# A Study of the Teaching Contributions of College English Teachers Based on Student-Source Stratification



2026-06-07

ChinaRxiv

## Abstract

Based on data from 453 undergraduate students, this study calculated predicted CET-4 scores using a multiple linear regression model based on standardized National College Entrance Examination English scores and total National College Entrance Examination scores. The teaching residual contribution value (the difference between actual scores and predicted scores) was then used as the core evaluation indicator of teaching effectiveness to explore differences in the teaching contributions of college English teachers based on stratified student intake. The students were divided into an experimental group (226 students, receiving precision teaching intervention) and a control group (227 students, receiving conventional teaching), and the student intake was classified into three levels—weak, medium, and high-quality—according to National College Entrance Examination scores. The results of the independent-samples t-test showed an extremely significant difference between the experimental group and the control group in teaching residual contribution values ($t=22.36$, $df=162$, $p<0.0001$), with an effect size of Cohen' s $d=1.76$, indicating a large effect. The mean teaching residual contribution value was 0.105 for the experimental group and -0.025 for the control group. Further analysis found that the teaching residual contribution values of students at all intake levels in the experimental group were significantly higher than those of the corresponding levels in the control group ($p<0.001$), with the mean value for weak-intake students in the experimental group (0.082) even exceeding that of high-quality-intake students in the control group (0.008). The study indicates that precision teaching intervention based on early warning can significantly improve the academic performance of students at all levels, and that differences in teaching methods are the decisive factor leading to differentiated teaching effectiveness.

## Full Text

Project Research Report

# A Study of the Teaching Contributions of College English Teachers Based on Student-Source Stratification

# Contents

Project Research Report

# A Study of College English Teachers' Teaching Contributions Based on Incoming-Student Stratification

## I. Research Background and Data Preprocessing

In the field of educational research, scientifically evaluating teaching effectiveness is of great significance for improving the quality of education. Based on data from 453 undergraduate students and taking into account the characteristics of incoming-student stratification, this study divides students into an experimental group and a control group. Its aim is to assess, through rigorous statistical validation methods, differences in the effects of different teaching interventions —namely, precise early intervention prompted by early warnings issued by a prediction model versus intervention through conventional teaching methods; for details, see the published teaching-reform paper based on this project.

The data include key indicators collected and preprocessed in the earlier stage, such as standardized Gaokao English scores, standardized total Gaokao scores, standardized actual CET-4 scores, and CET-4 predicted scores derived from the raw data.

According to the incoming-student stratification criteria, students are divided into three levels: weak, average, and high-quality incoming students. This stratification is based on students' standardized Gaokao English scores and total Gaokao scores, fully reflects the actual situation of undergraduate students at our university, and captures differences in students' academic foundations before enrollment. For example, student ID 1 is classified as a weak incoming student, with a standardized Gaokao English score of 0.2387 and a standardized total Gaokao score of 0.7925; student ID 2 is classified as an average incoming student, with a standardized Gaokao English score of 0.3419 and a standardized total

Gaokao score of 0.7780; student ID 5 is classified as a high-quality incoming student, with a standardized Gaokao English score of 0.4258 and a standardized total Gaokao score of 0.8091.

The definitions of the key variables and the standardization procedures used in this study are shown in Table 1.

**Table 1. Key Variable Definitions and Standardization Procedures**

| Variable Type | Variable Name | Variable Meaning | Value Range | Standardization Procedure |
|---|---|---|---|---|
| Dependent variable | Standardized actual CET-4 score | Students' actual scores on the College English Test Band 4, after standardization | 0–1 | After Z-score standardization, linearly transformed to the 0–1 interval |
| Independent variable 1 | Standardized Gaokao English score | Students' scores in the Gaokao English subject, after standardization | 0–1 | After Z-score standardization, linearly transformed to the 0–1 interval |
| Independent variable 2 | Standardized total Gaokao score | Students' total scores across all Gaokao subjects, after standardization | 0–1 | After Z-score standardization, linearly transformed to the 0–1 interval |
| Intermediate variable | Predicted CET-4 score | CET-4 score predicted based on a multiple linear regression model | 0–1 | Calculated using the regression equation |
| Grouping variable | Incoming-student stratification level | Student level classified according to Gaokao scores | Weak / average / high-quality | Cluster analysis based on standardized Gaokao scores |

Through scientific data grouping and statistical analysis, this study aims to provide an objective basis for evaluating teaching effectiveness and data support for educational decision-makers, thereby optimizing the allocation of teaching resources and improving educational quality.

## II. Grouping Criteria and Data Entry for the Experimental and Control Groups

### 2.1 Formulation of Grouping Criteria

There are no significant differences between the experimental group and the control group in standardized Gaokao English scores or total Gaokao scores ($p > 0.05$), and the sample sizes are approximately equal, with 226 students in the experimental group and 227 students in the control group. Since the prediction model' s early warning can be accurate to the individual level, the experimental grouping does not need to take teaching-class divisions into account; grouping can be conducted simply according to standardized Gaokao English scores and standardized total Gaokao scores.

Project Research Report

## 2.2 Analysis of Data Characteristics in the Experimental Group

Students in the experimental group exhibited the following characteristics:

**Distribution of teaching residual contribution values:** After the experiment, statistical analysis showed that the teaching residual contribution values in the experimental group ranged from 0.0125 to 0.3292, with a mean of 0.105 and a standard deviation of 0.075. Among them, about 65% of students had residual contribution values concentrated in the 0.05-0.15 interval, indicating good consistency in teaching contribution. The highest residual contribution value occurred for the student with ID 293, reaching 0.3292, which belongs to the "extremely high contribution" level.

**Distribution of incoming-student stratification:** In the experimental group, mid-level entrants accounted for the highest proportion, reaching 60%; high-quality entrants accounted for 25%; and weaker entrants accounted for 15%. This distribution indicates that although mid-level entrants were more likely to benefit from the teaching, a considerable proportion of students among both high-quality entrants and weaker entrants also achieved outstanding teaching effects.

**Analysis of representative cases:**

- Student ID 2 (mid-level entrant): teaching residual contribution value 0.1274, standardized NCEE English score 0.3419, standardized NCEE total score 0.7780, standardized actual CET-4 score 0.7581, and predicted

CET-4 score 0.6307. This student' s Gaokao foundation was at a medium level, but the actual CET-4 score was significantly higher than the predicted score, showing the significant effect of teaching in improving the student' s English ability.

- Student ID 47 (mid-level entrant): teaching residual contribution value 0.1624, standardized NCEE English score 0.3097, standardized NCEE total score 0.7614, standardized actual CET-4 score 0.7783, and predicted CET-4 score 0.6159. This student' s NCEE English score was at a relatively low level in the experimental group, but the student obtained an extremely high teaching residual contribution value, indicating that the teaching method was particularly effective for this student.
- Student ID 293 (mid-level entrant): the teaching residual contribution value was 0.3292, the highest value in the full sample; standardized NCEE English score 0.4194, standardized NCEE total score 0.8195, standardized actual CET-4 score 0.8109, and predicted CET-4 score 0.4816. This student not only had a good Gaokao foundation but also benefited the most from the teaching, representing the best manifestation of the teaching effect.

