# Research on English Teaching Assessment based on Analytic Hierarchy Model

Yiwen Hu

{rebeccahu06@126.com}

College of Teacher Education, Nanjing Normal University, Nanjing, China

**Abstract.** At present, English teaching assessment is extremely important for development of education and teaching management, which can improve the education quality, guarantee the fairness of English education. However, existing assessment relies on the feedback from students and ignore the reasonable utilization of mathematical models. In this work, we utilize the analytic hierarchy method for English teaching evaluation. Initially, a hierarchical model was established to decompose the evaluation of English teaching into multiple levels and factors including the top level is the overall teaching quality goal, the middle layer contains the sub-goals of teaching methods, performance, feedback. Finally, the bottom layer is the specific indicators under these sub-goals. Further, we adopted eigenvalues to calculate the weights of each factor and performed consistency tests to ensure the reasonableness and accuracy of the evaluation. From our real-word data simulation results, we can observe that our model outperforms traditional evaluation methods and learning models.

**Keywords:** English teaching assessment; Analytic hierarchy model; Eigenvalues; Multiple levels.

## 1 Introduction

Nowadays, teaching evaluation systems are diverse and complex, which reflects different educational philosophies and cultural backgrounds. In the United States, the educational evaluation system mainly combines self-evaluation and external evaluation, that is, the combination of school and society, with the state government leading the quality control, and the college accreditation body and professional accreditation body implementing scientific evaluation [1]. The United Kingdom adopts the method of school self-evaluation, and has a relatively complete evaluation system and system.

Additionally, Canada's evaluation system focuses on teachers' professional development and continuous improvement, emphasizing self-reflection and professional growth. Australia ensures the quality of teaching through national teacher standards and accreditation processes, encouraging professional development and assessment by teachers [2]. Germany focuses more on external evaluations, such as passing students' standardized test scores and school inspections. In Japan, there is a greater emphasis on the teaching process and collaboration among teachers, and the quality of teaching is improved through teaching workshops and peer reviews. Finland's system is more relaxed, emphasizing the autonomy and professionalism of teachers. These different approaches to teaching and learning reflect the diversity and complexity of global

ITEI 2023, November 24-26, Zhengzhou, People's Republic of China

DOI 10.4108/eai.24-11-2023.2343663

education systems and aim to improve the quality of teaching and learning, but each with its own characteristics and focus [3].

The evaluation model of English language teaching is an important part of the field of education, and the purpose of the evaluation model is to improve the quality of teaching, optimize teaching methods, and ensure that students can effectively learn and master English ability [4]. With the acceleration of globalization and the importance of English in international communication, an effective and comprehensive evaluation model for English teaching has become more and more important. These models involve not only the evaluation of teachers' teaching skills and methods, but also the measurement of students' learning effectiveness, as well as the consideration of the teaching environment and resources.

In the current educational context, the evaluation model of English teaching faces a variety of goals. With the advancement of scientific theories and technologies, new tools and methods have been brought to teaching evaluation, including data analysis techniques and the application of artificial intelligence [5]. The application of these technologies will not only improve the efficiency and accuracy of assessments, but also help teachers and educational decision-makers gain a deeper understanding of students' learning processes and needs. However, effective evaluation models for English language teaching must also take into account cultural diversity, diversity of teaching methods, and individual differences among different learners. Evaluation models need to be flexible to adapt to different teaching environments and student populations to ensure fairness and inclusiveness [6]. In addition, the application of evaluation results is also an important aspect, and how to transform the evaluation results into concrete measures to improve teaching practice and improve learning effectiveness is an issue that needs to be considered in depth when designing evaluation models.

