College Academic Self-efficacy Inventory from Moral Perspective (CASI-M): Validation of Moral Education Evaluation for Chinese BE Students

1st Jingjing Shi[1,a], 2nd Suthagar Narasuman[2,b]*, 3rd Wenxuan Ren[3,c], 4th Huichun Ning[4,d], 5th Jie Xiang[5,e]

[a]janejjshi@outlook.com; [b]suthagar@uitm.edu.my; [c]renwenxuan@zjou.edu.cn; [d]ninghuichun@163.com; [e]jennyxiang@126.com

[1]Institute of International Communication, Taizhou Vocational & Technical College, China; Faculty of Education Universiti Teknologi MARA Puncak Alam, Malaysia 0000-0002-5948-3591
[2]Faculty of Education Universiti Teknologi MARA Puncak Alam, Malaysia; 0000-0001-8202-4951
[3]Faculty of Education Universiti Teknologi MARA Puncak Alam, Malaysia
[4]Faculty of Education Universiti Teknologi MARA Puncak Alam, Malaysia
[5]Institute of International Communication, Taizhou Vocational & Technical College, China Faculty of Education Universiti Teknologi MARA Puncak Alam, Malaysia

Abstract. A bilingual 35-item College Academic Self-efficacy Inventory from a Moral Perspective (CASI-M) was created as a fundamental element of a moral evaluation system for Business English (BE). This inventory is an extension of CASI, and its reliability and validity were thoroughly analyzed to assess its scale quality. The findings indicate that the instrument is deemed suitable for application in future research endeavors. Moreover, the application of CASI-M was also conducted for BE students in a Chinese vocational college. The following ANCOVA analysis proves the validation of moral education with a tech-integrated teaching model as well as the utility of CASI-M.

Keywords: moral education, evaluation system, self-efficacy, CASI-M, ANCOVA

1   Introduction

The term "Kecheng Sizheng" (abbreviated as KCSZ) was initially introduced in 2014 and has evolved into a key focus of current teaching reforms in colleges and universities. To elaborate, it involves integrating moral development into the delivery of knowledge in professional courses [20] through creative methods, combining both "preaching" and "teaching" approaches [1]. In this study, the term is aligned with moral education for consistency.

This research centers on vocational college students during a crucial phase of skill and knowledge acquisition. Based on the training content and objectives of Business English (BE), the study examines moral education for Business English majors through three dimensions, as illustrated in Figure 1:
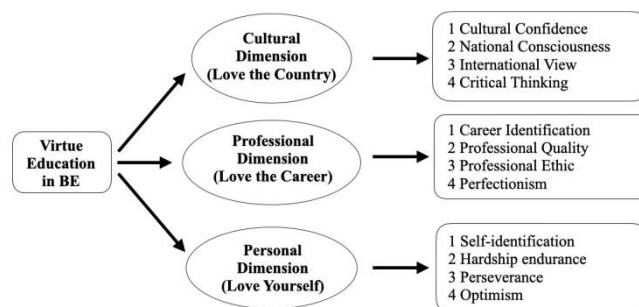
Fig. 1. 3 Dimensions of Moral Education in BE

Numerous investigations have identified a positive correlation between self-efficacy and the three dimensions of moral education in Business English (BE).

Regarding self-efficacy, cultural confidence, and critical thinking abilities, studies such as those conducted by Kim and colleagues[14] and Yada (2020) [21] revealed a positive correlation between self-efficacy and cultural confidence. Kim's study focused on international students in the United States[14], while Yada's study examined college students in Japan and Finland [21]. Additionally, insights from Bandura's team propose that cultural confidence could bolster self-efficacy by fostering a sense of belonging and connection to one's cultural heritage[2].

In the context of career identification, research by Nauta et al., (2002) [15] and Betz (2006) [3] demonstrated a positive association between self-efficacy and career identification in college students. Another study conducted by Yao's team (2021) [22] delved into the mediating effect of career identification on the ideological and moral values and self-efficacy of nursing students. The findings indicated positive correlations between ideological and moral values, career identification, and self-efficacy (r=0.179, r=0.698), as well as a positive correlation between ideological and moral values and professional identity (r=0.118). Career identification was identified as a partial intermediary, explaining 46.68% of the relationship between ideological and moral values and self-efficacy in nursing students. The researchers concluded that career identification enhances the impact of ideological and moral values education on undergraduate nursing students and has the potential to inspire their self-efficacy at an early stage.