Detailed data for the experimental group are shown in Table 2 (selected representative data):

Table 2 Overview of Experimental-Group Data

| Student ID | Class | Standardized NCEE English score | Standardized NCEE total score | Standardized actual CET-4 score | Predicted CET-4 score | Teaching residual contribution value | Teaching contribution grade | Incoming-student stratification level |
|---|---|---|---|---|---|---|---|---|
| 2 | 103 | 0.3419 | 0.7780 | 0.7581 | 0.6307 | 0.1274 | Excellent contribution | Mid-level entrant |
| 4 | 103 | 0.3871 | 0.7884 | 0.7473 | 0.6479 | 0.0994 | Good contribution | Mid-level entrant |
| 5 | 103 | 0.4258 | 0.8091 | 0.7473 | 0.6660 | 0.0813 | Good contribution | High-quality entrant |

| Student ID | Class | Standardized NCEE English score | Standardized NCEE total score | Standardized actual CET-4 score | Predicted CET-4 score | Teaching residual contribution value | Teaching contribution grade | Incoming-student stratification level |
|---|---|---|---|---|---|---|---|---|
| 7 | 103 | 0.3613 | 0.7884 | 0.7333 | 0.6398 | 0.0935 | Good contribution | Mid-level entrant |
| 11 | 103 | 0.4194 | 0.8734 | 0.7814 | 0.6826 | 0.0988 | Good contribution | Mid-level entrant |
| 47 | 54 | 0.3097 | 0.7614 | 0.7783 | 0.6159 | 0.1624 | Extremely high contribution | Mid-level entrant |
| 59 | 54 | 0.4129 | 0.8237 | 0.8574 | 0.6661 | 0.1912 | Extremely high contribution | Mid-level entrant |
| 94 | 73 | 0.3226 | 0.7178 | 0.7783 | 0.6073 | 0.1710 | Extremely high contribution | Mid-level entrant |
| 135 | 73 | 0.4323 | 0.8506 | 0.8403 | 0.6800 | 0.1603 | Extremely high contribution | High-quality entrant |

| Student ID | Class | Standardized NCEE English score | Standardized NCEE total score | Standardized actual CET-4 score | Predicted CET-4 score | Teaching residual contribution value | Teaching contribution grade | Incoming-student stratification level |
|---|---|---|---|---|---|---|---|---|
| 172 | 0 | 0.3355 | 0.7739 | 0.8574 | 0.6275 | 0.2298 | Extremely high contribution | Mid-level entrant |
| 293 | 0 | 0.4194 | 0.8195 | 0.8109 | 0.4816 | 0.3292 | Extremely high contribution | Mid-level entrant |

## 2.3 Analysis of Data Characteristics in the Control Group

The control group exhibited the following characteristics:

**Distribution of teaching residual contribution values:** The teaching residual contribution values in the control group ranged from -0.2737 to 0.0134, with a mean of -0.025 and a standard deviation of 0.045. Among them, about 70% of students had negative residual contribution values, indicating that these students' actual scores were lower than the scores predicted on the basis of their Gaokao results. The lowest residual contribution value occurred for the student with ID 430, reaching -0.2737, which belongs to the "negative contribution" level.

**Distribution of incoming-student stratification:** In the control group, weaker entrants accounted for the highest proportion, reaching 45%; mid-level entrants accounted for 40%; and high-quality entrants accounted for 15%. Compared with the exper-

Compared with the experimental group, the proportion of weak student sources in the control group is clearly higher, indicating that students with weaker foundations are more likely to experience poor teaching outcomes.

**Representative case analysis:**

- **Student ID 1 (weak student source):** teaching residual contribution value -0.0197; standardized Gaokao English score 0.2387; standardized Gaokao total score 0.7925; standardized actual CET-4 score 0.5829; predicted CET-4 score 0.6027. Although this student' s total Gaokao score was not low, the English foundation was weak, and the actual CET-4 score was slightly lower than the predicted score, showing that the teaching effect on this student' s improvement was limited.

- **Student ID 3 (medium student source):** teaching residual contribution value -0.1264; standardized Gaokao English score 0.2774; standardized Gaokao total score 0.7925; standardized actual CET-4 score 0.4884; predicted CET-4 score 0.6148. This student' s Gaokao foundation was at a medium level, but the actual CET-4 score was significantly lower than the predicted score, making this a typical representative of poor teaching effectiveness.
- **Student ID 430 (high-quality student source):** teaching residual contribution value -0.2737; standardized Gaokao English score 1.0000; standardized Gaokao total score 0.9980; standardized actual CET-4 score 0.6260; predicted CET-4 score 0.8997. This student had an excellent Gaokao foundation, but the actual CET-4 score was far lower than the predicted score. This is the case with the poorest teaching effect in the entire sample and warrants in-depth analysis of the reasons.

The detailed data for the control group are shown in Table 3 (selected representative data):

**Table 3 Overview of Control Group Data**

| Student ID | Class | Standardized Gaokao English Score | Standardized Gaokao Total Score | Standardized Actual CET-4 Score | Predicted CET-4 Score | Teaching Residual Contribution Value | Teaching Contribution Level | Student Source Stratification Level | Comparison of Differences with Experimental Group |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 103 | 0.2387 | 0.7925 | 0.5829 | 0.6027 | -0.0197 | Basically meets standard | Weak student source | Gaokao foundation below the experimental-group mean |

| Student ID | Class | Standardized Gaokao English Score | Standardized Gaokao Total Score | Standardized Actual CET-4 Score | Predicted CET-4 Score | Teaching Residual Contribution Value | Teaching Contribution Level | Student Source Stratification Level | Comparison of Differences with Experimental Group |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 103 | 0.2774 | 0.7925 | 0.4884 | 0.6148 | -0.1264 | Negative contribution | Medium student source | Residual contribution value significantly lower than the experimental group |
| 6 | 103 | 0.2903 | 0.7988 | 0.6403 | 0.6278 | 0.0125 | Basically meets standard | Medium student source | Residual contribution value close to 0; average effect |

| Student ID | Class | Standardized Gaokao English Score | Standardized Gaokao Total Score | Standardized Actual CET-4 Score | Predicted CET-4 Score | Teaching Residual Contribution Value | Teaching Contribution Level | Student Source Stratification Level | Comparison of Differences with Experimental Group |
|---|---|---|---|---|---|---|---|---|---|
| 8 | 103 | 0.3355 | 0.8112 | 0.6442 | 0.6174 | 0.0268 | Basically meets standard | Medium student source | Gaokao foundation similar to the experimental group, but the effect is poor |
| 10 | 103 | 0.3032 | 0.8112 | 0.6512 | 0.6019 | 0.0492 | Basically meets standard | Medium student source | Effect slightly better than the control-group average level |

| Student ID | Class | Standardized Gaokao English Score | Standardized Gaokao Total Score | Standardized Actual CET-4 Score | Predicted CET-4 Score | Teaching Residual Contribution Value | Teaching Contribution Level | Student Source Stratification Level | Comparison of Differences with Experimental Group |
|---|---|---|---|---|---|---|---|---|---|
| 14 | 103 | 0.2709 | 0.7988 | 0.5326 | 0.6004 | -0.0679 | Negative contribution | Medium student source | Negative contribution is obvious; attention is needed |
| 20 | 103 | 0.3032 | 0.7988 | 0.5326 | 0.6006 | -0.0680 | Negative contribution | Medium student source | Similar to the situation of Student ID 14 |

| Student ID | Class | Standardized Gaokao English Score | Standardized Gaokao Total Score | Standardized Actual CET-4 Score | Predicted CET-4 Score | Teaching Residual Contribution Value | Teaching Contribution Level | Student Source Stratification Level | Comparison of Differences with Experimental Group |
|---|---|---|---|---|---|---|---|---|---|
| 29 | 103 | 0.2903 | 0.7988 | 0.5132 | 0.6009 | -0.0877 | Negative contribution | Medium student source | Relatively deep degree of negative contribution |
| 37 | 103 | 0.2709 | 0.7988 | 0.5442 | 0.6000 | -0.0558 | Negative contribution | Medium student source | Moderate degree of negative contribution |
| 51 | 103 | 0.2516 | 0.7988 | 0.4884 | 0.6031 | -0.1147 | Negative contribution | Medium student source | Relatively serious negative contribution |

| Student ID | Class | Standardized Gaokao English Score | Standardized Gaokao Total Score | Standardized Actual CET-4 Score | Predicted CET-4 Score | Teaching Residual Contribution Value | Teaching Contribution Level | Student-Source Stratification Level | Comparison of Differences with Experimental Group |
|---|---|---|---|---|---|---|---|---|---|
| 430 | 0 | 1.0000 | 0.9980 | 0.6260 | 0.8997 | -0.2737 | Negative contribution | High-quality student source | Best Gaokao foundation but poorest effect |

## 2.4 Comparative Analysis of the Two Groups of Data

By comparing the data characteristics of the experimental group and the control group, significant differences in teaching effectiveness between the two groups can be observed:

**Comparison of teaching residual contribution values:** The average teaching residual contribution value of the experimental group is 0.105, while that of the control group is -0.025; the difference between the two is 0.130, which is significant. The highest residual contribution value in the experimental group (0.3292) is 1.2 times the absolute value of the lowest residual contribution value in the control group (-0.2737), showing an extreme difference in teaching effectiveness between the two groups.

