ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

Studies on the EFL Assessment Framework Driven by Data Mining Technology

Wen Gong

School of Foreign Studies, Lingnan Normal University, Zhanjiang, Guangdong, 524000, China
403312677@qq.com

Abstract: *This essay looks at how data mining technology is used in the assessment of EFL instruction. The accuracy of traditional assessment methods are constrained because they rely on subjective indicators rather than objective criteria. Data mining technology can be utilized as a useful method to assess college English instruction and offer in-depth analysis of the procedure. By seeing patterns and trends in the data, it can assist in enhancing the assessment's accuracy and validity. Data mining techniques are used to find patterns in datasets and reveal correlations between variables. The literature on data mining and its applications to the evaluation of college English education is reviewed in this study. It also discusses the advantages of data mining and looks into the difficulties associated with its application. Additionally, this study suggests and effectively executes a way for data mining implementation in college English teaching assessment. It implies that data mining is a valuable technique that can enhance the timeliness and accuracy of the evaluation process. With regard to the potential of data mining for the evaluation of college English education, the research offers helpful insights for administrators and educators.*

Keywords: *EFL; assessment framework; data mining technology; U Campus; big data*

1. Introduction

In recent decades, the assessment of language proficiency has become increasingly complex as the demands for more accurate and reliable evaluation of language skills have grown. This has led to the development of a variety of assessment approaches, such as standardized tests and performance-based measures. However, these approaches have their limitations, such as the lack of flexibility to accommodate the needs of individual learners and the reliance on a single source of information (i.e., the test score). In order to address these limitations, a new approach to language assessment has emerged in recent years: the data-driven approach. This approach involves the use of data gathered through a variety of sources, such as student performance, classroom observations, and surveys, to inform decisions on assessment design, feedback to students and teachers, and evaluation of instructional effectiveness.

In the context of big data, data mining technology—a type of data processing technology—is being utilized to change college English teaching assessment. This can aid in the realization of English as a Foreign Language (EFL) smart teaching as well as the correlation of the implicit and explicit teaching assessment objectives and the improvement of the detection effect on the English language acquisition of EFL students. Promoting the development of a college education assessment system and effective monitoring of teaching in many areas has been a significant component of China's reform of higher education in recent years. In order for the data to serve as the foundation for teaching reform, scientific educational decision-making must be based on reasonable statistical analysis and interpretation of data.

This paper provides an overview of the current research on the data-driven approach to EFL assessment and its potential implications for the field of language assessment. It begins by introducing the concept of data-driven assessment and its application in language learning. It then examines the theoretical background of the data-driven approach, including the concept of data-driven instruction, the principles for successful data-driven assessment, and the research that has been conducted on its implementation in the EFL classroom. The paper also considers the various challenges that need to be addressed in order for the data-driven approach to be effectively implemented in the EFL context. Finally, it proposes a framework for data-driven EFL assessment that is based on the principles identified in the research.

ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

2. Research Background

2.1 Definitions of Big Data and Data Mining

Big Data is a term used to describe any collection of large and complex data sets that require special handling to be analyzed and understood. Big Data characteristics such as volume, velocity, and variety differentiate it from traditional data models and make it difficult to be processed by conventional methods. Therefore, Big Data requires advanced technologies like artificial intelligence, machine learning, and data mining to process efficiently. Big Data is used for forecasting trends and behaviors, discovering unseen correlations, and optimizing services. Due to its massive size, velocity, and variety, Big Data has revolutionized organizations, businesses, and governments by providing a wealth of previously unavailable insights. As a result, Big Data has become an invaluable tool for making informed decisions in virtually every field.

Data mining technology is a powerful tool for analyzing large and complex datasets. It can be used to identify patterns and trends in student performance and to make predictions about future performance. Data mining technology can process large datasets quickly and can identify correlations that may not be apparent to the human eye. This makes it an ideal tool for precision teaching, as it can provide detailed insights into student learning. For example, it can be used to analyze student performance on standardized tests, performance assessments, and classroom observations. It can also be used to analyze student interactions with educational technology, such as online learning platforms. This information will be helpful to execute instruction and to identify areas of strength and weakness in student performance.

2.2 Overview of Data-driven EFL Assessment System

Data-driven EFL assessment framework is becoming increasingly popular in educational settings. This type of framework combines traditional assessment methods and big data to provide insight into students' language learning goals and progress. Using advanced analytics, data-driven EFL assessment framework can accurately measure and track student performance on a variety of language-related tasks. The system offers a comprehensive view of language knowledge and proficiency, allowing for more accurate and personalized feedback for students, teachers, and administrators.