As far back as 1999, Bandura's team and colleagues demonstrated a positive association between self-efficacy and self-identification in college students[2]. Correspondingly, a separate study conducted by Ferla and colleagues (2009) revealed that elevated levels of self-efficacy were linked to a clearer self-concept among young adults [7].

In summary, it can be deduced that the three dimensions of moral education in Business English are interconnected with academic self-efficacy, suggesting a reciprocal influence between personality factors and behavior, as described by Bandura. Figure 2 visually represents this interrelation.
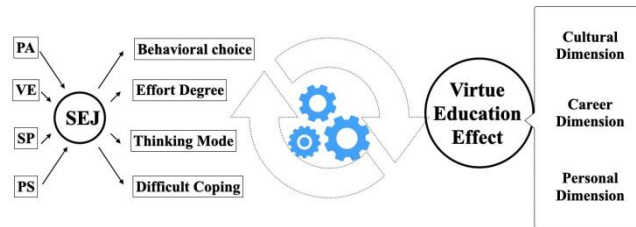
Fig. 2. Coupling Effect between SE and Moral Education

2 Method

Building upon the findings of the preceding study, the current research aims to construct a self-efficacy scale specifically tailored for moral education among students.

2.1 Instrument development

In this study, the CASI (College Academic Self-efficacy Inventory) developed by Owen & Froman (1988) was chosen as the foundational instrument [16]. The 33 items of the CASI were tested on a pilot sample of 40 students to assess reliability, with correlation coefficients measured across the entire range (0.26 – 0.67). All correlation coefficients demonstrated statistical significance. The reliability, evaluated through internal consistency using Cronbach's Alpha on an independent group of 40, was found to be 0.81. The scale's potential overall scores range from 33 points, representing the highest level of confidence, to 165 points, indicative of the lowest.

The adaptation process comprises four key steps: translation, expert review, integration of moral dimensions, and back-translation. To enhance comprehension among the local population (Chinese), the CASI must undergo translation (Hambleton et al., 2004; Schweizer et al., 2016)[10] [17]. This translation process is crucial to preserve the validity of the CASI, ensuring that the "meaning" of each item aligns with its original context (Geisser, 1975; Hambleton et al., 2004) [8] [10]. Simultaneously, objectivity is maintained in the adaptation process to prevent cultural biases (Canino & Bravo, 1999; Jones et al., 2001) [5] [13].

Subsequently, a panel of experts meticulously examined the translated scale, conducting a thorough review and screening process to eliminate items deemed unsuitable for the target student population.

The subsequent stage involves integrating the three dimensions of moral education for Business English into the refined scale. Following this integration, the version was subjected to back translation from Chinese into English, adhering to the process outlined by Brislin (1980)[4]. Back translation entails having a bilingual individual translate the scales from the target language (Chinese) back to the source language (English).

The concluding phase involves assessing the quality of the original instrument in English against the translated version, also in English [18]. The comparison between these instruments was carried out by evaluating the language used for each item. To mitigate potential bias and ensure the intended results, independent third parties were engaged to conduct the examination. As a result, the CASI has evolved into a bilingual 35-item College Academic Self-efficacy Inventory

from a Moral Perspective (CASI-M) designed for students. In order to avoid repetition, Table 4 provides a comprehensive list of all the items included in this inventory.

2.2   Sampling

The researcher selected a random sample comprising 158 students from Grade 1 and Grade 2 of the Business English major in a higher vocational college. The distribution of the samples is outlined in Table 1. To prevent any potential misunderstandings, the questionnaire was bilingual and distributed through "WenJuanXing," an online survey platform in China, ensuring efficiency. Data collection encompassed learners' demographic information, learning characteristics, academic self-efficacy scale, and moral awareness. Subsequently, the obtained data were analyzed to assess the validity and reliability of the CASI-V.

Table 1. Distribution of the Samples

Variables	Level	Frequency	Percentage
Gender	Female	135	85.4
	Male	23	14.6
	Total	158	100
Grade	Grade 1	81	51.3
	Grade 2	77	48.7
	Total	158	100

2.3   Reliability analysis

In the exploratory factor analysis (EFA) of social science research, the most commonly used scale is Likert scale, exactly what used in this research, while the most commonly used method for reliability is Cronbach's value [16]. The higher the reliability of a scale, the more stable it is. Cronbach's α value is between 0 and 1. The larger the value, the higher the reliability. Therefore, the Cronbach's α value should be at least greater than 0.5, and the best quantity in practice is $\alpha>0.7$ [17].

