



The Effectiveness of Public Speaking Instruction in Enhancing Students' Confidence and Communication Skills

Yuyu Liao^{1,*}

¹ Jiangxi University Of Technology, Nanchang, Jiangxi, China, 330098

SUMMARY: *Public speaking teaching is not only related to the completion quality of students 'oral expression, but also directly affects their self-presentation ability in classroom, academic communication and professional scenes. Aiming at the problems of single evaluation dimension, lagging process feedback and difficult to continuously identify ability changes in traditional teaching, this paper takes the public speaking course of colleges and universities as the object, constructs a teaching effect analysis framework integrating questionnaire, voice, video and performance data. Python was used to complete multi-source data preprocessing, standardization, paired sample t-test and multi-modal comprehensive evaluation. The results showed that after 12 weeks of teaching intervention, the total score of students 'self-confidence increased from 3.21 to 3.86, the comprehensive evaluation score increased from 70.84 to 81.27, and the public speaking test score increased from 72.64 to 84.38. The relevant changes showed that public speaking teaching had a significant promotion effect on the formation of students 'self-confidence, the improvement of expression fluency and the enhancement of field control ability. The significance of this paper is that the introduction of computer supported multimodal analysis into public speaking course evaluation provides a more detailed evidence base for teaching diagnosis and ability cultivation.*

KEYWORDS: *Public speaking teaching; Confidence; Communication skills; Multi-modal evaluation*

1 Introduction

Public speaking is not a simple manifestation of language knowledge, but a comprehensive ability that involves content organization, oral expression, emotional regulation and on-the-scene interaction. For college students, poor performance in public speaking is often not caused by "not being able to speak", but in the aspects of insufficient self-efficacy, elevated anxiety, feedback lag and single training situation, which also makes the teaching of public speaking face the practical problem of "a lot of practice, but unstable improvement" for a long time. Recent studies have shown that classroom presentation, situational tasks, collaborative adjustment and real audience participation can improve students' oral expression quality to varying degrees, and have a continuous impact on speech confidence [1]. This shows that the effectiveness of public speaking teaching cannot be judged by a single presentation, but should be examined in a more complete process of competence generation.

At the same time, digital technology is changing the way spoken language training is organized. Video recording tasks, virtual reality situations, online feedback platforms and

*kellyyuyu123@sina.com

<https://doi.org/10.65102/is2026707>

multimodal learning analysis tools have gradually entered the speech teaching scene, enabling teachers to observe students' performance from multiple dimensions such as speech clarity, speaking rate control, stop and connect rhythm, gaze allocation, posture stability and audience response [2]. Compared with the traditional practice of relying on teachers' experience to score, this kind of computer-supported teaching path can transform the scattered, subjective and difficult to track classroom performance into data records that can be collected, compared and backtracked, thus providing more detailed basis for teaching intervention [3]. Especially in college English and general expression courses, with the help of speech recognition, video slice analysis and multi-source data standardization processing, we can more clearly distinguish the two interrelated but different teaching objectives of "Dare to speak" and "can you speak well".

Based on this, this paper focuses on the promotion effect of public speaking teaching on students' self-confidence and communication ability. It not only focuses on the changes in students' psychological feelings and test scores before and after training, but also introduces the multi-modal computational analysis method to comprehensively evaluate the language, sound and behavioral signals in the speech, so as to construct a more complete teaching effect judgment framework. Research attempts to answer the question: when the teaching of public speaking changes from single-display training to continuous diagnosis supported by data, whether students' self-confidence improvement, expression optimization and improvement of comprehensive communication ability can be supported by clearer and more reliable evidence [4]. This discussion not only has the significance of curriculum reform, but also provides a verifiable practice path for computer technology to participate in the training of language and communication ability.

2 Public speaking teaching and ability improvement

The reason why public speaking teaching deserves to be discussed separately is that it does not correspond to a single speaking skill, but to a set of interconnected abilities, including information screening, structure organization, on-site expression, emotion regulation, audience perception and instant correction. In many cases, students' mistakes in speech are not due to lack of knowledge reserve, but due to the lack of stable coordination among language output, psychological state and non-verbal behavior. Traditional classroom often promotes training by the way of "assigning problems-students' show-teachers' comments", which can cover the basic teaching process, but it is difficult to continuously record students' progress trajectories at different periods of time, and it is not easy to distinguish whether the problem is in content construction, sound control or stage response. With the introduction of digital teaching tools into language courses, public speaking training gradually changes from experience-led to data-assisted. Video review, online platform, virtual situation simulation, voice transcription and multimodal feedback are used for speech performance analysis, which makes teaching change from one-time evaluation to process diagnosis [5]. In this context, the effectiveness of public speaking teaching is no longer only reflected in the final performance, but more reflected in whether the students' confidence level is improved, whether the expression is clearer, and whether the interaction is more relaxed, and whether these changes can be reliably presented by the evidence chain supported by the computer.

From the perspective of ability improvement mechanism, public speaking teaching acts on at least two levels. The first is psychological level, that is, through repeated exposure, peer cooperation, teacher feedback and task progression, students' avoidance tendency to public expression is reduced and their self-efficacy is gradually improved. The second is the behavioral level, that is, with the help of speech script training, discourse structure design,

video review, sound rhythm analysis and posture observation, students are encouraged to establish a more stable mapping relationship between language organization and expression execution [6]. Especially in computer supported environments, speech recognition can help extract speech rate, pauses, filler word frequency, and articulation, and video analysis can help identify eye distribution, body sway, gesture use, and stance stability. In this way, the original vague "expressiveness" and "confidence" may be broken down into a number of observable, quantifiable and comparable indicators, and then enhance the explanatory power of teaching evaluation.

2.1 Review of related research

Existing researches generally focus on three paths: contextual training, technical intervention and feedback optimization. Some studies emphasize the role of real audiences, dramatic tasks and display activities in promoting the growth of oral expression, and believe that the improvement of speech ability does not completely depend on the language knowledge itself, but is also closely related to the authenticity of the task and the depth of participation [7]. Some studies also focus on anxiety intervention, and reduce students' public expression tension by means of virtual reality, gaze exposure or simulated classroom, so as to create more stable psychological conditions for subsequent expression training [8]. Some achievements in the past two years show more obvious computer-aided characteristics, such as using the Blackboard platform to organize demonstration training, using video recording homework to strengthen self-observation, and using video-enhanced multimodal feedback to improve students' perception of speech details [9]. Together, these studies illustrate that the teaching of public speaking is moving from the primary goal of "speaking well" to the composite goal of "speaking clearly, appropriately, and with a sense of control".

However, there are still two obvious shortcomings in the existing results. First, many studies focus more on anxiety relief or single oral performance improvement, but less on confidence, communication skills and classroom performance in the same evaluation framework, resulting in scattered evidence of teaching effectiveness. Secondly, although technical tools have been widely introduced, many researches still stay at the level of tool use, and the discussion on how to preprocess, standardize, integrate and serve multi-source data for teaching decision is not sufficient. The entry of technology into the classroom does not automatically equate to more scientific evaluation. Only when the information such as voice, video, questionnaire, and score are transformed into structured indicators, the actual effect of public speaking teaching may be more accurately identified. Based on this understanding, this paper attempted to incorporate the change of students' confidence, composition of expression ability, diagnostic indicators and comprehensive scores into the unified analysis chain on the basis of existing research, so as to present the effect path of public speaking teaching more completely.