The data-driven EFL assessment framework works by collecting and analyzing data from a variety of sources. This includes, but is not limited to, student scores on assessments, classroom activities, and online activities. The data is then analyzed to identify patterns and trends in student performance. This analysis helps to identify areas of strengths and weaknesses, which allows teachers to tailor learning objectives and activities to better meet the needs of their students.

The data generated by the EFL smart teaching platform is also used to create detailed reports for teachers and administrators. These reports provide valuable insight into student performance and progress, as well as trends in language learning. This information can help teachers to better understand student performance and design more effective instruction. Additionally, administrators can use the reports to identify areas where additional resources or support may be needed.

This type of system is particularly beneficial for teachers, administrators, and students alike, as it can provide valuable feedback and insight into language learning and progress. By leveraging the power of big data, EFL assessment framework can provide a more comprehensive and accurate picture of student performance and learning progress.

3. Literature Review

3.1 Previous Research on the Use of Data Mining in EFL Teaching

Data mining has been used in EFL teaching in recent years. This is due to the fact that data mining technology offers the potential to improve the effectiveness of the learning process. Data mining can help teachers to identify patterns and trends in a student's performance and to then tailor their teaching to meet the student's needs. One of the earliest studies on the use of data mining in EFL teaching was conducted by Lin et al (2015). This study investigated the experience of six early-career teachers who team-taught grammar to EFL college students using data-driven learning (DDL) for the first time. Another study, conducted by Öztürk et al (2017), examined the main sources of the participant English as a foreign language (EFL) teachers' cognitions, their classroom practices and the impact of institutional context on

ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

these practices. Technological mediation and data-driven decision-making have a particularly significant impact in learning environments because the education process primarily consists of dynamic information exchange. In Zeide's (2017) overview, he highlighted three significant structural shifts that accompany school reliance on data-driven instructional platforms that perform core school functions: teaching, assessment, and credentialing.

Boontam et al (2018) aim to investigate the effectiveness of paper-based data-driven learning (DDL) method in developing young Thai students in learning three English prepositions, which are 'during', 'among' and 'between', and to examine the attitudes of the participants towards learning through the DDL activities. Teachers should follow the development trend of big data, use data information technology to promote modern teaching mode, and realize The Times demand of data-driven teaching practice. In this regard, teachers in higher vocational colleges should follow the trend of The Times, update their inherent concepts, and use data technology to improve their big data literacy so as to better serve the teaching work (Zhen, 2019). A more recent study conducted by Cui (2021), designed a college English writing teaching model based on big data technology to highlight the teaching focus of lexical chunks. It is clear from reviewing the preceding literature that there is very little research on data-driven EFL assessment. Consequently, it is urgent and necessary to enhance the efficacy of L2 instruction based on the big data of student learning record.

3.2 Advantages of Data Mining Technology in EFL Assessment Framework

Data mining is becoming an important tool for educational assessment and data analysis in EFL. Data mining can help teachers and administrators identify trends in student performance, target students who might be at risk for failure, and make informed decisions about interventions and teaching strategies. Data mining can also help identify patterns of student behavior that can be used to inform decisions about curriculum, instruction, and assessment.

Data mining can enable more efficient and effective EFL assessment by providing teachers, administrators, and researchers with access to large amounts of data and the ability to quickly analyze and interpret the information. Data mining can provide an in-depth understanding of student performance and help teachers and administrators target students who might need more support. For example, teachers can use data mining to identify and analyze patterns of student behavior in order to identify areas of learning difficulty or student strengths.

Data mining can also help teachers, administrators, and researchers identify emerging trends in EFL assessment. By analyzing student performance data over time, teachers can identify patterns of student growth or decline, and make informed decisions about how to adjust instruction and assessment strategies. Data mining can also help identify trends in student performance across different contexts and countries, allowing for comparison of assessment results and the development of more effective teaching and assessment strategies.

Data mining can also provide a more complete picture of student performance by allowing for the combination of multiple data sources. By combining data from multiple sources, teachers, researchers, and administrators can gain a holistic understanding of student performance and can make informed decisions about interventions or teaching strategies (Gong, 2023).

4. Research Design

4.1 Selection of appropriate data mining techniques for EFL teaching assessment

Data mining techniques is a powerful tool to assess and improve effectiveness of English as a Foreign Language (EFL) teaching. In this section, we will discuss the different data mining techniques which can be used in the assessment of EFL teaching.