In this research, the general Cronbach's α value and the Cronbach's α value for each dimension of the CASI-V scale is shown in Table 2:

Table 2. Reliability Analysis

Dimensions	Cronbach's Alpha	N of Item
ASE	.956	9
Cultural	.949	7
Career	.969	10
Personal	.963	8
Total	.988	35

Table 2 indicates that the overall scale has high reliability, $\alpha=0.988>0.7$, and the reliability of each dimension is also satisfied.

2.4   Validity analysis

Validity analysis is employed to examine whether the questions effectively capture the conceptual information related to the research variables or dimensions. Essentially, it assesses whether the research questions are well-designed and accurately represent the intended variables.

Analyzing the validity of the scale is crucial for enhancing the questionnaire's quality and the overall value of the study.

In the present research, validity analysis involves three steps: 1, KMO coefficient, the value range is 0-1, the closer to 1, the better the structural validity of the questionnaire. 2, Bartley Sphere Test, if the significance of the test is less than 0.05, we can also consider that the questionnaire has good structural validity. 3, if the overall questionnaire is valid, it is still necessary to further evaluate the rationality of the question through a component matrix.

The results are shown in Table 3:

Table 3. KMO and Bartlett's Test

	KMO	0.961
Bartlett's Test of Sphericity	Approx. Chi-Square	8004.558
	df	595
	Sig.	0

The results of KMO test and Bartlett test in Table 3 indicate that KMO=0.961>0.5, the sample size is sufficient, and in the sphericity test, P=0.000<0.05, conforms to the sphericity test. Combining the two indicators, this example is suitable for factor analysis.

Table 4. Pearson Correlation and Corrected Correlation of the 35-item CASI-V

Dimension	Code	Major Content	Pearson Correlation	Corrected Correlation
ASE	ASE1	Lecture notes	.777**	0.763
	ASE2	Engagement	.842**	0.832
	ASE3	Assignment	.790**	0.778
	ASE4	Examination	.879**	0.872
	ASE5	Attendance	.483**	0.461
	ASE6	Content study	.875**	0.868
	ASE7	Textbook understanding	.910**	0.904
	ASE8	Class presentation understanding	.893**	0.887
	ASE9	Library usage	.897**	0.891
	ASE10	Competitions	.855**	0.844
Cultural	Ctl1	Events	.743**	0.728
	Ctl2	Smart study	.921**	0.916
	Ctl3	Reasoning study	.876**	0.868
	Ctl4	Cross-culture	.884**	0.875
	Ctl5	Chinese culture confidence	.825**	0.812
	Ctl6	Challenging information	.882**	0.874
	Ctl7	Concept of right or wrong	.819**	0.808
Career	Crr1	Training programs	.891**	0.884
	Crr2	Local industries	.850**	0.840
	Crr3	Volunteer activities	.830**	0.819
	Crr4	Company visiting	.803**	0.790
	Crr5	Professional communication	.830**	0.818
	Crr6	Professional accreditation	.878**	0.870
	Crr7	Job hunting	.877**	0.869
	Crr8	Practical opportunities	.911**	0.905

	Crr9	Career development	.859**	0.849
	Crr10	Career preparation	.890**	0.883
	Ps1	Class discussion	.867**	0.859
	Ps2	Question answering	.862**	0.853
	Ps3	Asking for repetition	.844**	0.832
Personal	Ps4	Peer assistance	.858**	0.848
	Ps5	Classroom activities	.880**	0.872
	Ps6	Teachers communication	.880**	0.871
	Ps7	Challenging teachers	.852**	0.841
	Ps8	Challenging difficult tasks	.852**	0.843

If the content validity of the scale is high, the correlation between the score of each question and the total score should be high, and the criterion is r>0.4. In this case, all the results of Pearson Correlation are higher than 0.4, as shown in Table 4, which means all the items can represent the content to be studied.