Table 1: Review of research related to public speaking teaching

Reference	Research Method / Technical Approach	Main Findings	Limitations
[2]	Longitudinal task-based study using dramatized presentations and authentic audience settings	Long-term tasks help improve oral presentation quality and classroom participation	Strong context dependence; transferability to general classroom settings still needs verification
[4]	Public speaking training in a virtual reality environment with added distractors	Helps students adapt to high-pressure speaking situations and alleviates public speaking anxiety	Focuses more on anxiety intervention, with limited discussion of improvements in speech structure
[5]	Virtual reality therapy combined with counseling intervention	Positively reduces students' public speaking anxiety	Provides limited evidence on specific dimensions of communication competence
[6]	Presentation training program based on the Blackboard platform	Online platforms can increase practice frequency, feedback efficiency, and course organization	Platform effectiveness is influenced by learner engagement and digital literacy
[7]	Randomized controlled study of virtual classroom simulation	Simulated teaching scenarios help improve oral communication ability and self-efficacy	Sample contexts are relatively narrow, so external generalizability remains limited
[15]	Video-recorded assignment intervention in public speaking training	Repeated review enhances self-awareness and improves speaking confidence	More suitable for short-term training; long-term stability still requires follow-up
[17]	Video-enhanced multimodal feedback	Students can more easily identify detailed problems in speech, posture, and discourse organization	Feedback implementation is relatively costly and requires stronger teacher technical competence
[18]	Quasi-experimental study using presentation-based activities to improve speaking performance	Presentation tasks have positive effects on oral fluency and expressive proficiency	Analysis of nonverbal communication behaviors is relatively insufficient
[19]	Collaborative presentation study from the perspective of co-regulation	Peer collaboration can enhance public speaking self-efficacy	Limited discussion of individual differences and long-term maintenance effects

In general, the research on public speaking teaching has shifted from simply emphasizing "practice more" to "practice with support". Instructional design, feedback mode and technology environment are not separate variables, but jointly shape students' expression performance. Because of this, this paper does not regard the public speaking course as a

general demonstration training, but define it as a capacity training system that can be recorded, analyzed and optimized by computer technology. The actual effect of promoting students' confidence and communication skills is further analyzed by multi-source data processing and statistical test in the following paper.

3 Research Methods

3.1 Data collection and sample composition

In this study, a total of 64 students, including 26 male students and 38 female students, are selected as the research sample of undergraduates who are taking English public speaking courses in a university. The samples are all from two natural classes of the same grade. In order to ensure the comparability before and after the teaching intervention, the pre-test was implemented uniformly before the beginning of the course, including the public speaking confidence scale, the oral expression test in class and a short speech task with limited topics. After the course, the post-test was organized with the same structure and supplemented with the final public speaking composite grade. The whole data collection period was 12 weeks, covering multiple teaching sessions such as import training, keynote speech, improvisation, peer evaluation and summary presentation. In terms of data types, four types of information were collected simultaneously. The first was questionnaire data, which was used to describe students' self-efficacy, anxiety control and expression willingness in public expression. The second is the performance data, including pre-test speech scores, stage task scores and course test scores. The third is speech data, which extracts speech rate, pause frequency, average sentence length and filler word usage through recording. The fourth is video data, which is used to record gaze distribution, gesture amplitude, body sway and stance stability. All speech audio and video materials were imported into the analysis environment after unified slicing and numbering, and basic statistics, feature sorting and sample matching were completed under Python 3.10. In this way, the performance differences of students before, during and after teaching can be continuously tracked, which also provides a more stable data basis for subsequent paired sample t-test and multimodal evaluation.

3.2 Multi-source data preprocessing and standardization

Due to the simultaneous use of four types of data such as questionnaire, score, voice and video in this study, there are obvious differences in dimensions, sampling frequency, recording format and noise level among different sources. If directly entered into the statistical test, it is easy to amplify accidental fluctuations and weaken the comparability of results. Therefore, before the analysis of teaching effect, it is necessary to complete multi-source data preprocessing and standardization, so that students' performance at different time points can be uniformly mapped into the same computing framework.

The preprocessing stage mainly includes four steps: missing check, anomaly identification, time alignment and feature cleaning. The questionnaire and score data were verified by number matching method, and a small number of missing items were interpolated by the mean value of the same dimension. For speech data, noise reduction, silence segment removal and sentence segmentation were completed, and then the speech speed, average pause duration, filler word frequency and intonation fluctuation were extracted. After uniform frame rate, the video data are structured to encode gaze pauses, gesture changes, body swing amplitude, and stance stability. In order to avoid the misalignment caused by different collection periods, this study established a timestamp index centered on each formal speech task, and uniformly corresponding the questionnaire, audio, video and score records to the

same round of tasks of the same student, so as to form a horizontally comparable data sample.

In terms of standardization processing, considering the large differences in the dimensions of each index, this study used the range standardization method to linearly compress the original values. It is calculated as:

$$x_{ij}^* = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (1)$$

where x_{ij} represents the original value of the i th student on the JTH index, x_{ij}^* is the normalized result, and $\min(x_j)$ represents the maximum and minimum value of the index in all samples, respectively. After transformation, each index is uniformly mapped to the interval $[0,1]$, which is conducive to the subsequent comprehensive comparison.

For the negative indicators such as pause frequency, body shaking amplitude, and proportion of filler words, reverse processing is also needed to ensure that the indicators are in the same direction, and the calculation formula is:

$$x'_{ij} = 1 - x_{ij}^* \quad (2)$$

After this transformation, all indicators satisfy the interpretation rule of "the larger the value, the better the performance". This processing not only facilitates the subsequent paired sample t-test, but also provides a unified input for multi-modal comprehensive evaluation. After batch cleaning, normalization and data splicing were completed with the help of Python environment, the originally scattered classroom performance information was sorted into an analysis matrix with clear structure and consistent scale, which laid a solid data foundation for identifying the real impact of public speaking teaching on students' confidence and communication ability.

3.3 Paired sample t-test

In order to identify whether there are significant changes in students' self-confidence and communication ability before and after public speaking teaching, this paper uses paired sample t-test to statistically compare the pre-test and post-test results of the same group of students. This method is suitable for the research situation of "the same object, two measurements", and can focus on the individual difference rather than the discrete difference between samples, so it is more suitable for the design of teaching intervention in this study. Different from descriptive statistics that only compare the mean size, paired sample t-test can further determine whether the rise of indicators after teaching exceeds the range of random fluctuation, thus providing more explanatory evidence for the effectiveness of teaching.

In the specific calculation, the posttest value of each student is subtracted from the pretest value to obtain the difference sequence d_i , and then the significance test is performed on the mean value of the difference. The statistic is calculated as:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \quad (3)$$

Here, \bar{d} represents the sample mean of the paired difference, s_d represents the standard deviation of the difference, and n is the sample size. A large value of $|t|$ with $p < 0.05$ probability of significance indicates a statistically significant difference between the pre and post tests. In order to ensure the robustness of the test results, this paper calls pandas to complete the data processing in Python 3.10 environment, executes the paired test using

scipy.stats.ttest_rel(), and calculates the mean, standard deviation, and mean difference combined with numpy. The total score of confidence, expression fluency, non-verbal control score and comprehensive evaluation score were tested respectively. The significance level was uniformly set as 0.05, and the effect of teaching intervention was judged under 95% confidence interval.