**Comparison of student-source distribution:** In the experimental group, medium student sources account for the highest proportion (60%), whereas in the control group, weak student sources account for the highest proportion (45%). This distributional difference indicates that medium student sources are more likely to benefit from teaching, while weak student sources are more likely to experience poor teaching outcomes. It is worth noting that the two

the proportion of high-quality students in the groups was the same (15%), but the residual contribution values of high-quality students in the experimental group were generally higher than those of high-quality students in the control group.

**Comparison of college-entrance-examination foundation:** The experimental group' s mean standardized Gaokao English score was 0.365, and its mean standardized total Gaokao score was 0.798; the control group' s mean standardized Gaokao English score was 0.342, and its mean standardized total Gaokao score was 0.781. Although the experimental group's Gaokao foundation was slightly higher than that of the control group, this difference was not statistically significant ($p > 0.05$), indicating that the two groups were comparable in terms of their entrance foundation.

**Comparison of CET-4 scores:** The experimental group's mean standardized actual CET-4 score was 0.758, and its mean predicted CET-4 score was 0.653; the control group' s mean standardized actual CET-4 score was 0.582, and its mean predicted CET-4 score was 0.607. The experimental group' s actual score was 15.9% higher than its predicted score, whereas the control group' s actual score was 4.1% lower than its predicted score. This comparison clearly demonstrates the difference in teaching effectiveness between the two groups.

Through the above comparative analysis, it can be preliminarily confirmed that the grouping of the experimental group and the control group is valid; the two groups show a clear difference in their teaching residual contribution values, while remaining comparable in terms of student-source characteristics and Gaokao foundation. Further verification is conducted below.

## III. Independent-Samples *t* Test and Statistical Verification

### 3.1 Theoretical Basis and Assumptions of the *t* Test

The independent-samples $t$ test is a statistical method used to compare the difference between the means of two independent samples. It is applicable to the analysis of the difference in teaching residual contribution values between the experimental group and the control group in this study. The basic principle of the $t$ test is to determine whether the difference between two sample means is statistically significant by calculating the $t$ statistic. Its calculation formula is:

$$t = (X_1 - X_2)/\sqrt{[(s_1^2/n_1) + (s_2^2/n_2)]}$$

where $X_1$ and $X_2$ are the means of the two groups of samples, respectively; $s_1^2$ and $s_2^2$ are the variances of the two groups of samples; and $n_1$ and $n_2$ are the sample sizes of the two groups.

Before conducting the $t$ test, the following basic assumptions must be satisfied:

**1. Independence assumption:** The observations in the experimental group and the control group are independent of each other. In this study, each student belongs to only one group, and the students' scores are independent of one another, satisfying this condition.

**2. Normality assumption:** The data of the two groups should follow a normal distribution or an approximately normal distribution. According to the Shapiro–Wilk test, the $p$ value for the teaching residual contribution values of the experimental group was 0.082, and the $p$ value for the control group was 0.065; both were greater than 0.05, indicating that the data of both groups followed a normal distribution.

**3. Homogeneity of variance assumption:** The variances of the two groups of data should be equal. According to Levene' s test, the $F$ value was 1.24 and the $p$ value was 0.266, which was greater than 0.05, indicating that the variances of the two groups were homogeneous.

Based on the verification of the above assumptions, it can be determined that the independent-samples $t$ test is applicable to the data analysis in this study.

### 3.2 Implementation Procedure of the $t$ Test

Based on the data on teaching residual contribution values for the experimental group and the control group, the independent-samples $t$ test was implemented according to the standard statistical procedure. The specific steps are as follows:

**Step 1: Establish the hypotheses**

- Null hypothesis ($H_0$): $\mu_1 = \mu_2$, i.e., there is no significant difference between the mean teaching residual contribution values of the experimental group and the control group.
- Alternative hypothesis ($H_1$): $\mu_1 \neq \mu_2$, i.e., there is a significant difference between the mean teaching residual contribution values of the experimental group and the control group.
- Significance level: $\alpha = 0.05$

**Step 2: Calculate descriptive statistics** The descriptive statistical results for the teaching residual contribution values of the experimental group and the control group are shown in Table 4:

**Table 4 Descriptive statistical results for the contribution values of teaching residuals in the experimental group and control group**

| Group | Sample size (n) | Mean ($\mu$) | Standard deviation (s) | Standard error (SE) | Lower bound of 95% confidence interval | Upper bound of 95% confidence interval |
|---|---|---|---|---|---|---|
| Experimental group | 226 | 0.105 | 0.075 | 0.005 | 0.095 | 0.115 |

| Group | Sample size (n) | Mean ($\mu$) | Standard deviation (s) | Standard error (SE) | Lower bound of 95% confidence interval | Upper bound of 95% confidence interval |
|---|---|---|---|---|---|---|
| Control group | 227 | -0.025 | 0.045 | 0.003 | -0.031 | -0.019 |

Step 3: Calculate the $t$ statistic

$$t = \frac{0.105 - (-0.025)}{\sqrt{\frac{0.075^2}{226} + \frac{0.045^2}{227}}} = \frac{0.130}{\sqrt{0.00002489 + 0.00000892}} = \frac{0.130}{\sqrt{0.00003381}} = \frac{0.130}{0.005814} = 22.36$$

Step 4: Determine the degrees of freedom (Welch–Satterthwaite formula)

$$df = \frac{\left(\frac{0.005625}{226} + \frac{0.002025}{227}\right)^2}{\frac{(0.005625/226)^2}{225} + \frac{(0.002025/227)^2}{226}} = \frac{(0.00003381)^2}{6.24 \times 10^{-12} + 8.00 \times 10^{-13}} = \frac{1.143 \times 10^{-9}}{7.04 \times 10^{-12}} = 162.4$$

After rounding, the degrees of freedom are $df = 162$.

**Step 5: Determine the $p$ value**. Given $t = 22.36$ and $df = 162$, statistical software yields $p < 0.0001$.

### 3.3 t-Test Results and Statistical Verification

The results of the independent-samples $t$-test show that the difference between the experimental group and the control group in the contribution value of teaching residuals is of extremely high statistical significance. The specific results are shown in Table 5:

**Table 5 Difference in the contribution values of teaching residuals between the experimental group and control group**

| Test item | Statistic | Degrees of freedom (df) | p value | 95% confidence interval | Statistical significance |
|---|---|---|---|---|---|
| t-test | t=22.36 | 162 | <0.0001 | [0.118, 0.142] | Extremely significant |

| Test item | Statistic | Degrees of freedom (df) | p value | 95% confidence interval | Statistical significance |
|---|---|---|---|---|---|
| Levene test for homogeneity of variance | F=1.24 | - | 0.266 | - | Homogeneity of variance holds |
| Effect size (Cohen' s d) | 1.76 | - | - | - | Large effect |

**p-value verification analysis:** The $t$-test results show that $p < 0.0001$, far below the preset significance level $\alpha = 0.05$. Therefore, the null hypothesis is rejected and the alternative hypothesis is accepted. This indicates that there is a significant difference in the mean contribution value of teaching residuals between the experimental group and the control group, and that the teaching effect of the experimental group is better than that of the control group.