In order to further explore the relationship between different variables, this research also conducted a bivariate correlation analysis, and the results are shown in Table 5:

Table 5. Bivariate Correlation Matrix

		ASE	Cultural	Career	Personal	Gender	Grade
ASE	Pearson Correlation	1	.925**	.912**	.922**	.127	-.066
	Sig. (2-tailed)		.000	.000	.000	.111	.411
Cultural	Pearson Correlation	.925**	1	.935**	.910**	.169*	-.078
	Sig. (2-tailed)	.000		.000	.000	.034	.33
Career	Pearson Correlation	.912**	.935**	1	.910**	.123	-.096
	Sig. (2-tailed)	.000	.000		.000	.123	.228
Personal	Pearson Correlation	.922**	.910**	.910**	1	.168*	-.077
	Sig. (2-tailed)	.000	.000	.000		.035	.335
Gender	Pearson Correlation	.127	.169*	.123	.168*	1	-.007
	Sig. (2-tailed)	.111	.034	.123	.035		.925
Grade	Pearson Correlation	-.066	-.078	-.096	-.077	-.007	1
	Sig. (2-tailed)	.411	.33	.228	.335	.925	
** Correlation is significant at the 0.01 level (2-tailed).							
* Correlation is significant at the 0.05 level (2-tailed).							

It can be clearly seen that the significance values between every two dimensions are all 0.000, less than 0.05, which means there is a significant correlation among them, and the correlation values are all >0.9, defined as high correlation. While the correlation results between gender or grade and the four dimensions of the instrument deny the significance, which indicates that neither gender nor grade has impacts on the four dimensions of the instrument. Further research would be conducted to verify the effects of moral education in the testing class.

2.5 Results

Validity is the extent to which an instrument measures what it is intended to measure [18]. In addition, reliability is an essential characteristic of any good test: for it to be valid at all, a test must be reliable as a measuring instrument [19]. On the other words, the instrument can then be used in subsequent research.

The results from validity and reliability analysis tested 158 students spread across different classes from Business English Major shows that the items on the 35-itemd CASI-V, derived from CASI, are valid and highly reliable. In addition, the construct validity of the CASI-V tested also showed 35 items are valid, and for this reason, this instrument can be used to collect data in similar studies in the future. It is concluded that the construct of the CASI-V can help identify the moral awareness of Business English students in China.

3 Application

The present study aims to investigate the impact of the new teaching method on students' improvement in moral education with the newly developed CASI-V. Specifically, the research seeks to determine whether there is a significant difference in the mean scores between Group A and Group B. This research question is of particular interest as educational institutions strive to enhance instructional techniques and optimize the effectiveness of college students' moral education in professional courses.

Previous studies have suggested a positive relationship between self-efficacy and three dimensions of moral education in Business English. However, there is limited research specifically addressing the effects of tech-integrated teaching methods on moral education. Therefore, it is essential to conduct a rigorous analysis to evaluate whether the introduction of the new teaching method yields statistically significant differences in students' moral education compared to the traditional approach.

At present, there are three parallel classes in Grade 2 of Business English Major in our college. As part of the teaching reform in moral education, one class with 32 students was selected as the experimental group (Group A) that received the new teaching model which is based on a tech-integrated flipped classroom, while the other two classes with 77 students were taken as the control group (Group B), who still received the traditional teaching method. Each group took two rounds of questionnaire surveys with CASI-V. The first round was at the beginning of the semester, February 2023, and the second round at the end of the semester, June 2023. After data cleaning, 37 sets from Group B were finalized as the control group. This research adopts two-way repeated-measure ANCOVA to test the effect of moral education in the Business English courses applied with tech-integrated teaching model.

3.1 Method

The purpose of this analysis is to compare the effect of moral education of Group A and B with different teaching methods. The data collected twice are repeated measurements and there are two groups and one covariant, thus two factors (time and group) repeated measurement covariance analysis can be applied with the 7 assumptions as shown in Table 6 [9] [11] [12]:

Table 6. Assumptions for ANCOVA

	Assumptions	Descriptions
1	There is only one observed variable, which should be measured at the continuous level.	Met. CASI-V scale results.
2	There are two analysis factors.	Met. Time and group.
3	The observed variable is repeated measurement data and is not independent.	Met. Samples are the same in the two rounds of tests.
4	One or more covariates exist.	Met. The beginning level and gender.
5	There is no significant abnormal value in the observed variable.	To be tested.
6	The observed variables of each group and each level (time) are normal (or nearly normal) distributions.	To be tested.
7	There should be homogeneity of variances.	To be tested.

Since the data meets these first four assumptions, the two-way ANCOVA might be an appropriate statistical test to analyse the data. However, it is still necessary to carry out multiple procedures in SPSS Statistics and interpret the results from these procedures to check if the data passes each of the rest three assumptions and to determine whether it is the correct statistical test[19].

To verify Assumption 5, Cook distance and the normalized residual (ZRE) of the total scores were checked.