Considering that the paired-sample t-test requires that the difference series approximately follow a normal distribution, the study checked the normality of the main index difference before the formal test. The results showed that there was no obvious deviation in the difference distribution of each core variable, which met the requirements of subsequent analysis. The significance of this treatment is that the effect of public speaking teaching is not only reflected in the outstanding individual performance of a few students, but should be reflected in the overall sample showing a stable positive trend after teaching.

Table 2 presents the paired-sample t-test results for the main indicators. Statistically, the self-confidence level, expression ability and comprehensive performance of the students after teaching were significantly improved compared with the pre-test, indicating that the course training not only improved the mental state of public expression, but also promoted the synchronous optimization of language organization and field control ability to a certain extent.

Table 2: Summary of paired sample t-test results

Test Item	Pre-test Mean ± SD	Post-test Mean ± SD	Mean Difference	t Value	p Value
Confidence Total Score	3.21 ± 0.48	3.86 ± 0.44	0.65	8.37	<0.001
Speech Fluency Score	72.45 ± 8.16	80.92 ± 7.38	8.47	7.14	<0.001
Nonverbal Control Score	68.37 ± 9.02	76.15 ± 8.41	7.78	5.96	<0.001
Overall Evaluation Score	70.84 ± 7.65	81.27 ± 6.93	10.43	8.92	<0.001

From the methodological point of view, the paired sample t-test plays a connecting role in this study: on the one hand, it tests whether the change before and after teaching is significant, on the other hand, it also provides a statistical basis for subsequent multimodal index analysis and comprehensive evaluation modeling. In other words, only if the current post-test difference is statistically valid, the expansion of the ability composition, diagnostic indicators and comprehensive scores in the following article will have a more solid analysis foundation.

3.4 Multi-modal assessment methods for confidence and expressiveness

The effect of public speaking teaching is not suitable to be judged by only one class score. Whether students really get improved is often reflected in three levels: subjective self-feeling, language output quality and on-site behavior control. If only teacher ratings are retained, it is easy to confuse "content preparation" with "stable presentation state", and it is difficult to identify the subtle difference between students' confidence growth and communication ability improvement. Based on this consideration, this paper constructs a multi-modal assessment method for confidence and expression ability, which integrates questionnaire results, speech signals, video behaviors and classroom performance into the same analysis framework, and completes feature extraction, fusion calculation and result output in a computer environment. Its overall structure is shown in Figure 1.

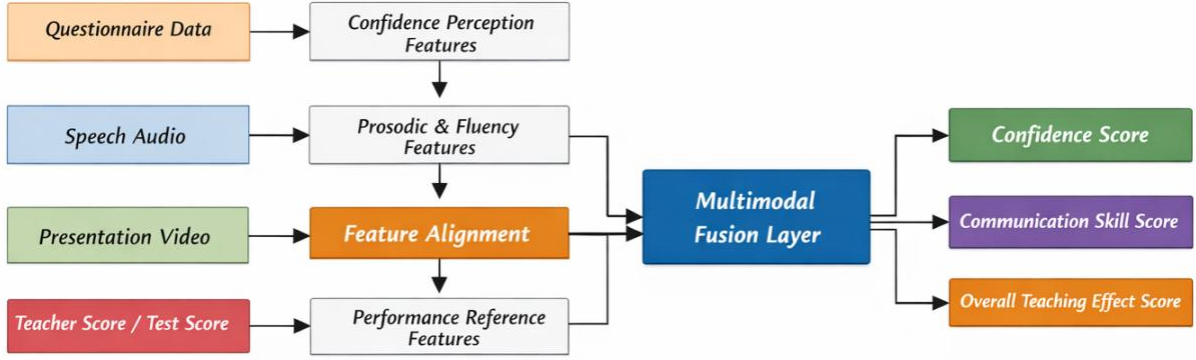


Figure 1: Structure diagram of multi-modal assessment of confidence and expressiveness

In the specific implementation, the questionnaire modality was mainly used to characterize students' self-efficacy, public expression of anxiety control and willingness to speak. Speech mode focuses on extracting speech rate, pause ratio, filler word frequency, loudness fluctuation and intonation change. The video modality is used to describe eye contact, gesture activity, body swing amplitude and stance stability. The performance modality assumes the role of calibration, which is used to test whether there is consistency between the model output and the actual performance of the course. The above data were uniformly numbered and entered into the Python analysis environment. The speech processing was completed by librosa for duration segmentation and acoustic feature extraction, and the video frame analysis was completed by OpenCV and pose keypoint recognition tools. The significance of this process is to transcribe the "self-confidence" and "expressivity" that were originally scattered in the classroom impressions into traceable and comparable digital evidence.

In order to ensure that different modalities can participate in the operation together, this paper constructs the voice feature vector A_i , video feature vector V_i , questionnaire feature vector Q_i and score reference vector R_i respectively after standardization. Let the fusion of the i th student be denoted F_i , then there is:

$$F_i = \alpha A_i + \beta V_i + \gamma Q_i + \delta R_i \quad (4)$$

Among them, $\alpha, \beta, \gamma, \delta$ represent the weight of each mode respectively, and satisfy $\alpha + \beta + \gamma + \delta = 1$. Considering that the explanatory power of each mode is not completely consistent in different teaching stages, this paper does not adopt fixed equal weight superposition, but dynamically assigns weights according to the correlation strength between each mode and the total performance. For speech and video behavioral data, this study pays special attention to their contributions to "explicit confidence" and "perceived expression quality". For example, the speaking rate of a student in a speech can be expressed as:

$$SR_i = \frac{N_i}{T_i} \quad (5)$$

where N_i is the number of effective words in a unit speech, T_i is the total length of the speech. The pause rate is then defined as:

$$PR_i = \frac{T_{\text{pause},i}}{T_i} \quad (6)$$

Here, $T_{\text{pause},i}$ denotes the total pause duration. Generally speaking, too low a speed is a sign of stagnation, too high a speed is a sign of stress, and too high a pause rate is often associated with tension and organizational instability. The proportion of eye contact in the video modality is calculated in a similar way:

$$GC_i = \frac{T_{\text{gaze},i}}{T_i} \tag{7}$$

Here, $T_{\text{gaze},i}$ is the cumulative time to gaze at the listener region. This index has a direct relationship with students' on-site engagement degree, self-stability and audience connection ability. After feature extraction, this paper further splits the fusion results into confidence dimension scores and expression dimension scores. The former emphasizes self-efficacy, pause control, gaze stability and body posture. The latter emphasizes content coherence, voice fluency, rhythm organization and interactive performance. The two types of results are then proportionately combined into the comprehensive evaluation score S_i :

$$S_i = \lambda C_i + (1 - \lambda)E_i \tag{8}$$

where C_i is the confidence score, E_i is the expressive ability score, and λ is the balance coefficient. This process can avoid "high achievement but unstable state" or "dare to speak but loose organization" being simply grouped into the same category, and also make subsequent teaching diagnosis more targeted. Its processing flow is shown in Figure 2.