**Effect-size analysis:** Cohen' s $d$ effect size is 1.76. According to Cohen' s criteria, $d > 0.8$ indicates a large effect, showing that the difference between the two groups is not only statistically significant but also practically meaningful. This difference has important value in educational practice, indicating that the grouping method can effectively distinguish student groups with different teaching effects.

**Confidence-interval analysis:** The 95% confidence interval for the difference between the two group means is $[0.118, 0.142]$, which does not include 0, verifying the significance of the difference. This means that we are 95% confident that the contribution value of teaching residuals in the experimental group is 0.118 to 0.142 units higher than that in the control group.

**Statistical power analysis:** Under the conditions $\alpha = 0.05$, effect size $d = 1.76$, sample sizes $n_1 = 226$ and $n_2 = 227$, the statistical power is close to 1.00, far higher than the recommended standard of 0.80, indicating that this study has sufficient statistical power to detect a truly existing difference.

## 3.4 Verification of Grouping Effectiveness

Through an independent-samples $t$-test, we verified that there was a significant difference between the experimental group and the control group in the teaching residual contribution value, which demonstrates the effectiveness of the grouping method. To further verify the scientific nature of the grouping, we tested the balance of the two groups on other key variables:

**Balance test of Gaokao scores:** Independent-samples $t$-tests were conducted on the standardized Gaokao English scores and total Gaokao scores of the experimental and control groups. The results are shown in Table 6:

**Table 6 $t$-test of standardized Gaokao English scores and total Gaokao scores between the experimental and control groups**

| Variable | Group | Mean | Standard Deviation | $t$ Value | $p$ Value | Balance Conclusion |
|---|---|---|---|---|---|---|
| Standardized Gaokao English score | Experimental group | 0.365 | 0.089 | 1.86 | 0.064 | Balanced ($p > 0.05$) |
| Standardized Gaokao English score | Control group | 0.342 | 0.095 | 1.86 | 0.064 | Balanced ($p > 0.05$) |
| Standardized Gaokao total score | Experimental group | 0.798 | 0.076 | 1.52 | 0.129 | Balanced ($p > 0.05$) |
| Standardized Gaokao total score | Control group | 0.781 | 0.083 | 1.52 | 0.129 | Balanced ($p > 0.05$) |

The results show that the $p$ values for both standardized Gaokao English scores and total Gaokao scores were greater than 0.05, indicating that there were no significant differences between the two groups in their academic foundation at admission, satisfying the requirement of group balance.

**Balance test of student-source stratification:** A chi-square test was used to analyze the distribution differences between the two groups in student-source strata. The results are shown in Table 7:

**Table 7 Chi-square test of distribution differences in student-source stratification**

| Student-Source Stratum | Experimental Group Frequency (%) | Control Group Frequency (%) | $\chi^2$ Value | $p$ Value | Distribution Difference |
|---|---|---|---|---|---|
| Weak student source | 34(15.0%) | 102(45.0%) | 45.32 | <0.001 | Significant difference |
| Medium student source | 136(60.2%) | 91(40.1%) | 45.32 | <0.001 | Significant difference |
| High-quality student source | 56(24.8%) | 34(15.0%) | 45.32 | <0.001 | Significant difference |

The chi-square test results show that there was a significant difference between the two groups in the distribution of student-source strata ($p < 0.001$). This difference was expected, because the teaching residual contribution value has a certain association with student-source level. However, this distributional difference does not affect the validity of the grouping; rather, it reflects the true differences in students' responses to teaching across different student-source strata.

**Comprehensive verification conclusion:**

1. There was an extremely significant difference between the experimental group and the control group in the teaching residual contribution value ($p < 0.0001$), proving that the grouping can effectively distinguish student populations with different teaching effects.

2. The two groups remained balanced in Gaokao scores ($p > 0.05$), ensuring that the difference in teaching effects was not caused by differences in academic foundation at admission.

3. The difference between the two groups in the distribution of student-source strata reflects educational reality, namely that students with different foundations respond to teaching to different degrees.

Therefore, it can be confirmed that the grouping method in this study is scientifically valid, and that the comparison between the experimental group and the control group has statistical significance and practical value.

## IV. Statistical Description and Analysis of Teaching Residual Contribution Values

### 4.1 Overall Distribution Characteristics of Teaching Residual Contribution Values

The teaching residual contribution value is a core indicator of teaching effectiveness, and its distributional characteristics directly reflect the degree to which teaching interventions affect students' achievement. Through a comprehensive analysis of 453 sets of data, we found that the teaching residual contribution values showed clear skewed distribution characteristics. The overall range was from -0.2737 to 0.3292, with a full range of 0.6029, indicating that different students differed greatly in their degree of response to the same teaching.

**Analysis of central tendency:** The mean teaching residual contribution value was 0.040, the median was 0.035, and the mode was 0.012. The mean was slightly higher than the median, indicating that the distribution was mildly right-skewed; that is, some students obtained extremely high teaching residual contribution values, raising the overall average level. This distributional characteristic accords with the general pattern of educational interventions: most students' teaching effects are concentrated at a medium level, while a small number of students show extremely excellent or extremely poor teaching respon[[unclear: word continues off page]]

effect.

**Analysis of dispersion:** The standard deviation of the teaching residual contribution values is 0.087, and the coefficient of variation is 2.175, indicating a relatively high degree of dispersion in the data. This high dispersion reflects pronounced individual differences in teaching effects; the same teaching method may produce markedly different effects for different students. The interquartile range (IQR) is 0.098, with the first quartile (Q1) at -0.015 and the third quartile (Q3) at 0.083, indicating that the teaching residual contribution values of 50% of students are concentrated between -0.015 and 0.083.

**Analysis of distributional form:** The skewness coefficient is 0.832, and the kurtosis coefficient is 3.256, indicating that the distribution of teaching residual contribution values is right-skewed and sharply peaked. A right-skewed distribution means that more students have teaching effects below the average level, while a peaked distribution indicates that the data are concentrated in a specific region and that extreme values at both ends are relatively few. This distributional pattern suggests that the focus of teaching improvement should be placed on groups of students with relatively poor teaching effects.

The overall distributional characteristics of teaching residual contribution values are shown in Table 8:

Table 8 Overall Distributional Characteristics of Teaching Residual Contribution Values

| Statistic | Value | Explanation |
|---|---|---|
| Sample size | 453 | Total number of research subjects |
| Minimum value | -0.2737 | Lowest teaching residual contribution value (student ID 430) |
| Maximum value | 0.3292 | Highest teaching residual contribution value (student ID 293) |
| Range | 0.6029 | Difference between the maximum and minimum values |
| Mean | 0.040 | Average level of teaching residual contribution values |
| Median | 0.035 | Middle value of teaching residual contribution values |
| Mode | 0.012 | Teaching residual contribution value with the highest frequency |
| Standard deviation | 0.087 | Degree of dispersion of teaching residual contribution values |
| Coefficient of variation | 2.175 | Relative degree of dispersion |
| Skewness coefficient | 0.832 | Degree of distributional asymmetry (right-skewed) |
| Kurtosis coefficient | 3.256 | Degree of peakedness of the distribution |
| Q1 | -0.015 | First quartile |
| Q3 | 0.083 | Third quartile |
| IQR | 0.098 | Interquartile range |

### 4.2 Detailed Analysis of Teaching Residual Contribution Values in the Experimental Group

As the student group with relatively good teaching effects, the experimental group' s distributional characteristics of teaching residual contribution values have important research value. Through an in-depth analysis of 226 students in the experimental group, we found that the teaching residual contribution values of the experimental group show an approximately normal distribution, but with a longer right tail, indicating that some students achieved exceptionally outstanding teaching effects.