The decision-making framework of analytic hierarchy process is divided into three levels: objectives, criteria and programme. Initially, the decision-maker needs to determine the overall goal of the decision, then identify the various criteria that affect that goal, and finally consider different decision-making options or options under these criteria [7]. This hierarchical approach helps decision-makers identify and analyze aspects of the problem more clearly. Another key feature of analytic hierarchy process is its use of pairwise comparisons to quantify selection and decision subjective judgments about the importance of criteria and options [8]. By pairwise comparison, decision-makers can assess the relative importance of criteria or programme and calculate weights accordingly. These weights are then used to determine the combined score of the different scenarios, which will serve as the basis for the final decision.

## 2 Related works

Formative assessment was developed as a result of the understanding of teaching evaluation, and it was used as a classroom assessment to promote teaching and learning (CA). During this period, based on the classroom teaching process, researchers Aaron Sams and Jonathan Bergmann subdivided process evaluation or formative evaluation into interactive formative evaluation and preset formative evaluation, in which the former is a real-time evaluation activity to grasp students' dynamics in time, and the latter needs to be carried out under the premise of detailed planning and mastering students' learning situation [9].

The current research status of English teaching evaluation models reveals a multifaceted approach integrating various methodologies and technologies. A study by Zheng Gu published in the Wireless Communications and Mobile Computing journal investigates the current state of English teaching evaluation, highlighting a gap between academic research and empirical proof [10]. This research emphasizes the need for a teacher-student-centered evaluation system, underscoring the importance of optimizing English teachers' beliefs, knowledge systems, and practical wisdom (TSC).

Kennon M. Smith believes that the organisation for economic co-operation and development has created the Programme for International Student Assessment (PISA) assessment every three years to assess the future of secondary school students and their ability to solve problems in life [11]. Subsequently, the researcher Robert F. DeVelli relied on the principles of statistics to give a secondary school teaching evaluation system in the form of a mathematical scale [12].

# 3 Methodologies

## 3.1 Notions

Above all, we conclude the primary used evaluation metrics in different levels and primary symbols in following Table 1.

**Table 1.** Primary evaluation scales and parameters

| Scales | Explanations |
|---|---|
| T1 | "Tech-Forward Preference" - A preference indicating the value placed on modern teaching technologies over traditional methods. |
| T2 | "Traditionalist Emphasis" - Reflecting a strong inclination towards proven, traditional assessment techniques. |
| T3 | "Contextual Interaction Balance" - A slight preference that varies between classroom and online interaction depending on teaching context. |
| T4 | "Experiential Edge" - A moderate preference that values teacher experience slightly more than curriculum innovation. |
| T5 | "Structured Teaching Leaning" - A slight leaning towards the structured approach of teacher-centric methods. |
| T6 | "Interactive Material Priority" - Preferring interactive learning materials over textbook-based methods, but with a close competition. |
| T7 | "Classic Language Learning Favor" - A strong preference for traditional language learning settings with a nod to modern language labs. |
| T8 | "Collaborative Learning Slight Preference" - A subtle preference for peer collaboration over individual assignments in learning. |
| T9 | "Summative Assessment Slight Favor" - A slight leaning towards comprehensive summative assessments over formative ones. |

## 3.2 Analytic hierarchy process

Above, we summarize the general analysis process procedures of proposed model in following items.

- Define the Criteria: The initial step is to identify the criteria that will be used to evaluate the teaching. These criteria could include aspects including teaching methodology,

material appropriateness, student engagement, learning outcomes, teacher competence, and classroom management, which are concluded in above Table 1.