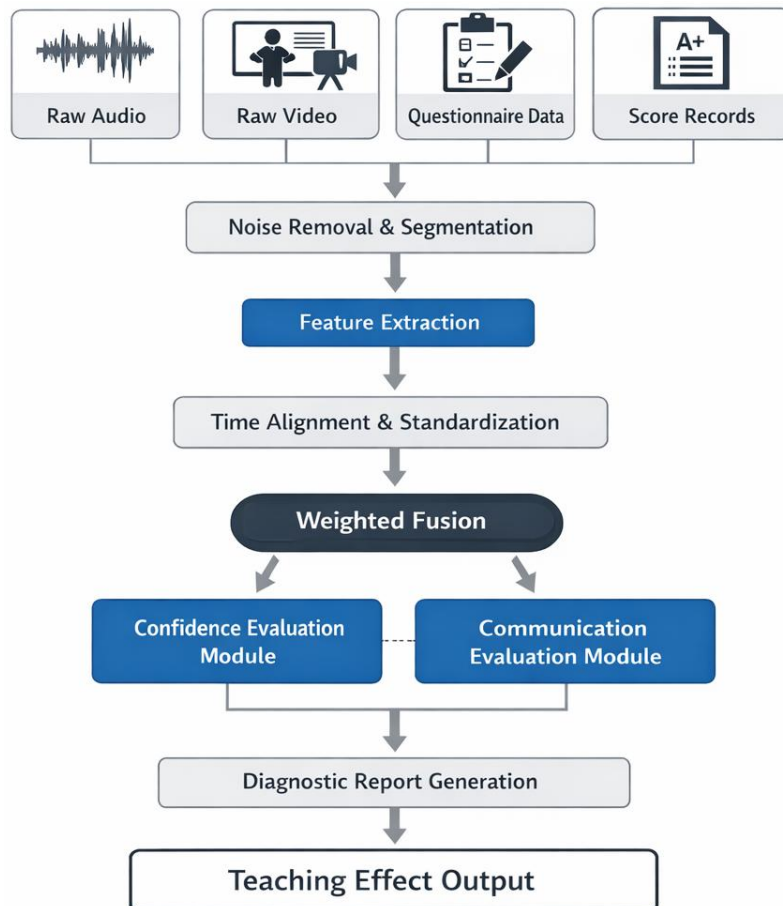


Figure 2: Flow chart of multimodal feature fusion and result output

Compared with the single scoring model, this multi-modal assessment approach is closer to the real scenario of public speaking teaching. It can not only retain the experience value in the teacher's judgment, but also make joint analysis of speech, behavior and psychological variables with the help of computer processing power, so as to improve the fineness of the evaluation. For this paper, the significance of this method is not only to generate a total score, but also to reveal whether students make more obvious progress in the self-confidence dimension, or more prominent improvement in the expression organization level, and thus provide a basis for subsequent teaching adjustment.

3.5 Construction of teaching effect evaluation index

After the multi-source feature extraction and fusion, the recognition of teaching effect still cannot stay on the parallel expression of several discrete indicators. The improvement of self-confidence and communication ability in the public speaking course belongs to a hierarchical comprehensive change in nature: some students' voice is more stable, but eye contact is still weak; Some students dare to speak up, but the improvement in structure and organization is limited. If the judgment is only based on the single score, it is easy to cause the evaluation distortion. Based on this, this paper constructs the teaching effect judgment index system on the basis of the standardized feature matrix, and converts the multimodal results into computable, scalable and interpretable teaching conclusions.

The index system is composed of three levels, namely, confidence dimension, expression dimension and comprehensive judgment dimension. Let the value of the i th student on the JTH standardized index be x_{ij} and the corresponding weight be w_j , then the confidence dimension score C_i can be expressed as:

$$C_i = \sum_{j=1}^m w_j x_{ij}, \quad \sum_{j=1}^m w_j = 1 \quad (9)$$

Among them, m represents the number of indicators belonging to the confidence dimension, which mainly includes self-efficacy, gaze stability, pause control rate and body posture stability. The expression dimension score E_i is calculated in the same way:

$$E_i = \sum_{k=1}^n v_k x_{ik}, \quad \sum_{k=1}^n v_k = 1 \quad (10)$$

where n represents the number of indicators related to expressive ability, and v_k is the corresponding weight, covering the appropriateness of speaking rate, linguistic coherence, filler word control, intonation fluctuation and content organization integrity. Considering that the teaching goal is not only to pursue "dare to speak", nor to emphasize "speak normatively", this paper further combines the two kinds of results into the comprehensive teaching effect score T_i :

$$T_i = \alpha C_i + \beta E_i, \quad \alpha + \beta = 1 \quad (11)$$

where, α and β represent the contribution coefficients of the confidence dimension and the expression dimension in the overall evaluation respectively. According to the course training objectives and the focus of teacher evaluation, this paper sets the two as similar weights, so as to avoid that one side of the index is too strong and the other side of the short board is covered.

In order to enhance the interpretation of the results, this paper does not directly take the continuous score as the final conclusion, but introduces a hierarchical decision rule. Let the set of teaching effect levels be $L = \{l_1, l_2, l_3, l_4\}$, corresponding to "significant improvement, relatively obvious improvement, slight improvement, and not obvious improvement" respectively. When the comprehensive score T_i satisfies different interval conditions, the level mapping can be completed:

$$L_i = \begin{cases} l_1, & T_i \geq 0.85 \\ l_2, & 0.70 \leq T_i < 0.85 \\ l_3, & 0.55 \leq T_i < 0.70 \\ l_4, & T_i < 0.55 \end{cases} \quad (12)$$

The advantage of this regularization processing is that it can convert the multimodal calculation results into a form of discourse closer to teaching decision-making, and facilitate teachers to identify whether students belong to "stable progress" or "partial improvement". At the same time, in order to test the consistency between the model judgment results and the teacher's rating, this paper also calculates the deviation value between the comprehensive score and the teacher's overall rating:

$$D_i = |T_i - S_i| \quad (13)$$

Among them, S_i represents the results after the standardization of teacher ratings, and a smaller D_i indicates that the output of the indicator system is closer to the judgment of classroom experience. Through this construction process, the information originally scattered in the questionnaire, voice, video and score is integrated into a clear structure of teaching effect judgment framework, so as to provide a unified basis for the analysis of the results in the following paper.

4 Results and discussion

4.1 Changes in students' self-confidence level

The change of students' confidence level is a direct incision to test whether the teaching of public speaking is effective. Combining the pre-test and post-test questionnaire data, and comparing the pause proportion, eye stability and body shaking amplitude in classroom audio and video with the results of the scale, it can be found that after 12 weeks of teaching intervention, students' subjective confidence and visibility in public expression have been significantly improved. The Python summary results showed that the total self-confidence score increased from 3.21 in the pre-test to 3.86 in the post-test, with an increase of 20.2%, indicating that most students had changed from "able to complete expression" to "relatively stable completion of expression". This change did not stay on a single item, but appeared simultaneously in multiple dimensions such as speaking intention, presence control and interaction ease.

According to the sub-item results, the most obvious improvement was in "improvising response confidence" and "expressing confidence to the audience". This suggests that time-limited presentation, peer review, and video review in the course not only improved students' psychological anticipation of the presentation task, but also enhanced their ability to maintain continuity of presentation under incomplete preparation conditions. In contrast, "Recovery confidence after sentence errors" has improved, but the increase is slightly lower, indicating that some students still have some instability in the face of verbal errors, short

pauses or disconnection of content. Combined with the results of video behavior recognition, it can be seen that although these students have reduced head bending avoidance and frequent swing in the post-test, their self-repair ability in high-pressure situations has not been fully consolidated.