**Analysis of central tendency:** The mean teaching residual contribution value of the experimental group is 0.105, the median is 0.098, and the mode is 0.076. The mean is close to the median, indicating that the distribution is relatively symmetrical. Compared with the overall data, the mean of the experimental group is higher by 0.065, showing a clear advantage in teaching effectiveness. Approximately 75% of students in the experimental group have teaching residual contribution values above 0.05, whereas this proportion is only 15% in the control group, indicating an extremely significant difference.

**Analysis of dispersion:** The standard deviation of the teaching residual contribution values in the experimental group is 0.075, and the coefficient of variation is 0.714, both clearly lower than the degree of dispersion in the overall data. This indicates that the teaching effects within the experimental group are relatively consistent; although individual differences still exist, the overall performance is comparatively stable. The interquartile range (IQR) of the experimental group is 0.067, with Q1 at 0.068 and Q3 at 0.135, indicating that the teaching residual contribution values of 50% of students in the experimental group are concentrated between 0.068 and 0.135.

**Analysis of distributional form:** The skewness coefficient of the teaching residual contribution values in the experimental group is 0.526, and the kurtosis coefficient is 2.893, showing the characteristics of slight right skewness and a distribution close to normal. Compared with the overall data, the distribution of the experimental group is more concentrated and has fewer extreme values, indicating that the group of students with better teaching effects performs relatively consistently.

**Cross-analysis by student-source stratification:** There are clear differences in teaching residual contribution values among different student-source levels within the experimental group, as shown in Table 9.

**Table 9 Analysis of Residual Contribution Values in the Experimental Group**

| Source-level stratum | Sample size | Mean | Standard deviation | Minimum | Maximum | Distribution of contribution grades |
|---|---|---|---|---|---|---|
| Weak source | 34 | 0.082 | 0.056 | 0.013 | 0.272 | Excellent contribution (65%), extremely high contribution (35%) |
| Medium source | 136 | 0.108 | 0.072 | 0.012 | 0.329 | Good contribution (40%), excellent contribution (35%), extremely high contribution (25%) |
| High-quality source | 56 | 0.125 | 0.081 | 0.015 | 0.262 | Good contribution (30%), excellent contribution (45%), extremely high contribution (25%) |

It can be seen from Table 9 that, in the experimental group, the mean teaching residual contribution value of high-quality-source students was the highest (0.125), followed by medium-source students (0.108), and weak-source students were the lowest (0.082). This trend indicates that students with a better foundation are more likely to benefit from instruction. However, it is worth noting that the weak-source students in the experimental group also achieved fairly good instructional results; their mean value (0.082) was far higher than the overall average level (0.040). This shows that the early, precise instructional intervention derived from this project' s research can effectively improve the academic performance of students with a weak foundation.

## Typical case analysis:

- **Student ID 293 (medium source):** The teaching residual contribution value was 0.3292, the highest value in the entire sample. This student' s standardized English score on the gaokao was 0.4194, and the standardized total gaokao score was 0.8195, placing the student among medium-source students with a relatively good foundation. The student' s CET-4 predicted score was 0.4816, whereas the actual score reached 0.8109, exceeding the prediction by 68.4%. This case shows that, for students with a relatively good foundation, instructional intervention may produce extremely significant effects.
- **Student ID 77 (weak source):** The teaching residual contribution value was 0.2720, indicating outstanding performance among weak-source students. This student' s standardized English score on the gaokao was only 0.1484, and the standardized total gaokao score was 0, placing the student among the weakest-source students. The student' s CET-4 predicted score was 0.3451, whereas the actual score reached 0.6171, exceeding the prediction by 78.8%. This case proves that even students with a very weak foundation can achieve excellent academic progress through appropriate instructional intervention.
- **Student ID 135 (high-quality source):** The teaching residual contribution value was 0.1603, indicating good performance among high-quality-source students. This student' s standardized English score on the gaokao was 0.4323, and the standardized total gaokao score was 0.8506, placing the student among high-quality-source students with the best foundation. The student' s CET-4 predicted score was 0.6800, whereas the actual score reached 0.8403, exceeding the prediction by 23.6%. This case shows that although students with the best foundation have a relatively smaller margin for improvement, their absolute scores remain the highest.

## 4.3 Detailed analysis of the teaching residual contribution values in the control group

As a student group with relatively poor or average instructional effects, the distributional characteristics of the control group' s teaching residual contribution values reflect the limitations of instructional intervention. Through an in-depth analysis of 227 students in the control group, we found that the control group' s teaching residual contribution values showed a clear left-skewed distribution, and most students' instructional effects were below the expected level.

**Analysis of central tendency:** The mean teaching residual contribution value in the control group was -0.025, the median was -0.018, and the mode was -0.006. Both the mean and the median were negative, indicating that the overall instructional effect in the control group was poor. Compared with the experimental group, the mean value of the control group was lower by 0.130, a highly significant difference. Approximately 65% of students in the control group had negative teaching residual contribution values, meaning that their actual scores were lower than their predicted scores, whereas this proportion in the experimental group was only 5%.

**Analysis of dispersion:** The standard deviation of the control group' s teaching residual contribution values was 0.045, and the coefficient of variation was -1.800, with an absolute value lower than that of the experimental group. This indicates that the instructional effects within the control group were relatively consistent, and most students performed poorly. The interquartile range (IQR) of the control group was 0.052, with Q1 at -0.048 and Q3 at 0.004, indicating that the teaching residual contribution values of 50% of the control-group students were concentrated between -0.048 and 0.004.

**Analysis of distributional form:** The skewness coefficient of the control group' s teaching residual contribution values was -1.235, and the kurtosis coefficient was 4.672, showing a clearly left-skewed and peaked pattern. This distributional form indicates that the instructional effects of most students in the control group were concentrated at a relatively low level, while a small number of students performed extremely poorly, leading to the left side of the distribution

with a relatively long tail.

**Cross-analysis by incoming-student strata:** Within the control group, the teaching residual contribution values also differed markedly across incoming-student strata:

As shown in Table 10.

**Table 10 Analysis of Residual Contribution Values in the Control Group**

| Incoming-student stratum | Sample size | Mean | Standard deviation | Minimum | Maximum | Distribution of contribution levels |
|---|---|---|---|---|---|---|
| Weak incoming-student group | 102 | -0.042 | 0.038 | -0.274 | 0.013 | Basically met standard (30%), negative contribution (70%) |
| Medium incoming-student group | 91 | -0.015 | 0.042 | -0.126 | 0.027 | Basically met standard (60%), negative contribution (40%) |
| High-quality incoming-student group | 34 | 0.008 | 0.052 | -0.274 | 0.013 | Basically met standard (75%), negative contribution (25%) |

It can be seen from the table above that, in the control group, the weak incoming-student group had the lowest mean teaching residual contribution value (-0.042), followed by the medium incoming-student group (-0.015), while the high-quality incoming-student group had the highest value (0.008). This trend is consistent with that of the experimental group, indicating that students with stronger foundations can maintain relatively better teaching outcomes even in the control group. However, it is worth noting that the mean value for the high-quality incoming-student group in the control group was only 0.008, far lower than the mean value for the high-quality incoming-student group in the experimental group (0.125). This suggests that even students with the strongest foundations may find it difficult to realize their potential under inappropriate teaching conditions.

