTICS

4.1 Build a Multimodal and Data-driven English Teaching System

One of the strategies to enhance English teaching in higher vocational colleges based on multimodal learning analytics is to build a multimodal and data-driven English teaching system. The first and foremost is the digitalization of teaching resources. The teaching resources may not be limited to text books. They can also include pictures, videos, animation, micro lessons related to the teaching theme, so that the teaching content can be presented in a variety of ways to stimulate students' interests and catch their attention with multiple senses. The second is the intellectualization of teaching methods. The existing smart teaching environment of the school can carry out theoretical teaching and practice. Through screen recording, manual observation and other methods, it can meet the needs of multimodal data collection. Through Chaoxing Learning platform, the smart classroom is built by integrating pre-class activities and guidance, in-class activities, and after-class assignment. A variety of information teaching means can be adopted, such as micro class, Praat voice software, iwrite software, to realize online and offline blended teaching. For example, in the teaching design of the module "Home and House" in the course of Higher Vocational English, we had online and offline teaching assisted by Chaoxing Learning platform with such information means as iwrite

software, micro-class, animation and video. Meanwhile, the current popular topics such as rural poverty alleviation and craftsmanship initiative were also introduced into topic discussion. Through this teaching design, students can understand the beauty of architecture, experience the craftsmanship and traditional virtues contained in traditional Chinese architecture, so that their cross-cultural awareness can be cultivated, and the ideological and political education can be realized.

4.2 Build a Reasonable Evaluation System of English Learning Abilities

The weighting and interpretation of multiple data sources have always been a difficult problem in multimodal learning analytics. Therefore, a reasonable evaluation system of English learning abilities based on multimodal learning analytics is suggested to build. The examination and evaluation of English learning in higher vocational colleges is an important means to test the quality of English teaching and promote the construction and development of courses. In view of multiple learning data on platform, the evaluation of English in higher vocational colleges combines formative evaluation and summative evaluation, in which formative evaluation accounts for 40% and summative evaluation accounts for 60%. Based on the learning data of Chaoxing learning platform,

iwrite writing platform and Praat acoustic software, the students' comprehensive language abilities of listening, speaking, reading and writing are evaluated, and their learning effects are also evaluated. The main intention of a formative evaluation is to help the students and their teachers to measure where they stay in the process of study. With this measure, the formative evaluation will take into consideration students' homework completion, platform clickstream and classroom performance before class, during and after class, with scoring based on teacher evaluation, platform evaluation and peer evaluation (See Table 2). Summative evaluation typically takes the form of tests and quizzes after they conclude the lesson. With this measure, the summative evaluation is project-oriented, and evaluates the project drills of each module from five aspects: students' knowledge acquisition, language practice, moral character, team cooperation, and task participation, and finally obtains the assessment score through manual evaluation of peer evaluation, teachers and foreign teachers. Taking the teaching module "Food" in higher vocational English as an example, the formative assessment of oral English is mainly composed of two parts, and the assessment objects include the completion of self-study of pre-class resources, and the video of local cuisine introduction.

Table 2
Formative Assessment of Speaking on the Module FOOD

Evaluation carrier	Evaluation phrase	Evaluation objects	Evaluation means	Evaluation scores
Chaoxing Learning APP	pre-class self-study	resources for self-study	platform evaluation (40%)	12
		video of local cuisine	teacher evaluation (60%)	18
Chaoxing Learning APP	in-class exploration	group discussion	peer evaluation	20
Praat acoustic software		pronunciation and fluency	platform evaluation	30
Chaoxing Learning APP	after-class knowledge expansion	assignment	platform evaluation	20
Total				100

4.3 Integrate the Characteristics of the Development of Foreign Language Teaching in Higher Vocational Colleges

If multimodal leaning analytics could be implemented in English teaching in higher vocational colleges, it is necessary to fully integrate the characteristics of foreign language courses in higher vocational colleges. Higher vocational education is not only a part of higher education, but also a part of vocational education. In recent years, the rapid development of vocational education has also put forward new requirements for the reform of foreign language curriculum in higher vocational colleges. One of the basic characteristics of foreign language teaching in higher vocational colleges is that the overall level of teaching objects is not that high, and the foreign language ability is uneven, which bring difficulties to multimodal leaning analytics, as the degree of participation and persistence of learners on

the intelligent learning platform is the prerequisite for data recording and data analysis. Secondly, multimodal leaning analytics also puts forward higher requirements for English teachers in higher vocational colleges. These teachers need to strengthen the training of multimodal technology and information literacy, at the same time higher vocational colleges should also continuously optimize faculty structure. In addition, with the high-quality development of vocational education, English teaching in higher vocational colleges should also continue to innovate with the development of the times. Today's higher vocational education advocates "school-enterprise cooperation and study-practice integration". If the reform of the English course can be integrated into the general background of higher vocational education, it is necessary to specify talent training programs and curriculum standards according to the job requirements of enterprises and the needs of students' career development.

Meanwhile, the whole teaching implementation process is closely related to learners' needs, combining general English courses with professional knowledge, introducing actual work cases of graduates in the workplace to teaching situations, and improving students' problem-solving ability and English communication ability in the workplace.

5. CONCLUSION

Multimodal learning analytics is the study which uses explicit data representation and implicit physiological information obtained by computer technology and intelligent platform to analyze learners' learning habits and push personalized learning resources. Common types of multimodal data in in the field of intelligent education include external behavior representation data, internal neurophysiological information data, human-computer interaction data and learning situational awareness data (Wang and Zheng, 2022). Applying multimodal data and intelligent technology to teaching has become an inevitable trend in the reform of higher vocational foreign language education in the new era. Through the analysis, the research finds that the multimodal application for English teaching in higher vocational colleges currently has problems such as the weighting and interpretation of multiple data resources, the protection of ethical privacy and the acquisition of platform data. Therefore, a multimodal and data-driven English teaching system and a reasonable evaluation system are suggested to build. The characteristics of the development of foreign language teaching in higher vocational colleges should also be considered. Only in this way can we promote the empowerment of education big data technology and promote the high-quality development of intelligent education project and higher vocational education.

REFERENCES

Johns, T. (1991). Should you be persuaded: Two samples of data driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing* (pp. 2). Birmingham: ELR.

Kress, G., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design.* London: Routledge.

Ma, Y. F., Yue, T. Y., & Di, X. (2020). The reform of classroom teaching in vocational education: A study of deep learning supported by multimodal data. *Communication of Vocational Education, 2020*(12), 17-25.

Peng, H. C., & Jiang, Y. Q. (2022). Multimodal data-enhanced research in education sciences: Development trajectory and challenges. *Chinese Journal of Distance Education, 2022*(09), 19-26;33;78.

Scherer, S., Worsley, M., & Morency, L. P. (2012). *1st International Workshop on multimodal learning analytics: Extended abstract. In Proceedings of the 14th ACM international conference on multimodal interaction* (pp. 353-356). New York: ACM.

Wang, Q. (2021). Exploring the teaching reform of higher vocational education based on the cognitive characteristics of higher vocational students. *Xue Zhou Kan Magazine, 2021*(27), 3-4.

Wang, Y. Y., & Zheng, Y. H. (2022). Multimodal data fusion: The core driving force to solve the key problems of intelligent education. *Modern Distance Education Research, 34*(02), 93-102.

Zhang, Z. H., He, M., & Han, Z. (2014). Construction and application of College English Multimodal-Corpus. *Shandong Foreign Language Teaching, 35*(03), 50-55.