Table 3: Results of pre-test and post-test of students' self-confidence level

Indicator	Pre-test Mean \pm SD	Post-test Mean \pm SD	Improvement Rate (%)
Confidence in Speaking to an Audience	3.18 \pm 0.56	3.91 \pm 0.47	23.0
Confidence in Impromptu Response	3.02 \pm 0.61	3.79 \pm 0.50	25.5
Confidence in Speech Process Control	3.27 \pm 0.49	3.88 \pm 0.45	18.7
Confidence in Recovering from Verbal Errors	3.11 \pm 0.54	3.63 \pm 0.48	16.7
Total Confidence Score	3.21 \pm 0.48	3.86 \pm 0.44	20.2

Table 3 presents the pre-test and post-test changes of confidence-related indicators. On the whole, the mean value of each index in the post-test was higher than that in the pre-test, and the difference in the total score was consistent with the results of the paired sample t-test above, indicating that public speaking teaching had a relatively clear role in promoting students' self-confidence. More importantly, this improvement was not just a subjective "feeling better" at the questionnaire level, but was mutually corroborated by objective changes in voice, posture, and classroom performance.

4.2 Composition analysis of expressive ability indicators

After completing the confidence level test, the further question that needs to be answered is: which components of expressive ability of the students are actually improved by the teaching of public speaking. According to the multi-modal assessment framework constructed in the previous section, the expressive ability is divided into five dimensions: content organization, language fluency, voice control, non-verbal expression and audience interaction, and the standardized speech, video and rating data are weighted and summarized in the Python environment. The results showed that the composition of students' expressive ability in the post-test stage was not evenly distributed, but showed clear hierarchical differences. According to the contribution rate of each dimension to the comprehensive expression score in the post-test stage, content organization accounted for 27.4%, and language fluency accounted for 24.1%, and the two totaled 51.5%, indicating that speech teaching had a more direct effect on "what to say" and "how to say coherently". The proportions of voice control, non-verbal expression and audience interaction were 18.6%, 16.8% and 13.1%, respectively, which were slightly lower than the previous two items, but steadily improved compared with the pre-test.

In terms of specific values, the content organization score increased from 72.8 to 81.6, and the language fluency score increased from 70.4 to 80.9, with an increase of 12.1% and 14.9% respectively. Voice control improved from 68.7 to 77.1, non-verbal expression improved from 66.9 to 75.8, and audience interaction improved from 64.3 to 74.2. Such results suggest that the promotion of course training to students is not limited to the text preparation level, but gradually extends to speech rhythm, posture management, and field connection ability. Especially after the intervention of video review and segmented feedback, the improvement of students in pause compression, gaze maintenance and gesture coordination was more obvious. At the same time, audience interaction is still the dimension with the lowest proportion, which means that although some students have been able to complete the personal statement

relatively completely, there is still room for improvement in question response, emotional contagion and on-site resonance establishment.

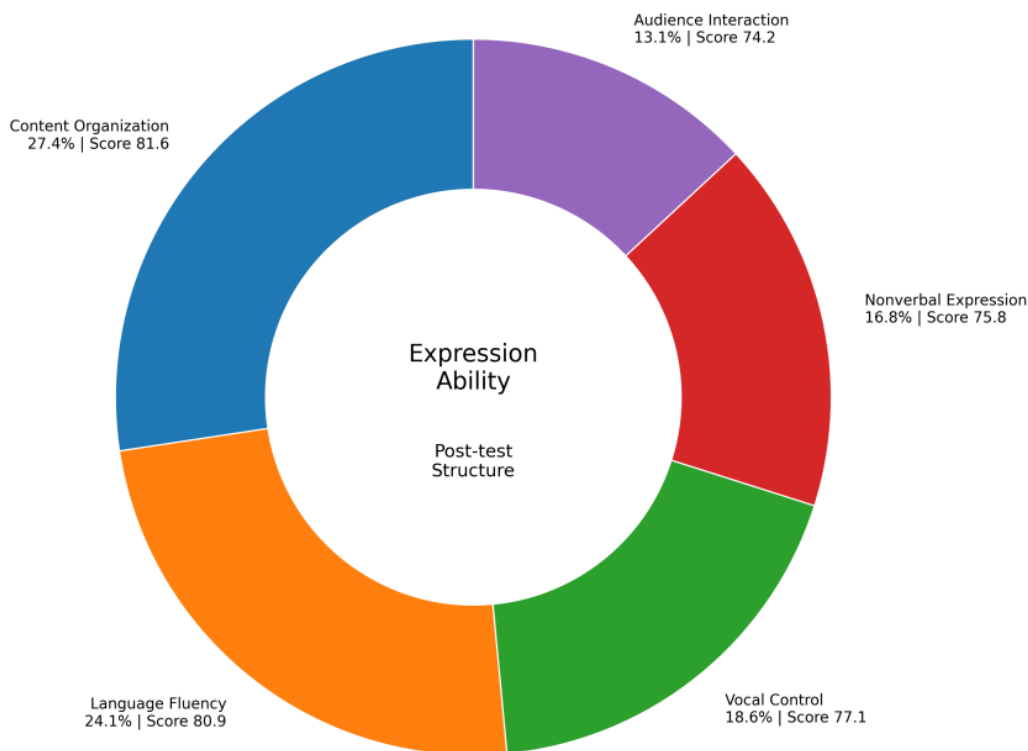


Figure 3: Composition of expressive ability indicators and distribution of post-test scores

As shown in Figure 3, the improvement of expression ability shows a change path of "structure first stability, expression then live, interaction after follow-up", which is basically consistent with the implementation logic of public speaking teaching from script training, oral output to on-site adjustment, and also provides a more specific explanation basis for the analysis of teaching diagnostic indicators in the following article.

4.3 Analysis of teaching diagnosis indicators

After completing the sub-item test of self-confidence and expression ability, the analysis of teaching diagnostic indicators is not only concerned with whether students' scores are improved, but also in which weak links teaching has produced more obvious correction effects. Based on the multimodal assessment results, this paper further summarizes the key features in the questionnaire, speech, video and classroom grading into five indicators: content organization diagnostic index, verbal fluency diagnostic index, non-verbal control diagnostic index, audience interaction diagnostic index and present adjustment diagnostic index, and completes the standardized calculation in the Python environment. The higher the index value, the more adequate problem identification in this dimension and the more obvious improvement after the teaching intervention.

The results showed that all the diagnostic indicators in the post-test stage were higher than those in the pre-test stage, and the greatest improvement was in the presence adjustment index and non-verbal control index. As shown in Figure 4, the on-the-spot adjustment index increased from 0.54 to 0.78, with an increase of 44.4%, indicating that students were better able to maintain the continuity of expression when they paused, forgot or made minor

mistakes after stage improvising statements, video review and peer feedback. The non-verbal control index increased from 0.57 to 0.77, indicating that the problems of body shaking, head bending avoidance and gesture imbalance were significantly alleviated. In contrast, although the audience interaction index increased from 0.52 to 0.68, it was still the lowest among the five items, which meant that students were relatively lagging behind in building audience response and strengthening the scene impact.