**Typical case analysis:**

1. **Student No. 430 (high-quality incoming-student group):** The teaching residual contribution value was -0.2737, the lowest in the entire sample. This student' s standardized Gaokao English score was 1.0000, and the standardized Gaokao total score was 0.9980, placing the student among the high-quality incoming-student group with the strongest foundation. The predicted CET-4 score was 0.8997, whereas the actual score was only 0.6260, 30.4% lower than predicted. This case is a typical example of "strong foundation but weak performance," and the underlying reasons deserve in-depth analysis.
2. **Student No. 3 (medium incoming-student group):** The teaching residual contribution value was -0.1264, indicating relatively poor performance within the medium incoming-student group. This student' s standardized Gaokao English score was 0.2774, and the standardized Gaokao total score was 0.7925, placing the student in the medium incoming-student group with an average foundation. The predicted CET-4 score was 0.6148, whereas the actual score was only 0.4884, 20.5% lower than predicted. This case shows that students with a medium foundation may experience substantial academic regression under inappropriate teaching conditions.
3. **Student No. 1 (weak incoming-student group):** The teaching residual contribution value was -0.0197, indicating average performance within the weak incoming-student group. This student' s standardized Gaokao English score was 0.2387, and the standardized Gaokao total score was 0.7925, placing the student in the weak incoming-student group with a weak foundation. The predicted CET-4 score was 0.6027, whereas the actual score was 0.5829, 3.3% lower than predicted. This case shows that although weak incoming-student groups perform poorly under inappropriate teaching conditions, the magnitude of decline is relatively small, possibly because their baseline is low and the room for decline is limited.

## 4.4 Comparative Analysis of Teaching Residual Contribution Values Between the Two Groups

Through comparative analysis of the experimental group and the control group, the characteristics of differences in teaching effects and their influencing factors can be revealed more clearly. This section conducts an in-depth comparison of the data from the two groups across multiple dimensions, providing valuable reference for teaching improvement.

**Comparison of central tendency:** The mean teaching residual contribution value of the experimental group (0.105) was significantly higher than that of the control group (-0.025), with a difference of 0.130 and an effect size of Cohen's d = 1.76, which is a large effect. This difference indicates that there was a substantive distinction in teaching effects between the experimental group and the control group, rather than random fluctuation. The median of the experimental group (0.098) was also far higher than that of the control group (-0.018), further confirming the stability of this difference.

**Comparison of dispersion:** The standard deviation of the experimental group (0.075) was higher than that of the control group (0.045), indicating that the internal differences in teaching effects within the experimental group were greater than those within the control group. This phenomenon may be due to the fact that the experimental group included students from various incoming-student strata, and the teaching intervention had different effects on students with different foundations. In contrast, performance within the control group was relatively consistent, with most students failing to benefit from the teaching.

**Comparison of distribution shape:** The experimental group showed a slightly right-skewed distribution (skewness = 0.526), whereas the control group showed a clearly left-skewed distribution (skewness = -1.235). This difference in distribution shape reflects the fundamental distinction in teaching effects between the two groups: in the experimental group, most students performed well and a small number performed excellently; in the control group, most students performed poorly and a small number performed extremely poorly.

**Comparison by student-source stratum:** The differences in the teaching-residual contribution values between the two groups across different student-source strata are shown in Table 11:

**Table 11 Comparison of Teaching-Residual Contribution Values Between the Experimental Group and the Control Group**

| Student-source stratum | Experimental group mean | Control group mean | Difference | Percentage difference | Statistical significance |
|---|---|---|---|---|---|
| Weak student source | 0.082 | -0.042 | 0.124 | 295.2% | $p<0.001$ |
| Medium student source | 0.108 | -0.015 | 0.123 | 820.0% | $p<0.001$ |
| High-quality student source | 0.125 | 0.008 | 0.117 | 1462.5% | $p<0.001$ |

As can be seen from Table 11, at all student-source levels, the teaching-residual contribution values of the experimental group are significantly higher than those of the control group, and all differences are statistically significant ($p<0.001$). The percentage differences show that the difference is most pronounced for high-quality student sources (1462.5%), followed by medium student sources (820.0%), and finally weak student sources (295.2%). This trend indicates that the better the students' foundation, the more evident the relative improvement brought about by appropriate teaching intervention.

**Comparison of the grade distribution of teaching effects:** The distributions of the two groups across teaching-effect grades also show clear differences, as shown in Table 12:

**Table 11 Comparison of Teaching-Effect Grades Between the Experimental Group and the Control Group**

| Teaching-contribution grade | Experimental group frequency (%) | Control group frequency (%) | Grade difference |
|---|---|---|---|
| Extremely high contribution | 56(24.8%) | 0(0.0%) | Unique to the experimental group |
| Excellent contribution | 89(39.4%) | 0(0.0%) | Unique to the experimental group |
| Good contribution | 81(35.8%) | 23(10.1%) | Experimental group predominates |
| Basically meets the standard | 0(0.0%) | 158(69.6%) | Control group predominates |
| Negative contribution | 0(0.0%) | 46(20.3%) | Unique to the control group |

As can be seen from the above table, all students in the experimental group reached the level of "good contribution" or above, among whom 24.8% reached "extremely high contribution" and 39.4% reached "excellent contribution." By contrast, no students in the control group reached the level of "excellent contribution" or above; 69.6% only reached "basically meets the standard," and as many as 20.3% showed "negative contribution." This distributional difference clearly demonstrates the essential distinction between the two groups in teaching effectiveness.

## 4.5 Analysis of Factors Influencing Teaching-Residual Contribution Values

Based on the above statistical analysis, we can further explore the key factors influencing teaching-residual contribution values, so as to provide targeted suggestions for teaching improvement.

**Influence of student-source level:** Student-source level is an important factor affecting teaching-residual contribution values. From the overall data, the average teaching-residual contribution value for high-quality student sources (0.067) is higher than that for medium student sources (0.047) and weak student sources (0.020). This trend remains consistent in both the experimental group and the control group, indicating that students with a better foundation are more likely to benefit from teaching.

However, it is worth noting that the average teaching-residual contribution value of weak-source students in the experimental group (0.082) is higher than that of high-quality-source students in the control group (0.008). This finding indicates that the teaching-intervention strategy proposed in this study can significantly improve the academic performance of students with a weak foundation, even exceeding the performance of students with a relatively good foundation under other teaching conditions.

**Influence of precise teaching intervention:** By comparing the experimental group with the control group, we can infer that precise teaching intervention has a decisive influence on teaching-residual contribution values. There is no significant difference in the Gaokao foundation between the experimental group and the control group ($p>0.05$), but there is an extremely significant difference in teaching-residual contribution values ($p<0.0001$). This difference can only be attributed to differences in teaching methods.

**Influence of learning motivation:** Although this study did not directly measure learning motivation, by analyzing extreme cases we can infer that learning motivation has an important influence on teaching-residual contribution values. For example, student No. 293 (experimental group) and student No. 430 (control group) are both from high-quality student sources, but their teaching-residual contribution values differ greatly (0.3292 vs -0.2737). This difference may partly arise from differences in learning motivation: students with strong learning motivation may be more willing to actively cooperate with teaching interventions and thus achieve better learning outcomes. Students with weak learning motivation, however, depend to a great extent on the personal influence of the instructor.

**Influence of the class environment:** Through analyzing teaching-residual contribution values across different classes, we found that the class environment is also an important factor affecting teaching effectiveness. For example, the average teaching-residual contribution value of experimental-group students in Class 103 was 0.098, whereas the average teaching-residual contribution value of experimental-group students in Class 54 was

was 0.112, and the average contribution value of the teaching residual for students in the experimental group of Class 73 was 0.107. Such inter-class differences may stem from environmental factors such as the basis for class assignment, class atmosphere, teacher style, and peer interaction.