Data mining techniques can be broadly classified into two categories: supervised and unsupervised. Supervised data mining techniques are used when there is labeled data available, while unsupervised data mining techniques are used when the data is unlabeled. Supervised data mining techniques can be used to assess the performance of EFL learners. For example, classification algorithms can be used to identify patterns in student performance data. Logistic regression, decision trees, and support vector machines are some of the most commonly used classification algorithms for this purpose. These algorithms can be used to classify student performance into different categories, such as high achievers, average achievers, and low achievers. This can help in identifying areas where the learners need more help and attention.

ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

Clustering algorithms can also be used to assess the performance of EFL learners. These algorithms are used to group together similar students based on their performance data. This can help in identifying the strengths and weaknesses of different groups of students, as well as any common patterns among them. K-means clustering and hierarchical clustering are some of the most commonly used clustering algorithms for this purpose.

Unsupervised data mining techniques can also be used to assess the performance of EFL learners. Association rule mining is one of the most commonly used unsupervised data mining techniques for this purpose. This technique can be used to identify associations between different student performance metrics, such as grades, attendance, and test scores. These associations can help in understanding the different factors that influence student performance, as well as in identifying areas where further improvement is needed. In addition, natural language processing techniques can be used to analyze student feedback and assess the effectiveness of teaching methods. Similarly, text mining techniques can also be utilized to analyze student essays and identify any areas of improvement.

4.2 Gathering appropriate data for analysis

Gathering appropriate data for analysis of EFL teaching involves a process of collecting and analyzing information from multiple sources. The data collection process can include quantitative and qualitative methods, as well as a variety of data collection techniques.

A professional English learning data and processing center is established to record and store huge amounts of teaching and learning data. The U Campus Smart Education Cloud Platform (see Figure 1) can meet the demand, with a multi-dimensional and full-chain construction of a smart foreign language teaching platform that is connected online and offline and integrated inside and outside the classroom. The smart learning environment senses learning situations, identifies learner characteristics, automatically records the learning process and evaluates learning outcomes, collects personalized learning data such as students' classroom data, assignment completion data, the degree of mastery of each knowledge point, and the length of learning in a senseless manner, and closely integrates learning behaviors with evaluation objectives. By breaking the time and space constraints of the classroom, it builds an integrated learning interaction space before - during - and after class. It also provides uninterrupted access to massive, heterogeneous, highly concurrent and multi-dimensional data to understand the complexity and time-varying nature of real learning situations, track the learning process from all angles and provide a basis for teachers' teaching decisions.

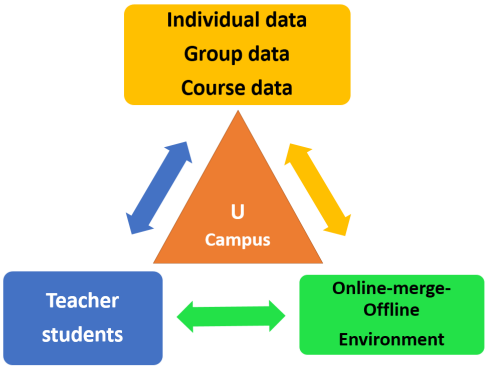

Figure 1. U Campus Smart Teaching Platform

Data collection techniques involve the use of instruments to collect information. These can include checklists, questionnaires, and rating scales. These techniques can be used to measure student learning, evaluate the quality of instruction, and identify areas of improvement. Data analysis can include statistical analysis, content analysis, and visual methods. By analyzing data, the EFL practitioners can uncover trends and patterns that can be used to make sound decisions about EFL teaching (Gong, 2022 a).

4.3 Flowchart of data-driven EFL Assessment Framework

The flowchart outlines the specific steps to integrate data mining into a college English teaching assessment system. The first step is to collect and analyze college English teaching data, including student answers and grades, test scores, and other information related to college English teaching. This data can be collected from a variety of sources, such as online resources, student surveys, and existing

ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

assessment systems. The collected data must then be processed to extract meaningful insights. This can be done using a variety of methods, such as text mining, natural language processing (NLP), and machine learning. Once the data is collected and processed, the next step is to identify areas for improvement in the assessment system. This can be done by analyzing the data to identify any gaps or deficiencies in the assessment system. The data can also be used to identify any trends or patterns that can be used to improve the assessment system. The fourth step is to develop data mining models and algorithms to analyze the data. This includes developing algorithms to identify patterns, trends, and insights from the data. These models and algorithms can then be used to analyze the data and extract meaningful information. The fifth step is to integrate data mining models into the assessment system. This involves integrating the data mining models into the assessment system so that the data can be analyzed and used to inform decisions about student performance. This integration can also include training stakeholders on the use and implications of data mining. The sixth step is to monitor the data mining models and adjust them according to changes in the data. This is an ongoing process that requires regular monitoring and updating of the data mining models to ensure that they are providing accurate and up-to-date insights. The seventh step is to utilize data mining models in the assessment system. This involves making use of the data mining models in the assessment system to identify patterns and trends in EFL student performance and to inform decisions about student success. The final step is to apply the results of data mining models to improve the assessment system.

By following the steps outlined in this flowchart, educators can integrate data mining into their college English teaching assessment framework to generate more meaningful insights and improve the assessment system.

5. Applications of Data Mining in EFL settings

5.1 FiF Spoken English Training System

One application of data mining technology is automated speech recognition (ASR). This technology uses Natural Language Processing (NLP) to analyze spoken responses and detect the presence of specific linguistic elements or words. It can also detect intonation and pitch, which can provide valuable feedback on the fluency and accuracy of student's speech. Natural Language Processing (NLP) is a branch of artificial intelligence that focuses on analyzing natural language data. NLP can analyze spoken responses and detect the presence of specific linguistic elements or words. It can also detect intonation and pitch, which can provide valuable feedback on the fluency and accuracy of student's speech.

NLP can be used to analyze spoken language by breaking it down into its component parts, such as individual phonemes, syllables, words, and phrases. It can then detect the presence of specific linguistic elements or words. This can be done by using a variety of algorithms, such as Hidden Markov Models (HMMs), which can be used to detect both the presence of certain words as well as the order in which they were spoken. Furthermore, NLP can be used to analyze the intonation and pitch of speech. This can provide valuable feedback on the fluency and accuracy of a student's speech (see Figure 2). NLP can also be used to identify and classify speech patterns. This can be used to detect the presence of certain linguistic elements or words. It can also be used to identify certain emotions or feelings. This can provide valuable feedback on the effectiveness of a student's communication.

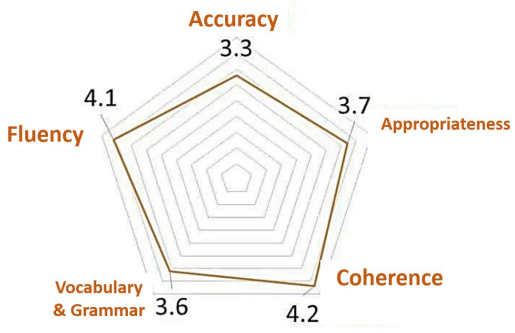

Figure 2. Profiles of Students' Spoken English Proficiency

Furthermore, NLP can be used to detect errors in spoken language (see Figure 3). This can be done by using a variety of algorithms, such as neural networks, which can be used to detect irregularities in a person's speech. This can be used to provide feedback on a student's pronunciation and fluency.

ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

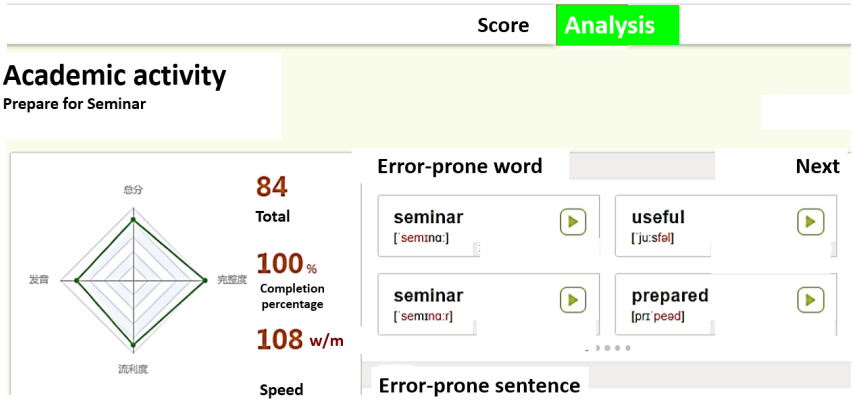

Figure 3. Detected Errors in Spoken Language

Finally, the data mining will provide teachers with horizontal comparison results of spoken language training values for different classes (Figure 4). This generates a clear and precise profiling of each class for teachers and will enable them to offer students with adaptive learning suggestions in different classes.