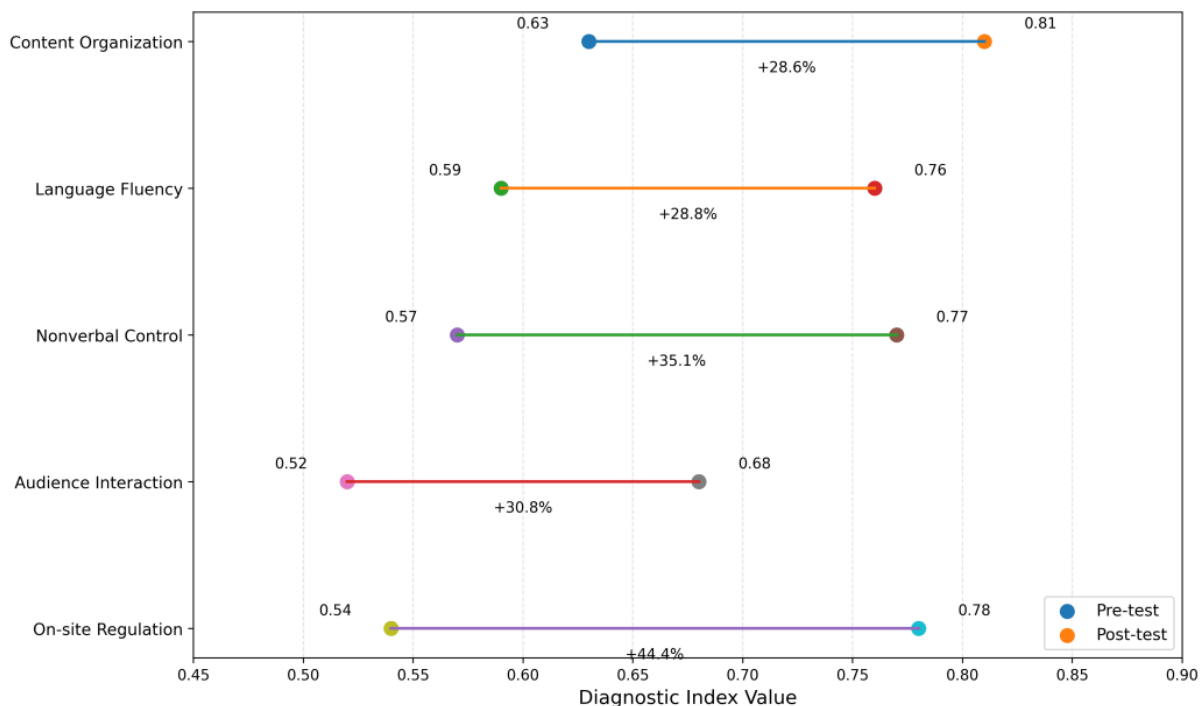


Figure 4: Change chart of teaching diagnosis index before and after test

Content organization increased from 0.63 to 0.81, while language fluency increased from 0.59 to 0.76. This indicates that public speaking teaching not only improves students' explicit state, but also enhances their ability to control speech structure and language cohesion. Combined with the classroom observation records, it can be found that most students have been able to clearly complete the introduction of the opening, the expansion of the key points and the end of the wrapping in the post-test stage, and the phenomenon of sentence interruption and repeated winding is significantly reduced. The overall rise of teaching diagnostic indicators indicates that the multi-modal evaluation system constructed in this study can more accurately capture the specific direction of student ability change, rather than just giving a single total score.

4.4 Analysis of comprehensive evaluation score

The comprehensive evaluation score is used to reflect the changes in the overall performance of students after the public speaking teaching intervention. Different from measuring the self-confidence scale or a certain type of expression index alone, the comprehensive score simultaneously incorporates multi-source information such as questionnaire results, speech fluency, video behavior performance and course test scores, which can more completely present the real progress of students in the speech task. In this paper, the weighted fusion of multi-modal features is completed in the Python environment, and the output results are uniformly mapped into hundred-point scores. The results showed that after 12 weeks of teaching training, the average of students' comprehensive evaluation increased from 70.84

points in the pre-test to 81.27 points in the post-test, with an average increase of 10.43 points and an increase of 14.7%. Combined with the results of paired sample t-test above, the difference reached a significant level, indicating that public speaking teaching had a stable role in promoting students' overall ability.

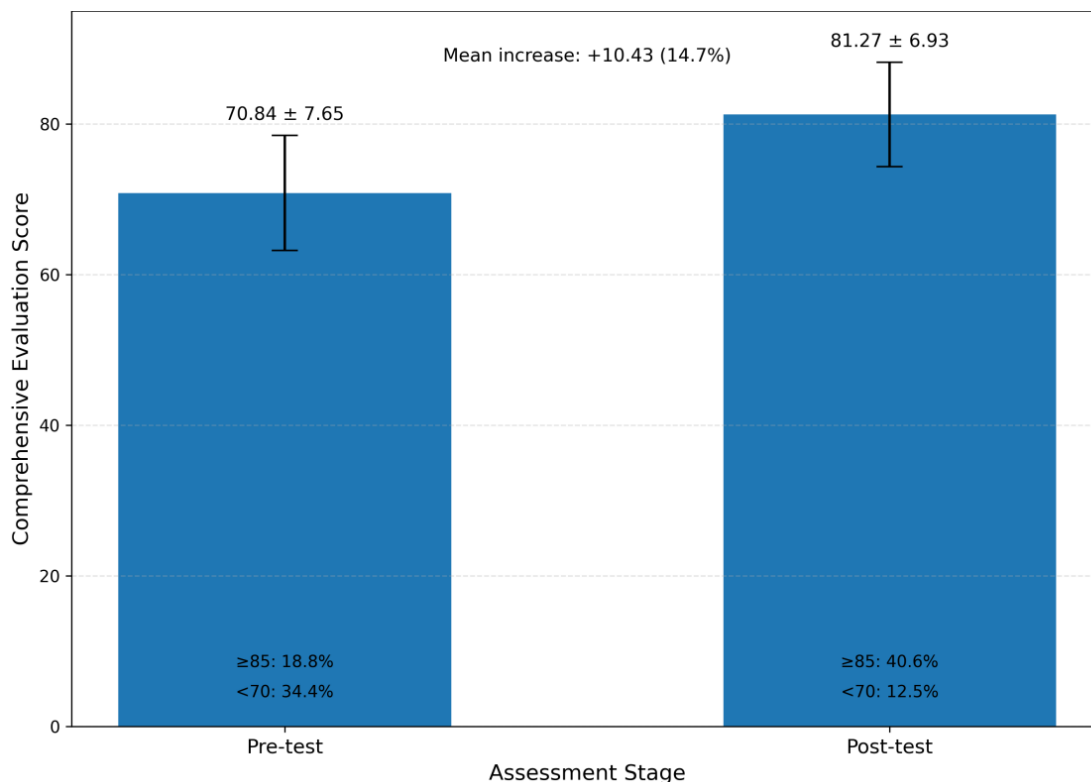


Figure 5: Comparison of comprehensive evaluation scores before and after test

From the score distribution, the number of high sections increased significantly after the teaching intervention. As shown in Figure 5, the proportion of students with scores of 85 or above in the posttest increased from 18.8% to 40.6%, and the proportion of students with scores of 70 or below decreased from 34.4% to 12.5%. This change shows that the teaching effect is not only reflected in a few high-level students, but the overall performance structure moves to the middle and high score interval. At the same time, the standard deviation of the post-test decreased from 7.65 to 6.93, indicating that the performance dispersion between students had converged, and some students who were in the low level gradually approached the course requirements after training, and their comprehensive performance became more stable.