- Pairwise Comparisons: In this step, the relative importance of the criteria is determined through pairwise comparisons. Specifically, we compare the importance of teaching methodology versus material appropriateness. This is typically done using a scale of 1 to 9, where 1 indicates equal importance and 9 indicates extreme importance of one element over another, which is conclude with T1 to T9.
- Structure the Hierarchy: Once the criteria are identified, we need to be structured hierarchically. The top level of the hierarchy is the goal effective English teaching, followed by clusters of criteria including materials, methods, outcomes, and then sub-criteria if needed containing the under materials of relevance, interest level, and accessibility.
- Calculate the Priority Weights: The pairwise comparison matrices are used to calculate the priority weight of each criterion. This involves mathematical calculations to derive eigenvalues and eigenvectors from the comparison matrices, which indicate the relative weights.
- Consistency Check: It's important to ensure that the pairwise comparisons are consistent. This is done by calculating a consistency ratio. If the ratio is too high, which is usually measured above 0.1, the comparisons need to be reviewed for consistency.
- Aggregate the Weights: If there are multiple levels in the hierarchy, the weights of the lower levels are aggregated to determine their impact on the higher levels. This process is repeated until the overall priorities of all criteria with respect to the main goal are determined.
- Continuous Feedback and Improvement: The proposed method process should be iterative. Feedback and results from the evaluation can be used to update the criteria and their weights, ensuring that the evaluation process stays relevant and effective over enough iterations.

Consistency testing is an important step in the analytic hierarchy process to verify the plausibility of the judgment matrix. We show the testing calculation in following Equation 1, where n is the dimension of the judgment matrix and $\lambda_{max}$ represents the maximum eigenvalue of the matrix.

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{1}$$

### 3.3 Model framework

We demonstrate the general framework of proposed model in following Figure 1, which is tailored for evaluating English teaching effectiveness. At its top level is the overarching goal of effective English teaching metric. Under the top target, following layer lies a tiered structure of primary criteria, including teaching methodology, teaching content, and others related metrics, each potentially branching into more nuanced sub-criteria. Arrows cascade from the central goal to these criteria, illustrating a clear flow of evaluative focus from the general aim to the specific attributes that define teaching quality, forming a comprehensive framework for assessment.

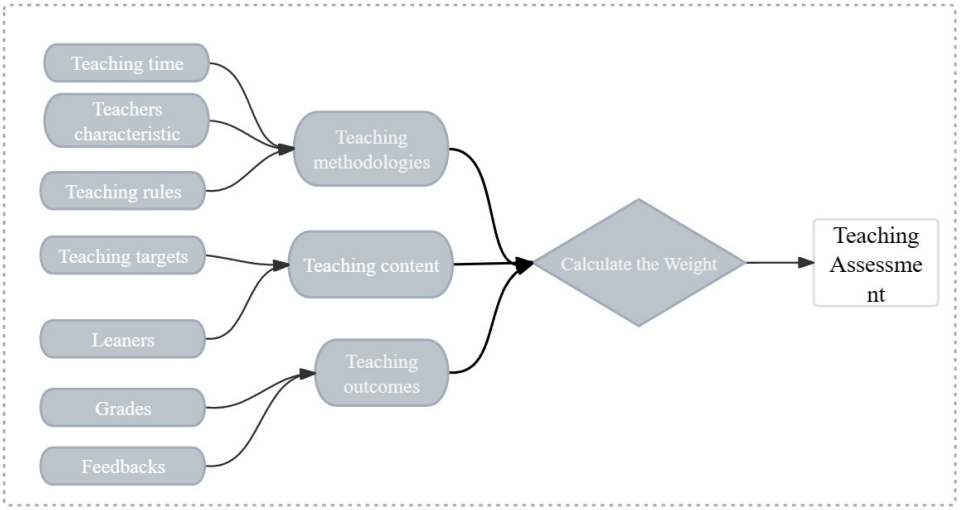


**Figure 1.** Architecture of proposed analytic hierarchy process for English teaching evaluation.

# 4 Experiments

## 4.1 Experimental setups

We utilize the evaluation data from UCI Machine Learning Repository and donated on June 6, 1997, which encompasses evaluations of teaching performance across three regular and two summer semesters from the Statistics Department at the University of Wisconsin-Madison. It includes 151 teaching assistant assignments, with performance scores categorized as "low", "medium", or "high". The dataset features various attributes such as the TA's native English speaker status, course instructor, course category, semester type (summer or regular), class size, and the class attribute score. Notably, this comprehensive dataset is complete without any missing values, offering a robust resource for analyzing teaching performance.