The Cook distance is generally considered to be without an outlier when D<0.5, but with outliers when D>0.5. While the normalized residual (ZRE) is checked to see if there are outliers greater than 3 or less than -3.

Table 7 shows that all the values of Cook distance are less than 0.5, and there is no absolute value exceeding 3 in ZRE, so no special treatment is needed.

Table 7. ZRE and COOK Distance

NO	ZRE	Cook	NO	ZRE	Cook	NO	ZRE	Cook
1	-2.51	.26	24	.02	.00	47	.08	.00
2	1.80	.09	25	.73	.02	48	-.40	.01
3	-1.27	.02	26	1.09	.02	49	.46	.00
4	-1.53	.03	27	1.22	.06	50	.48	.00
5	-1.51	.02	28	.94	.01	51	.67	.00
6	-1.18	.01	29	.20	.00	52	1.32	.01
7	-1.01	.01	30	1.08	.06	53	1.77	.02
8	-.99	.01	31	.54	.01	54	1.77	.02
9	-.54	.00	32	-.69	.03	55	1.81	.03
10	-.54	.00	33	.86	.01	56	1.18	.01
11	-.35	.00	34	.86	.01	57	.30	.00
12	-.48	.00	35	.21	.00	58	.81	.01

13	-.31	.00	36	.21	.00	59	.43	.00
14	-.31	.00	37	-.17	.00	60	.30	.00
15	.51	.00	38	-.17	.00	61	-.71	.02
16	.53	.00	39	-.39	.00	62	-.71	.02
17	1.28	.01	40	-.37	.00	63	-.48	.00
18	.65	.00	41	-.61	.00	64	-.33	.00
19	.82	.01	42	-.82	.04	65	-2.27	.10
20	.59	.00	43	-.44	.00	66	-2.40	.12
21	.76	.01	44	-.29	.00	67	-1.40	.06
22	.13	.00	45	-.10	.00	68	-1.45	.08
23	.32	.00	46	-.07	.00	69	.06	.00

Table 8 shows the normality test results of the total scores. It can be seen that the residuals of Group B do not obey the normal distribution, and the residuals of Group A obey the normal distribution, so the normality is generally acceptable and satisfies Assumption 6.

Table 8. Normality Test

	Group	Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
PreT	A	0.14	32	0.11	0.957	32	0.228
	B	0.146	37	0.046	0.927	37	0.018
ProT	A	0.115	32	.200*	0.959	32	0.261
	B	0.149	37	0.037	0.899	37	0.003
* This is a lower bound of the true significance.							
a Lilliefors Significance Correction							

Figure 3 to Figure 6 show the normality test results in a more visual way.
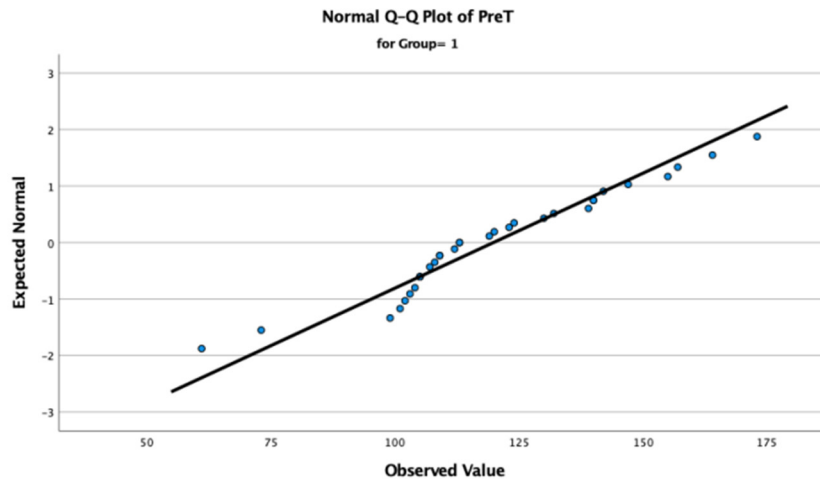