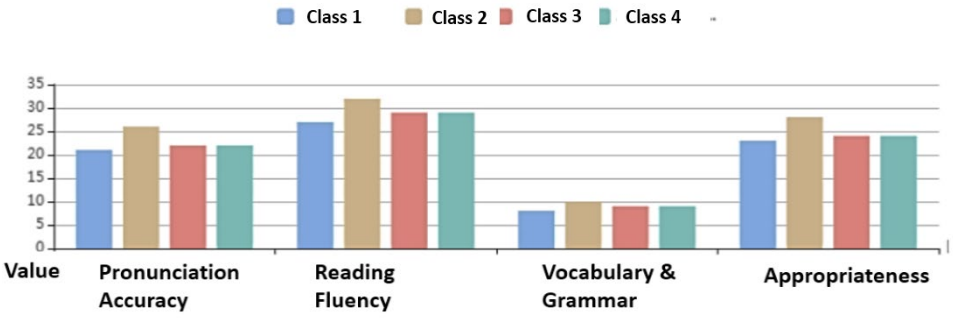

Figure 4. Horizontal Comparison of Spoken Language Skills Training Values

5.2 iWrite English Writing Teaching and Grading System

Another application of data mining technology in the assessment system of college English teaching is the use of automated essay scoring (AES). This technology uses natural language processing (NLP) to analyze the structure, grammar, syntax, and content of student writing. It can quickly grade essay answers and provide feedback on specific areas of improvement. With the integration of artificial intelligence (AI) and machine learning, AES can be trained to identify subtle aspects of writing and score accordingly. iWrite English Writing Teaching and Grading System (Figure 5) is a case in point. The four dimensions of language, content, chapter structure, and technical specifications can all be intelligently reviewed by machines. It can also combine human grading and machine grading, encourage feedback, and enhance teacher-student interaction in education.

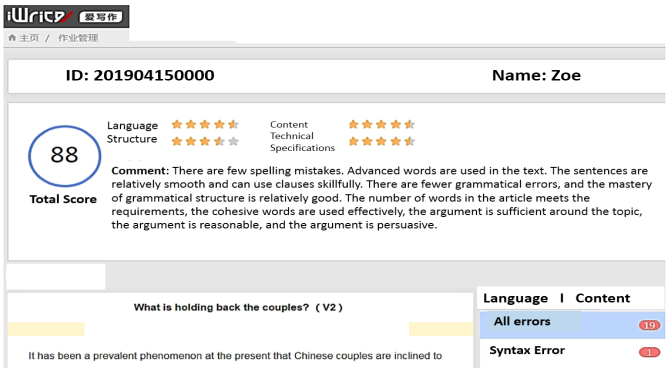

Figure 5. iWrite English Writing Grading System

ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

6. Discussion

6.1 Significance of Data-Driven EFL Assessment Framework

Data mining enables more effective and efficient assessment of English language proficiency, allowing instructors to better target instruction and support for their students. By applying sophisticated algorithms to analyze large volumes of student data, data mining can identify correlations and trends between student performance, course content and assessment criteria. This helps instructors to more accurately assess student proficiency and progress, informing feedback and the development of corrective strategies to address any issues.

Data mining also allows for more comprehensive and in-depth analysis of student performance. By tracking and analyzing student performance across multiple assessments, instructors can monitor the effect of classroom instruction on student learning. This provides valuable insight that can be used to inform improvements to the instruction and assessment system. Data mining can also be used to identify areas of weakness or difficulty in the course and provide teachers with deeper understanding of their students' performance.

Furthermore, data mining is a source of insight into the effectiveness of assessment criteria. By analyzing the effectiveness and consistency of assessment rubrics and standards, data mining can help instructors to refine their criteria and ensure the quality of their assessments. Data mining can also be used to identify valid and reliable measurements for student proficiency, and to develop consistent and effective assessment systems for EFL teaching. Therefore, data mining is a powerful tool for improving and refining the assessment system of EFL teaching. By leveraging sophisticated algorithms to track and analyze student performance and assessment criteria, data mining can help instructors to more accurately assess student proficiency and progress, target instruction and support, and develop effective assessment systems (Gong, 2022 b).

6.2 Challenges of Data-Driven EFL Assessment

Although data-driven assessment has the potential to improve the quality of language assessment, there are several challenges that need to be addressed in order for it to be effectively implemented in the EFL context. First, data-driven assessment requires a significant amount of time and resources, as it involves the collection and analysis of data from a variety of sources. Second, the use of data-driven assessment can lead to an over-reliance on assessment data, which can lead to a lack of focus on other important aspects of language learning, such as communication, social interaction, and cultural knowledge.