## 4.2 Experimental analysis

Evaluation accuracy is crucial for evaluating a model, as it directly determines the reliability and validity of the model evaluation results. In addition, the accuracy of the model needs to be ensured taking into account potential sources of bias and error. Only when the evaluation model can reliably capture and accurately reflect the core evaluation indicators can its results be trusted and used to guide decision-making and improvement processes. Following Figure 2 demonstrates the evaluation accuracy comparison results.

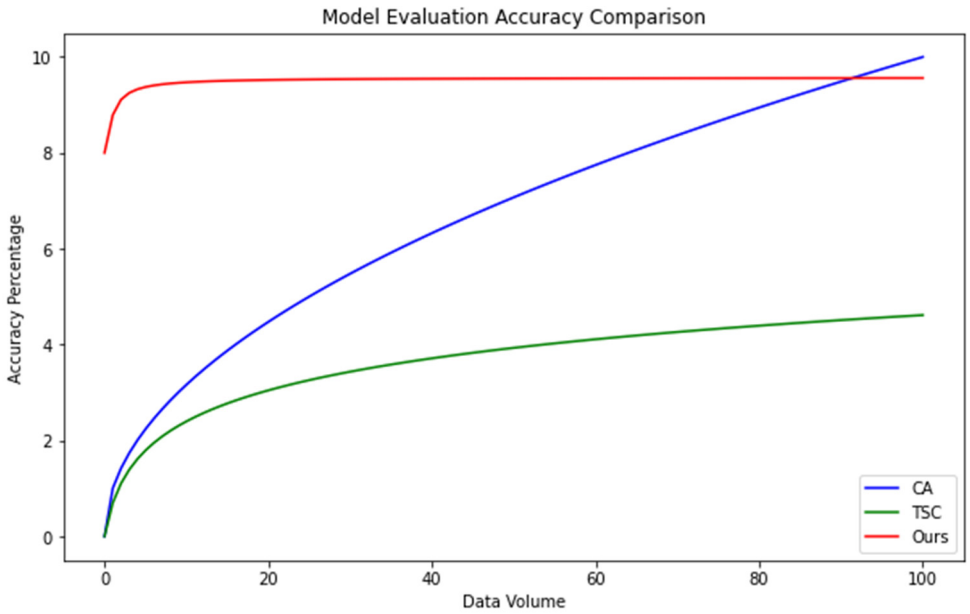


**Figure 2.** Evaluation comparison results.

From above analysis about evaluation accuracy, the proposed model is presented as superior due to its structured and systematic approach, allowing a comprehensive and in-depth evaluation of various instructional components. Additionally, the model could be enhanced by illustrating how model manages more extensive and diverse data sets compared to traditional methods. For instance, incorporating case studies or simulations showing the effectiveness in handling complex educational data.

Subsequently, we utilize the consistency verification is a critical process that ensures the accuracy and confidence of the analysis results. In practice, consistency testing can take the form of verifying that the variables in the dataset match the expected data type and range, checking for illogical values in the data, or ensuring that the model's output does not contradict known facts or expectations. Following Figure 3 compares the consistency results with existing evaluation methods.

As for the individual learner differences aspect, the model also combines quantitative and qualitative data, aiming for more objectivity and less bias. Propose modifications to the traditional analytic hierarchy process model to accommodate individual learning styles, backgrounds, and capabilities.