Fig. 3. Normal Q-Q Plot of PreT for Group A
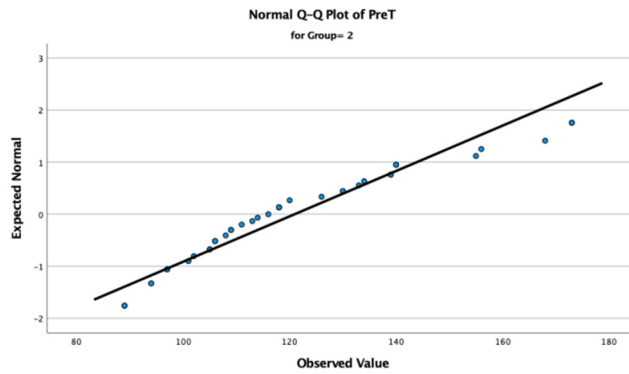
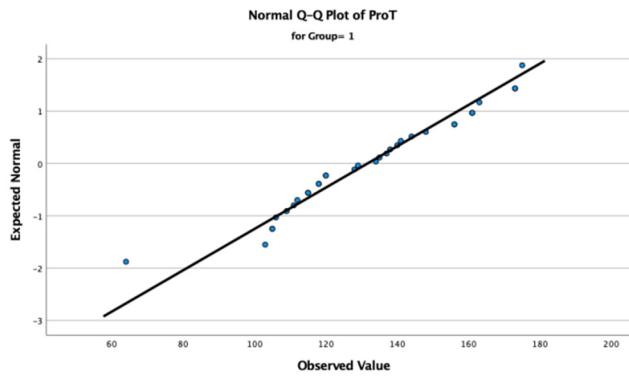**Fig. 4.** Normal Q-Q Plot of PreT for Group B

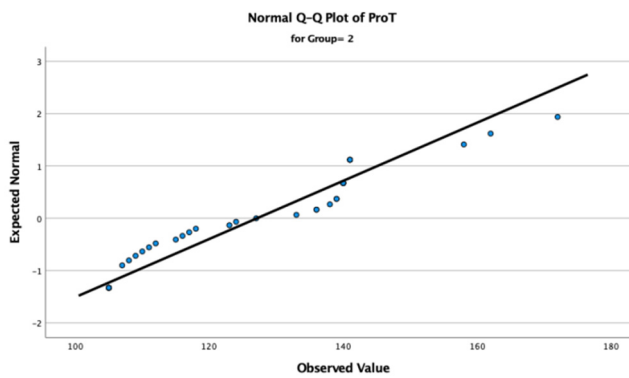**Fig. 5.** Normal Q-Q Plot of ProT for Group A

Fig. 6. Normal Q-Q Plot of ProT for Group B

Table 9 is the test result of homogeneity of variance between groups. It can be seen that the residuals are homogeneous at the pro-test. The population can be considered homogeneous in variance, meeting Assumption 7.

Table 9. Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
ProT	0.413	1	67	0.523
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.				
a Design: Intercept + Sex + Begin + Group				
Within Subjects Design: time				

Based on the above results, the method of ANCOVA is verified to be appliable to this analysis. Through conducting the analysis, the mean scores of the Control and Experimental groups would be compared to determine if any observed differences are statistically significant.

3.2 Results

(1) Descriptive Analysis

Table 10 lists the means of each dimension at pre-test and pro-test in group A and B. It can be clearly seen that the average total score of Group A jumped from 119.81 to 131.66 after the study of one complete semester. Comparatively, the beginning level of Group B was even a little bit higher than that of Group A, and the average total score increased likewise from 120.95 to 127.11, ending up with 4.55 points backwardness.

The following Figure 7 and Figure 8 provide more details of Table 10, presenting the differences of gained scores of each group in a visualized way.

Table 10. Descriptive Means Analysis

Dimension	Group	Pre Means	Pro Means	Cases
Total	A	119.81	131.66	32
	B	120.95	127.11	37
ASE	A	3.6469	3.9063	32
	B	3.6243	3.7514	37
Ctl	A	3.4286	3.8080	32
	B	3.5521	3.7066	37
Crr	A	3.2375	3.6031	32
	B	3.3108	3.5054	37
Ps	A	3.3711	3.7383	32
	B	3.3412	3.5743	37

The estimated marginal mean chart in Figure 7 shows the changes of the total CASI-V scores in two groups at pre-test and pro-test. It can be seen that the average score of the two groups has increased, and the increased range is quite different.
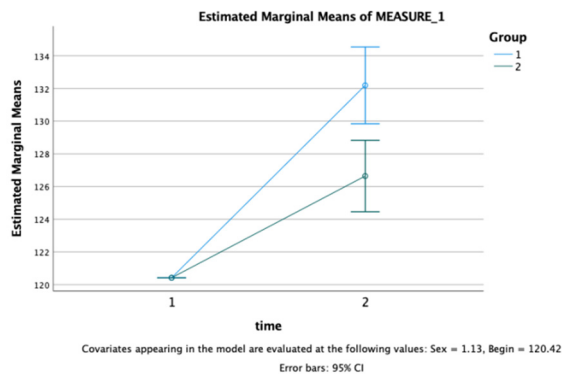
Fig. 7. Estimated Marginal Means