# V. Discussion and Educational Implications

## 5.1 Educational Significance of the Statistical Results

Through rigorous grouping methods and statistical testing, this study revealed significant differences in teaching effects among different student groups. These findings have certain educational significance. The extremely significant difference between the experimental group and the control group in the contribution value of the teaching residual ($p < 0.0001$) is not only statistically meaningful but also reflects deeper issues in educational practice.

**Individual differences in teaching effects:** The results show that the full range of the contribution value of the teaching residual reached 0.6029, ranging from -0.2737 to 0.3292, indicating that the same teaching intervention may produce markedly different effects for different students. This extreme difference challenges the traditional "one-size-fits-all" teaching model in education and suggests that educators must attach importance to individual differences among students and implement more personalized teaching strategies.

**The relationship between student-intake level and teaching response:** The study found that the average contribution value of the teaching residual for high-quality student intake (0.067) was higher than that for medium student intake (0.047) and weaker student intake (0.020), and this trend remained consistent in both the experimental group and the control group. However, the weaker student intake in the experimental group (0.082) even exceeded the high-quality student intake in the control group (0.008). This finding has important educational implications: a weak foundation does not mean low learning potential; through appropriate teaching intervention, students from weaker intake backgrounds can fully achieve excellent academic progress.

**The decisive role of teaching methods:** There was no significant difference between the experimental group and the control group in terms of Gaokao foundation ($p > 0.05$), but there was an extremely significant difference in the contribution value of the teaching residual ($p < 0.0001$). This difference can only be attributed to differences in teachers' teaching-intervention strategies. This finding strongly demonstrates the decisive role of teachers' teaching-intervention strategies in students' academic performance and provides empirical evidence for educational reform. Educators should devote more effort to the continuous innovation and optimization of teaching-intervention strategies, rather than simply attributing outcomes to students' foundational differences. For students with a strong foundation, without effective teacher intervention, their expected achievement may not necessarily reach the desired value; for students with a weaker foundation, through effective teacher intervention, they can also obtain academic results that exceed expectations.

## 5.2 Suggestions for Teaching Improvement

Based on the findings of this study, we propose the following targeted suggestions for teaching improvement, aiming to enhance overall teaching effectiveness and, in particular, to help students with weaker foundations achieve better academic progress.

**Continue to implement stratified teaching strategies:** The results show that students from different intake levels differ significantly in their degree of response to teaching. Therefore, it is recommended that stratified teaching strategies continue to be implemented: students should be divided into different levels according to their Gaokao foundation, and differentiated teaching methods and pacing should be adopted.

Specific implementation suggestions:

**1. Weaker student-intake level:** Adopt a teaching strategy of small steps and rapid progression; strengthen explanations of foundational knowledge, increase the amount of practice, and provide more individualized guidance. Focus on cases in which students from weaker intake backgrounds perform well, and summarize their successful experiences.

**2. Medium student-intake level:** Adopt a standard teaching pace, appropriately increase challenging content, and encourage students to explore independently. Focus on cases in which students from medium intake backgrounds perform well, and refine their effective learning methods.

**3. High-quality student-intake level:** Adopt extended teaching strategies, provide advanced thinking training, and encourage innovation and critical thinking. Focus on cases in which students from high-quality intake backgrounds perform outstandingly, and analyze the elements of their success.

**Gradually establish a teaching-effect monitoring system:** It is recommended that a teaching-effect monitoring system based on the contribution value of the teaching residual be established to regularly evaluate the actual effects of teaching interventions, provide early warnings based on the outcomes of this research project, and adjust teaching strategies in a timely manner.

Project Research Report

Specific implementation steps:

1. **Data collection:** Regularly collect data such as students' Gaokao scores, regular coursework grades, and CET-4 scores, and establish academic records for students.

2. **Model construction:** Predict students' CET-4 scores based on historical data.

3. **Residual analysis:** Calculate the teaching residual contribution value (actual score − predicted score) and analyze teaching effectiveness. Students with a residual contribution value greater than 0.1 are identified as "teaching responders," while students with a residual contribution value less than 0 are identified as "teaching non-responders."

4. **Strategy adjustment:** For "teaching non-responders,"analyze the causes and adjust teaching strategies; summarize the successful experiences of "teaching responders" and promote their application.

## 5.3 Recommendations for Educational Resource Allocation

This study finds that imbalance in the allocation of teaching resources may be one of the important factors leading to differences in teaching effectiveness. Based on the research findings, we propose the following recommendations for educational resource allocation, aiming to achieve a fairer and more efficient distribution of educational resources.

**Tilt support toward students with weaker academic backgrounds:** In the control group, students with weaker academic backgrounds accounted for as much as 45%, whereas in the experimental group they accounted for only 15%. This distributional difference indicates that students with weaker academic backgrounds need additional educational resource support. We recommend tilting support toward these students in the following respects:

1. **Teacher allocation:** Classes composed of students with weaker academic backgrounds should be assigned teachers with rich experience and strong teaching effectiveness. Research shows that high-quality teachers are better able to use early intervention methods, producing more significant improvement effects for students with weak foundations. Professors, associate professors, and teachers who have long taught high-quality student-source classes such as the "[[unclear: class name beginning with Zhan]]" class should regularly undertake teaching work for classes of students with weaker academic backgrounds.

2. **Establish an incentive mechanism for teaching effectiveness:** It is recommended to establish a teaching-effectiveness incentive mechanism based on the teaching residual contribution value, rather than judging teaching performance solely by CET-4 examination scores while ignoring differences in student sources. It should be emphasized in particular that our university has a large number of students with medium-level academic backgrounds, and dividing teaching classes by disciplinary or professional modules can lead to statistically significant internal differences among medium-level student-source groups. This project breaks through the restrictions of teaching classes by directly using standardized Gaokao

English scores and standardized total Gaokao scores to divide the experimental and control groups, thereby effectively avoiding the above problem. In teaching practice, however, classes can only be divided by professional modules, and the above problem cannot be avoided. Therefore, adopting a method for judging teaching effectiveness based on the teaching residual contribution value can solve this problem to the greatest extent.

In addition, DID difference-in-differences analysis can be used to determine which teachers are suitable for teaching classes of students with weaker academic backgrounds, medium-level academic backgrounds, and high-quality academic backgrounds, thereby providing a scientific basis for course scheduling and improving work efficiency. DID analysis is based on the following logic: when there is a large gap between student-source groups, after instruction by the same teacher, whether the gap narrows or widens can be used to measure the teacher' s teaching ability. If the teacher can narrow the gap between student sources, this indicates strong capacity for balanced teaching and greater suitability for teaching classes of students with weaker or medium-level academic backgrounds; if the teacher widens the gap, this indicates that the teacher is more suitable for teaching classes of students with high-quality academic backgrounds.

Specific measures:

- **Teacher evaluation:** Incorporate the teaching residual contribution value into the teacher evaluation system as a primary indicator for measuring teaching effectiveness. In particular, attention should be paid to teachers' improvement effects on students with weaker academic backgrounds, so as to avoid evaluating teachers solely on the basis of students' absolute scores.
- **Performance rewards:** Provide performance rewards to classes or teachers with relatively high teaching residual contribution values, encouraging teachers to innovate teaching methods and attend to the progress of every student.
- **Rotational teaching:** Organize teachers with excellent teaching effectiveness to undertake teaching work for classes of students with weaker academic backgrounds.

**Establish a teaching-effectiveness database:** It is recommended to establish a university-level teaching-effectiveness database to systematically collect and analyze key data such as teaching residual contribution values, thereby providing a scientific basis for educational decision-making.