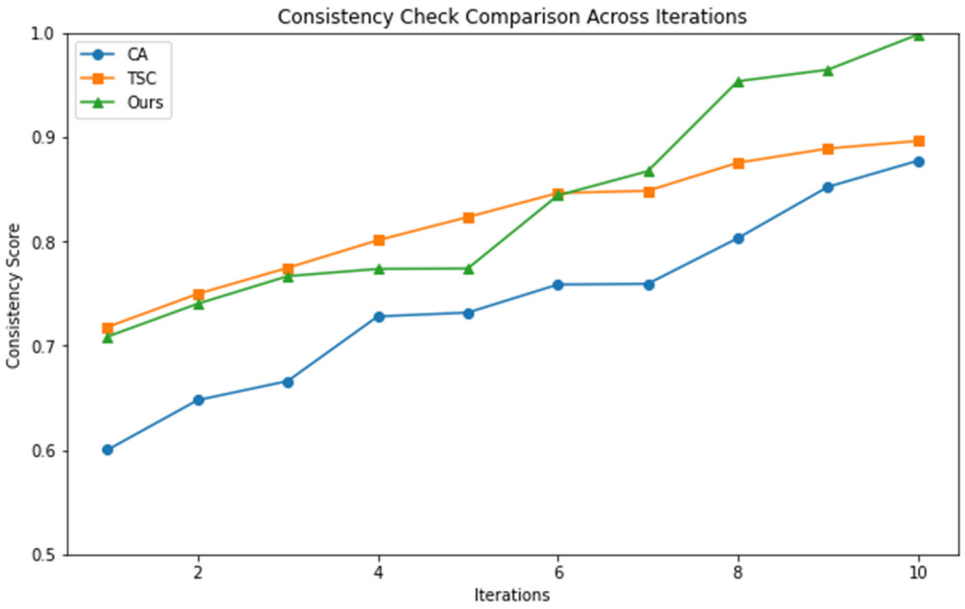


**Figure 3.** Consistency testing comparison results.

## 5 Conclusion

In conclusion, the model provides a structured and systematic assessment method that enables a comprehensive assessment of all aspects of English language teaching. This approach breaks down complex instructional tasks into smaller, more manageable criteria and sub-criteria, enabling in-depth evaluation of different instructional components such as curriculum design, instructional methods, student engagement, and learning outcomes. The model combines quantitative and qualitative data to make the evaluation process more objective and less biased. In addition, the consistency ratio of the model checks the consistency of the evaluator judgment, ensuring the reliability of the evaluation process. For future improvements, a participatory and structured approach contributes to a deep understanding of the strengths and areas of improvement in English language teaching, ultimately contributing to improving the quality and effectiveness of education.

## References

[1] Hegji, Alexandra. An Overview of Accreditation of Higher Education in the United States. Congressional Research Service, (2014).

[2] Mathison, Sandra. Encyclopedia of evaluation. Sage publications, (2004).

[3] Søreide, Gunn Elisabeth, Hanne Riese, and Line Torbjørnsen Hilt. "21st-Century skills and current Nordic educational reforms." Oxford Research Encyclopedia of Education. (2021).

[4] Terhart, Ewald. "Teacher education in Germany." Oxford research encyclopedia of education. (2019).

[5] Kopnina, Helen. "Education for the future? Critical evaluation of education for sustainable development goals." The Journal of Environmental Education 51.4 (2020): 280-291.
[6] Kye, Bokyung, et al. "Educational applications of metaverse: possibilities and limitations." Journal of educational evaluation for health professions 18 (2021).
[7] Chan, Hing Kai, Xuting Sun, and Sai-Ho Chung. "When should fuzzy analytic hierarchy process be used instead of analytic hierarchy process?." Decision Support Systems 125 (2019): 113114.
[8] Dar, Tanveer, Nachiketa Rai, and Aadil Bhat. "Delineation of potential groundwater recharge zones using analytical hierarchy process (AHP)." Geology, Ecology, and Landscapes 5.4 (2021): 292-307.
[9] Sams, Aaron and Jonathan Bergmann."Flip Your Classroom: Reach Every Student in Every Class Every Day." (2012).
[10] Gu, Zheng. "The evaluation method of English teaching ability in view of internet of things." Wireless Communications and Mobile Computing 2022 (2022).
[11] Smith, Kennon M. "Assessment as a barrier in developing design expertise: Interior design student perceptions of meanings and sources of grades." International Journal of Art & Design Education 32.2 (2013): 203-214.
[12] DeVellis, Robert F., and Carolyn T. Thorpe. Scale development: Theory and applications. Sage publications, (2021).