Figure 8 compares the means of each dimension at pre-test and pro-test in Group A and Group B. It can be clearly seen that students of both Group A and Group B have improved in all four dimensions, while Group A gained a greater margin, especially in the dimensions of cultural awareness, career identification, and personal identification.
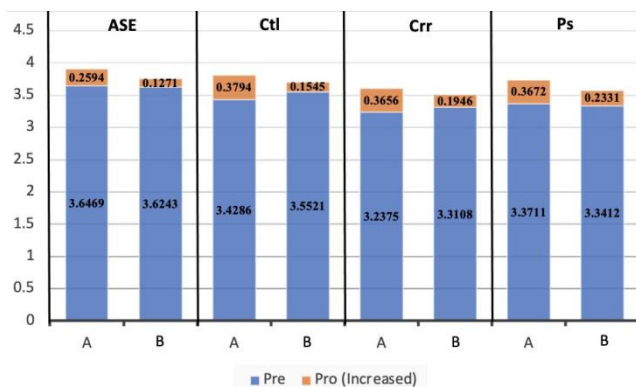

Fig. 8. Means Increased by Dimensions

(2) Interactive Analysis

Because there are two factors in this research(time and group), it is necessary to judge whether there is interaction between the two factors. If the interaction is statistically significant, it is necessary to take further analysis to explore the individual effects.

Table 11. Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	578.131	1	578.13	26.137	0
	Greenhouse-Geisser	578.131	1	578.13	26.137	0

	Huynh-Feldt	578.131	1	578.13	26.137	0
	Lower-bound	578.131	1	578.13	26.137	0
time * Sex	Sphericity Assumed	21.056	1	21.056	0.952	0.333
	Greenhouse-Geisser	21.056	1	21.056	0.952	0.333
	Huynh-Feldt	21.056	1	21.056	0.952	0.333
	Lower-bound	21.056	1	21.056	0.952	0.333
time * Begin	Sphericity Assumed	334.546	1	334.55	15.124	0
	Greenhouse-Geisser	334.546	1	334.55	15.124	0
	Huynh-Feldt	334.546	1	334.55	15.124	0
	Lower-bound	334.546	1	334.55	15.124	0
time * Group	Sphericity Assumed	263.989	1	263.99	11.935	0.001
	Greenhouse-Geisser	263.989	1	263.99	11.935	0.001
	Huynh-Feldt	263.989	1	263.99	11.935	0.001
	Lower-bound	263.989	1	263.99	11.935	0.001
Error (time)	Sphericity Assumed	1437.771	65	22.12		
	Greenhouse-Geisser	1437.771	65	22.12		
	Huynh-Feldt	1437.771	65	22.12		
	Lower-bound	1437.771	65	22.12		

Table 11 shows the results of within-subjects effects. The first line is the result of assumed sphericity, which shows that the sphericity assumption is satisfied. Thus, it is meaningful to view the statistical inference results, which read the F value of interaction between time and group is F(time *group)=11.935, P<0.001, suggesting that the interaction between time and group is statistically significant. Therefore, it is necessary to explore the individual effects. While the F value of interaction between time and sex is F(time *sex)=0.952, P=0.333>0.001, suggesting that the interaction between time and sex is not statistically significant.

The single effect of time refers to comparing the differences of time factors in different groups. The results of "Multivariate Test" (Table 12) give four kinds of statistical inference results, and it is known that there are significant differences in the total score of CASI-V at under different teaching conditions of A and B (P<0.001).

Table 12. Multivariate Tests - Time

	Value	F	Hypothesis df	Error df	Sig.
Pillai's trace	0.659	125.621a	1	65	0
Wilks' lambda	0.341	125.621a	1	65	0
Hotelling's trace	1.933	125.621a	1	65	0
Roy's largest root	1.933	125.621a	1	65	0
Each F tests the multivariate effect of time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.					
a. Exact statistic					

Further, the data of pre-test and pro-test of group A and group B were compared by Bonfroni method, and the results were shown in "Paired Comparison" (Table 13 and Table 14). It is can be seen that the differences were statistically significant (P < 0.001).