4.5 Analysis of students' public speaking test scores

Students' public speaking test scores can directly reflect whether the teaching intervention is really transformed into explicit performance. Different from the comprehensive evaluation scores in the previous section, this section analyzes the unified test results of the course, and the scoring dimensions include five aspects: structure organization, language fluency, voice performance, non-verbal control and on-site response. After the test, the teacher completed the main scoring, and the computer side extracted the speech rate, pause rate, eye contact length and body stability of the recording and video materials synchronously, which were used to cross-check the scoring results. The results showed that the total average score of the public

speaking test increased from 72.64 in the first diagnostic test of the course to 84.38 in the final formal test, with an average increase of 11.74 points (16.2%), indicating that the teaching training had been significantly translated into performance improvement in real speech tasks.

In terms of sub-scores, the most obvious improvement was in structure organization and language fluency, the former from 74.15 to 86.72, and the latter from 71.83 to 84.91. This indicates that after several rounds of thesis statement, outline revision and video review, students have been able to complete the introduction of the opening, the expansion of the main points and the end of the wrapping more stably, and the repetition of sentences and invalid pauses have also been reduced. The speech performance was improved from 73.26 to 83.57, and the non-verbal control was improved from 70.48 to 82.06, indicating that the reading demonstration, rhythm training and video feedback in the course played a continuous role in correcting voice control and body management. In contrast, although the score of live response increased from 68.94 to 79.85, it was still a relatively low score among all items, reflecting that some students still have further room for improvement in improvising response and audience interaction.

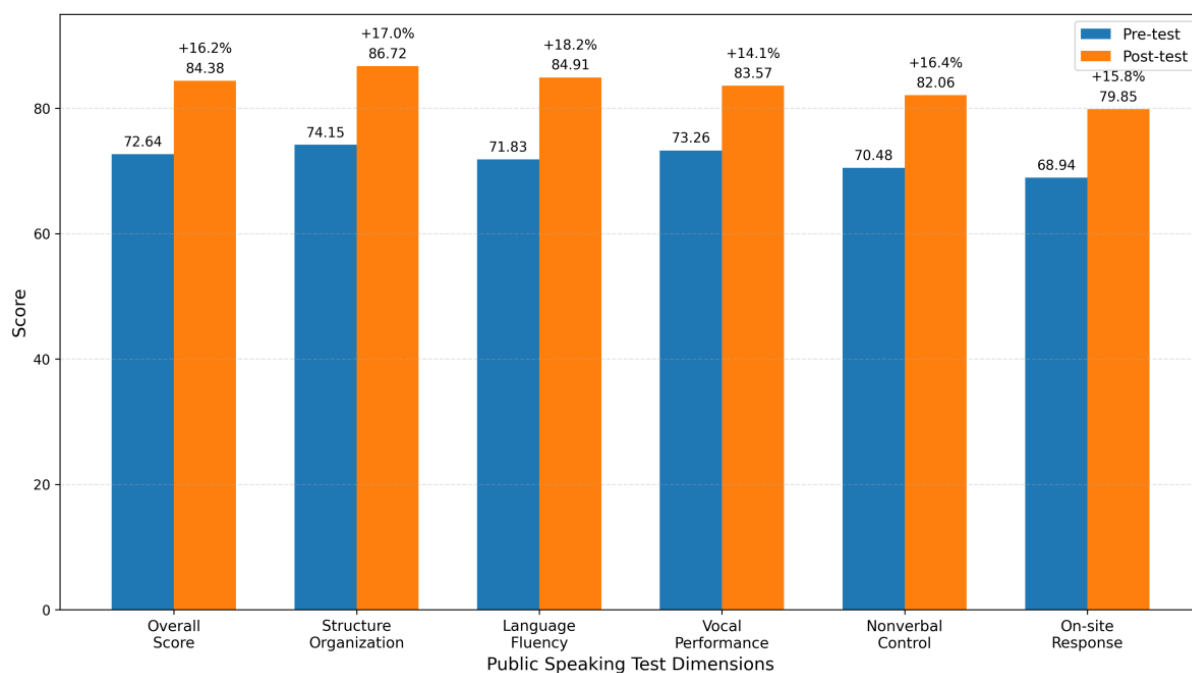


Figure 6: Comparison of students' public speaking test scores before and after test

As shown in Figure 6, the proportion of students with 80 points or above increased from 21.9% in the pre-test to 57.8% in the post-test, and the proportion of students with 70 points or below decreased from 31.3% to 9.4%. This means that performance improvement is not concentrated in a few high-performing students, but shows a shift in the overall distribution towards the middle and high score range. Combined with the above changes in self-confidence and diagnostic indicators, it can be seen that public speaking teaching not only improves students' subjective expression willingness, but also substantially improves their completion quality in formal tests.

4.6 Discussion

Combined with the above results, it can be seen that the promotion of public speaking teaching to students' ability does not come from the accidental rise of a certain index. More

precisely, it manifests as a synchronized improvement in mental state, verbal output, and field behavior. The total score of self-confidence increased from 3.21 to 3.86, the comprehensive evaluation score increased from 70.84 to 81.27, and the score of public speaking test increased from 72.64 to 84.38, which indicated that the teaching intervention had been further promoted from "reducing the fear of difficulty" to "optimizing the quality of expression completion". In other words, students are not only more willing to speak at the end of the course, but also more able to organize content, maintain pace, and complete effective presentation to the audience within the time limit. From the perspective of improvement path, content organization and language fluency are always in the high contribution range, while the improvement of audience interaction and live response is relatively lagging behind. This result indicates that classroom training promotes structured expression and controllable output more directly, because these two abilities are easier to continuously revise through outline design, segmented drill, video review and teacher feedback. In contrast, interactive competence relies more on immediate judgment, context awareness, and integration of audience feedback, and its formation is usually slower than that of scripted expressive competence. This also explains why audience interaction accounted for only 13.1%, and the live response score rose to 79.85, but it was still lower than the other items. In a methodological sense, multimodal assessment provides a more nuanced view of public speaking instruction than a single teacher rating. If we only look at the final grades, we often only know that students "speak better than before"; With the help of speech feature extraction, video action recognition and questionnaire structuring processing, it can be further determined whether the improvement comes more from the reduction of pauses, the stabilization of gaze, or the enhancement of language organization. The results of this paper show that computer supported analysis does not weaken the humanistic nature of teaching judgment, but complements and verifies classroom experience through data, so that "confidence improvement" and "communication ability enhancement" are no longer general descriptions, but can be explained more clearly. In this regard, the value of public speaking teaching lies not only in improving the performance of one course, but also in helping students gradually form a transferable and sustainable structure of expression ability.