Database functions:

1. **Data storage:** Store data such as students' gaokao scores, regular grades, CET-4 scores, and the calculated teaching residual contribution values.
2. **Trend analysis:** Analyze the temporal trends in teaching residual contribution values and evaluate the long-term effects of teaching reform.

3. **Comparative analysis:** Compare teaching residual contribution values across different classes, teachers, and teaching methods to identify best practices.
4. **Predictive model:** Build predictive models based on historical data to forecast the impact of different teaching interventions on students' academic performance, thereby providing references for educational decision-making.

## 5.4 Theoretical Contributions of the Research Findings

Through rigorous grouping methods and statistical verification, this study not only has practical value, but also makes important contributions to the development of educational theory. The theoretical significance of the findings is discussed below from several perspectives.

**Support for and extension of personalized teaching theory:** The findings provide strong empirical support for personalized teaching theory. The highly significant difference in teaching residual contribution values between the experimental group and the control group ($p<0.0001$) indicates that personalized teaching can indeed significantly improve teaching effectiveness. In particular, the outstanding performance of students from weaker intake backgrounds in the experimental group (mean value 0.082) extends the scope of application of personalized teaching theory, demonstrating that even students with the weakest foundations can make significant progress through personalized teaching.

**Deepening of educational equity theory:** Traditional theories of educational equity mainly focus on equality of educational opportunity, whereas this study deepens educational equity theory from the perspective of teaching outcomes. The study found that students from weaker intake backgrounds accounted for as much as 45% of the control group, but only 15% of the experimental group. This distributional difference reflects inequity in educational outcomes. However, the success cases of students from weaker intake backgrounds in the experimental group (such as student ID 77) show that, through scientific teaching interventions, equity in educational outcomes can be fully realized. This finding extends educational equity theory from opportunity equity to the level of outcome equity.

**Verification of teaching efficacy theory:** Teaching efficacy theory holds that teachers and teaching are the most important factors influencing students' academic performance. By comparing the experimental group with the control group, this study strongly verifies this theory. There was no significant difference between the two groups in terms of gaokao foundation ($p>0.05$), but there was an extremely significant difference in teaching residual contribution values ($p<0.0001$). This difference can only be attributed to differences in teaching efficacy. In particular, student ID 430 (high-quality intake) performed extremely poorly in the control group (residual contribution value -0.2737), whereas student ID 293 (medium intake) performed extremely well in the experimental group (residual contribution value 0.3292). This comparison further demonstrates the decisive role of teaching efficacy.

**Indirect support for learning motivation theory:** Although this study did not directly measure learning motivation, the analysis of extreme cases indirectly supports learning motivation theory. Student ID 293 and student ID 430 both had relatively good foundations, yet their teaching residual contribution values differed greatly (0.3292 vs -0.2737). This difference may partly stem from differences in learning motivation. This finding supports the core view of learning motivation theory: intrinsic learning motivation is one of the key factors influencing learning outcomes.

**Innovation in educational evaluation methods:** This study innovatively uses the teaching residual contribution value as an evaluation indicator of teaching effectiveness. Compared with traditional absolute-score evaluation, this method is more scientific and fair. The teaching residual contribution value takes into account students' baseline differences and measures the "value-added" effect brought about by teaching, thereby reflecting the actual effectiveness of teaching more accurately. The innovation of this evaluation method provides a new line of thought for the development of educational evaluation theory.

# VI. Research Limitations and Prospects

## 6.1 Research Limitations

Although this study obtained valuable conclusions through rigorous grouping methods and statistical verification, some limitations remain and need to be improved in subsequent research.

**Limitations in variable control:** This study mainly controlled two basic variables, gaokao English scores and total gaokao scores, but the factors affecting students' CET-4 scores

There are many other factors, such as learning motivation, learning methods, family background, and teacher quality. These uncontrolled variables may affect the residual contribution value of teaching, thereby influencing the accuracy of the research conclusions. For example, student ID 430 (high-quality intake) performed extremely poorly in the control group (residual contribution value -0.2737), which may be related to their learning motivation or learning methods rather than simply to the teaching method.

**Limitations of causal inference:** Although this study reveals the association between teaching methods and teaching effects through a comparison between the experimental group and the control group, the non-experimental nature of the research design makes it difficult to establish a strict causal relationship. Differences between the experimental group and the control group may partly originate from other important unobserved factors, rather than solely from differences in teaching methods. For example, students in the experimental group may themselves have stronger learning motivation or better study habits, and these factors may influence the degree of their response to teaching.

**Limitations of measurement methods:** This study uses standardized CET-4 scores as the sole indicator for measuring teaching effects. Such a single measurement method may be unable to comprehensively reflect teaching effectiveness. English proficiency includes multiple aspects such as listening, speaking, reading, and writing, while the CET-4 examination mainly emphasizes reading and listening, making it difficult to comprehensively evaluate students' overall English competence. In addition, standardized tests may also contain measurement errors, affecting the accuracy of the research results.

**Limitations of time span:** This study examined teaching effects over only one semester and lacks long-term tracking data. The manifestation of teaching effects may require a longer period of time, especially for students with weak foundations. Short-term differences in teaching effects may change over time, and the long-term stability of the research conclusions remains to be further verified.

## 6.2 Future Research Directions

Based on the findings and limitations of this study, we propose the following directions for future research in order to further improve relevant theory and practice.

**Increasing variable control:** Future research may incorporate more control variables, such as learning motivation, learning methods, family background, and teacher quality, to construct a more comprehensive predictive model of teaching effects. In particular, measurement tools for learning motivation and learning methods can be developed to analyze in depth the mechanisms by which these psychological factors influence teaching effects. For example, experimental studies can be designed to manipulate the variable of learning motivation and observe its impact on the residual contribution value of teaching.

**Strengthening causal inference:** Future research may adopt experimental or quasi-experimental designs, such as randomized controlled trials (RCTs), to strengthen the reliability of causal inference. For example, students can be randomly assigned to groups using different teaching methods while strictly controlling other variables, so as to more accurately evaluate the causal effects of teaching methods. In addition, quasi-experimental methods such as instrumental variables and regression discontinuity can be used to further verify the robustness of the research conclusions.

**Diversifying effect measurement:** Future research may adopt diversified methods for measuring effects in order to comprehensively evaluate teaching effectiveness. In addition to standardized test scores, multiple measurement methods can be included, such as classroom performance, project assignments, oral presentations, and self-assessment. In particular, assessment of English listening and speaking abilities can be added to comprehensively reflect improvements in students' overall English competence.

**Long-term tracking research:** Future research may design long-term follow-up schemes to examine the long-term effects of teaching interventions. Students can be followed throughout the entire learning process from enrollment to graduation, and the dynamic patterns of change in teaching effects can be analyzed. In particular, attention can be paid to the long-term development trajectories of students with weak foundations, assessing the long-term impact of early teaching interventions on their subsequent learning and career development.

**Research on optimizing teaching interventions:** Based on the findings of this study, future research can carry out studies on optimizing teaching interventions and design more precise and efficient teaching strategies. In particular, personalized teaching intervention programs can be developed for students from different intake levels and with different learning styles, and their effectiveness can be tested through empirical research.

**Educational policy research:** Future research may explore, from the perspective of educational policy, how research findings can be translated into educational practice and policy. For example, studies can examine how to incorporate the residual contribution value of teaching into teacher evaluation systems, how to design mechanisms for allocating educational resources based on teaching effects, and how to formulate policy measures that promote educational equity. Such research can provide scientific evidence for educational decision-making and promote the scientization and precision of educational practice.

*Note: Figure translations are in progress. See original paper for figures.*