6.3 Recommendations for further research

The use of data mining in EFL teaching is an interesting and promising area of research. The potential of data mining to increase student engagement, foster teacher productivity, and enhance learning outcomes are often seen in other areas of education, and the same could be applied to EFL teaching. One possible area of research would be to explore the impact of data mining on student motivation. As more and more data become available to teachers, it is important to understand how this data can be used to inform the teaching process and influence student motivation. Research could focus on the effects of using data mining to personalize instruction, foster better feedback loops, and track student engagement. By understanding how data mining can be applied to teaching, more effective approaches to EFL teaching can be developed.

Another possible research topic could be to analyze the use of data for formative assessment. As more information about student performance is collected, it is essential to understand how this data can be used to track student progress and inform instruction. Research into how data mining can be used to improve formative assessment and track progress over time could help teachers craft more effective instruction and make better use of their teaching time. Finally, research into how data mining can be used to facilitate cross-cultural communication would be a valuable addition to the current research. As EFL students become increasingly mobile and connected, the need to understand how data can help bridge cultural gaps grows in importance. Research into this area could help teachers design more effective lesson plans and activities that focus on improving cross-cultural communication.

ISSN 2522-6398 Vol. 6, Issue 4: 40-47, DOI: 10.25236/FER.2023.060408

7. Conclusion

Overall, data mining technology has the potential to revolutionize the way that EFL teachers approach the classroom. Research into the impact of data mining on student motivation, formative assessment, and cross-cultural communication could help teachers make better use of their teaching time and improve the overall quality of their instruction. The logical mechanism of data mining technology lies in the summary and analysis of students' English learning performance. The core of realizing the use of data mining technology in the college English teaching assessment system is to clarify the functional mechanism of data mining technology, establish a database of students' English performance to correlate data cleaning, data summarization and data conversion, and at the same time, promote the innovation of the EFL teaching evaluation mode, improve the data collection channels and realize the smart processing of college English teaching information and data (Gong, 2021). Based on the principles identified in the research, the paper has proposed a framework for data-driven EFL assessment. It is hoped that this framework will provide guidance to language teachers and assessment experts in the development of more effective data-driven EFL assessment practices.

Acknowledgement

This research was funded by "Intelligent Technology-Driven Research on Data Mining for Large-scale Testing of Foreign Languages", Teaching Reform Grant of Lingnan Normal University, 2021.

References

[1] Boontam P. & Phoocharoensil, S. (2018) "Effectiveness of English Preposition Learning Through Data-Driven Learning (DDL)". 3L The Southeast Asian Journal of English Language Studies, vol. 24, number 3.
[2] Cui Y. (2021) "Research on The Design of Lexical-chunks Centered Mode of Writing under Artificial Intelligence in College English Course". 2021 2nd International Conference on Big Data and Informatization Education, April.
[3] Gong W. (2023). "Reshaping the EFL Formative Assessment with Blockchain Technology". International Journal of English Linguistics*, vol. 13, number 1.*
[4] Gong W. (2022 a). "An Empirical Study of College English Smart Teaching Driven by Intelligent Technology. Asia-Pacific Journal of Humanities and Social Sciences*, vol. 2, number 3, 165-172.*
[5] Gong W. (2022 b). "Research on Foreign Language Cloud-based Examination Driven by Intelligent Technology." Advances in Social Science, Education and Humanities Research, volume 637, 205-210.
[6] Gong W. (2021). "Design and Implementation of EFL Blended Smart Teaching Based on Rain Classroom." Frontiers in Educational Research, volume 4, issue 9, 43-49.
[7] Öztürk G. & Gürbüz, N. (2017) "Re-defining Language Teacher Cognition Through A Data-driven Model: The Case of Three EFL Teachers", Cogent Education, vol. 4, issue 1.
[8] Lin M. H. & Lee J.Y. (2015) "Data-driven Learning: Changing the Teaching of Grammar in EFL Classes". ELT Journal, vol. 69, number 3.
[9] Zeide E. (2017) "The Structural Consequences of Big Data-Driven Education". Big Data, vol. 5, number 2.
[10] Zhen Li. (2019) "The Formation of Core Qualities of Teachers in Higher Vocational Colleges under Big Data". Science Innovation, vol. 7, issue 1.