4.7 Limitations

Although the results of this paper showed that the total self-confidence score of students increased from 3.21 to 3.86, the comprehensive evaluation score increased from 70.84 to 81.27, and the final public speaking test score increased from 72.64 to 84.38, this conclusion still needs to be understood within the boundaries of specific research. The research sample only included 64 students from two natural classes in the same university, and the sample sources were relatively concentrated. The teaching progress, course requirements and teachers' styles were consistent to a certain extent, which helped to control the interference variables, but also limited the interpretation strength of the results when extending to different regions, different majors and different types of speech courses. In other words, the results obtained in this paper are closer to the results of a clear course test, rather than a general conclusion applicable to all teaching scenarios of public speaking. At the same time, although this paper introduces multi-source data such as questionnaires, voice, video and scores, and uses Python to complete standardization processing and fusion calculation, the multi-modal analysis itself is still subject to technical conditions. Speech feature extraction is sensitive to the recording environment, and video action recognition may also be affected by shooting Angle, illumination change and occlusion, which means that some students' eye contact length, body swing range and other indicators may still have certain measurement errors. Especially in the real classroom environment, computer tools are more suitable for assisting analysis, but

it is difficult to completely replace the teacher's comprehensive judgment of context, emotion and expression intention. In addition, the observation period of this paper is 12 weeks, which is suitable for identifying short-term changes brought about by the course intervention, but not enough to show whether the improvement can be sustained and stable over a longer period of time. Abilities such as audience interaction and live response were improved in this study, but their post-test scores were still relatively low, suggesting that their development may require longer training periods and more complex real context support. Therefore, the results of this paper can clearly illustrate the effectiveness of public speaking teaching in phased training, but the long-term transfer effect and cross-scene adaptation ability still need to be verified by subsequent research.

5 Conclusion

This paper focuses on the analysis of the influence of public speaking teaching on students' confidence and communication ability. On the basis of course intervention, multi-source data such as questionnaires, voice, video and grades are introduced, and preprocessing, standardization, paired sample t-test and multimodal fusion evaluation are completed with Python to construct a more complete teaching effect judgment framework. The results showed that after 12 weeks of training, the total self-confidence score of students increased from 3.21 to 3.86, the comprehensive evaluation score increased from 70.84 to 81.27, and the public speaking test score increased from 72.64 to 84.38. It shows that public speaking teaching has obvious promoting effect on students' psychological stability, language organization ability and on-site expression control. Further analysis found that content organization and language fluency were more prominent, audience interaction and on-site response were improved, but they still needed to be continuously strengthened in subsequent teaching. It can be seen that embedding computer supported multimodal analysis into the public speaking course can not only improve the detail degree of teaching evaluation, but also help to show the change process of students' "dare to express" and "can express" more clearly.

Funding

This research was supported by the 2024 Jiangxi Province Higher Education Teaching Reform Research Provincial Project "Teaching Reform and Practice of Integrating Ideological and Political Education into English Public Speaking Course Based on the 'Three-into' of Xi Jinping: The Governance of China" (JXJG-24-24-5).

References

- [1] Belaman J A X, Zakaria S F B, Nasharudin S N S, et al. Balancing the use of oral communication strategies and fear for oral presentation: The case for language learning[J]. *International Journal of Academic Research in Business and Social Sciences*, 2022, 12(6): 849-868.
- [2] Lee Y J, Liu Y T. Promoting oral presentation skills through drama-based tasks with an authentic audience: A longitudinal study[J]. *The Asia-Pacific Education Researcher*, 2022, 31(3): 253-267.
- [3] Pradhananga P, ElZomor M, Santi Kasabdji G. Advancing minority STEM students'

- communication and presentation skills through cocurricular training activities[J]. *Journal of Civil Engineering Education*, 2022, 148(2): 04022001.
- [4] Rodero E, Larrea O. Virtual Reality with Distractors to Overcome Public Speaking Anxiety in University Students[J]. *Comunicar: Media Education Research Journal*, 2022, 30(72): 85-96.
- [5] Sarpourian F, Samad-Soltani T, Moulaei K, et al. The effect of virtual reality therapy and counseling on students' public speaking anxiety[J]. *Health Science Reports*, 2022, 5(5): e816.
- [6] Algouzi S, Alzubi A A F, Nazim M. Strengthening English language undergraduates' presentation skills: A blackboard-mediated intervention program[J]. *Plos one*, 2023, 18(8): e0289936.
- [7] Remacle A, Bouchard S, Morsomme D. Can teaching simulations in a virtual classroom help trainee teachers to develop oral communication skills and self-efficacy? A randomized controlled trial[J]. *Computers & Education*, 2023, 200: 104808.
- [8] Bachmann M, Subramaniam A, Born J, et al. Virtual reality public speaking training: effectiveness and user technology acceptance[J]. *Frontiers in virtual reality*, 2023, 4: 1242544.
- [9] Fehlmann B, Mueller F D, Wang N, et al. Virtual reality gaze exposure treatment reduces state anxiety during public speaking in individuals with public speaking anxiety: a randomized controlled trial[J]. *Journal of Affective Disorders Reports*, 2023, 14: 100627.
- [10] Alshammari N A, Mugaddam A H. Oral Presentations as Assessment Criteria for EFL Learners at University Level: A Saudi Instructors' Perspective[J]. *Eurasian Journal of Applied Linguistics*, 2023, 9(2): 132-141.
- [11] Docena V T, Buenaventura K H, Lobingco W, et al. Investigating oral presentation anxiety of freshmen English major students in a Philippine higher educational institution[J]. *Journal of Language and Pragmatics Studies*, 2023, 2(3): 173-189.
- [12] Zakaria S F, Rusli R, Mat N H C, et al. An insight to attitudes and challenges in oral presentations among university students[J]. *European Proceedings of Educational Sciences*, 2023.
- [13] Penrose R B. Increasing student self-efficacy: A case for assigning oral communication activities before written assignments[J]. *College Teaching*, 2024, 72(2): 98-105.
- [14] Akhir N M, Nasrudin N, Aziz M N. Perceptions on the role of YouTube in improving speaking skills among undergraduate students in Malaysia[J]. *Asian People Journal (APJ)*, 2024, 7(1): 154-163.
- [15] Ngadiran N M, Kadir H, Mizad M. Improving students' level of confidence in oral presentation via video recording task[J]. *Development in Language Studies*, 2024, 4(2): 11-14.

- [16] BINTI M D N N, BINTI M M, MD A N. ENHANCING STUDENTS'CONFIDENCE IN ORAL PRESENTATIONS WITH A VIDEO RECORDING ASSIGNMENT[J]. *International Journal of Academic Research in Business and Social Sciences*, 2024, 14(10).
- [17] Park J. Students' perceptions toward providing video-enhanced multimodal feedback on oral presentations[J]. *Computer Assisted Language Learning*, 2025, 38(8): 1934-1960.
- [18] Al-Khresheh M H. The role of presentation-based activities in enhancing speaking proficiency among Saudi EFL students: A quasi-experimental study[J]. *Acta Psychologica*, 2024, 243: 104159.
- [19] Hao X, Chen H. The effect of co-regulation on English public speaking self-efficacy in collaborative oral presentations[J]. *Frontiers in psychology*, 2024, 15: 1423607.
- [20] Wu H. The more, the better? A multivariate longitudinal study on L2 motivation and anxiety in EFL oral presentations[C]//*Frontiers in Education*. Frontiers Media SA, 2024, 9: 1394922.
- [21] Wanchid R, Charoensuk V. Effects of Feedback Types and English Achievement on EFL Oral Presentation Performance of Thai Undergraduates[J]. *PASAA: Journal of Language Teaching and Learning in Thailand*, 2024, 69: 153-194.
- [22] Quvanch Z, Qasemi A S, Na K S. Analyzing levels, factors and coping strategies of speaking anxiety among EFL undergraduates in Afghanistan[J]. *Cogent Education*, 2024, 11(1): 2413225.