Table 13. Pairwise Comparisons Group

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	2.774*	0.803	0	1.171	4.378
2	1	-2.774*	0.803	0	-4.378	-1.171
Based on estimated marginal means						
* The mean difference is significant at the .05 level.						
b. Adjustment for multiple comparisons: Bonferroni.						

Table 14 Pairwise Comparisons Time

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-8.998*	0.803	0	-10.6	-7.395
2	1	8.998*	0.803	0	7.395	10.602
Based on estimated marginal means						
* The mean difference is significant at the .05 level.						
b. Adjustment for multiple comparisons: Bonferroni.						

According to the results of "univariate test" (Table 15), there is significant difference in score gain between the two groups at pro-test (P<0.001), it can be seen that the score gain difference between the two groups is statistically significant (P<0.001).

Table 15. Univariate Test Group

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	131.995	1	131.995	12	0.001
Error	718.886	65	11.06		
The F tests the effect of Group. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.					

3.3 Discussion

The results of this study have the potential to contribute to educational practices by providing evidence-based insights into the effectiveness of the new teaching method and its implications for Business English instruction.

Overall, this study aims to fill the existing research gap and provide valuable information to educators, policymakers, and researchers interested in improving Business English education. By examining the impact of the new teaching method on student performance, we can gain a deeper understanding of its potential benefits and inform future instructional decisions.

4 Conclusion

As research on moral education in China progresses, the theoretical and historical foundations become increasingly clear, and practical challenges have been effectively addressed. Efforts, including the establishment of a moral education system, the development of teacher awareness, curriculum and teaching reforms, and the creation of teaching resources, are becoming more systematic. However, the construction and enhancement of an evaluation system for the implementation of moral education remain incomplete, and a self-improvement mechanism for moral education has yet to be established.

Building on the theory of self-efficacy, this study aims to explore a progressive evaluation system for professional ESP courses from the perspective of moral education. By seamlessly integrating moral education into daily teaching activities, the study seeks to stimulate students' self-efficacy in English learning, thereby enhancing the effectiveness of moral education. However, the current research acknowledges limitations, such as a small sample size for the questionnaire and only a simple reliability and validity test[6]. Subsequent ANCOVA analysis was conducted to examine variances among different dimensions, with positive results supporting the effectiveness of the new teaching method in moral education.

In the future, ongoing research will focus on optimizing the questionnaire, conducting broader studies, and developing measurement tools such as semi-structured interviews with teachers and students. These efforts aim to complete the evaluation kit for moral education and enhance its broader adoption[1].

References

[1]   M. J. Allen, "Assessing Academic Programs in Higher Education", San Francisco: Jossey-Bass, 2004.

[2]   A. Bandura, "Self-Efficacy: The Foundation of Agency," in Control of Human Behavior, Mental Processes, and Consciousness, W. J. Perrig and A. Grob, Eds., Psychology Press, 2000, pp. 16–30.

[3]   N. E. Betz, "Developing and using parallel measures of career self-efficacy and interests with adolescents," Self-efficacy beliefs of adolescents, pp. 225–244, 2006.

[4]   R. W. Brislin, "Cross-cultural research methods: Strategies, problems, applications," in Environment and culture, Springer, 1980, pp. 47–82.

[5]   G. Canino and M. Bravo, "The translation and adaptation of diagnostic instruments for cross-cultural use.," 1999.

[6]   M. Clifford, R. Menon, T. Gangi, C. Condon, and K. Hornung, "Measuring School Climate for Gauging Principal Performance: A Review of the Validity and Reliability of Publicly Accessible Measures. A Quality School Leadership Issue Brief.," American Institutes for Research, 2012.

[7]   J. Ferla, M. Valcke, and Y. Cai, "Academic self-efficacy and academic self-concept: Reconsidering structural relationships," Learn Individ Differ, vol. 19, no. 4, pp. 499–505, 2009.

[8]   S. Geisser, "The predictive sample reuse method with applications," J Am Stat Assoc, vol. 70, no. 350, pp. 320–328, 1975, doi: 10.1080/01621459.1975.10479865.

[9]   G. V Glass, P. D. Peckham, and J. R. Sanders, "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance," Rev Educ Res, vol. 42, no. 3, pp. 237–288, 1972.

[10]   R. K. Hambleton, P. F. Merenda, and C. D. Spielberger, Adapting educational and psychological tests for cross-cultural assessment. Psychology Press, 2004.

[11]   J. Jamieson, "Analysis of covariance (ANCOVA) with difference scores," International Journal of Psychophysiology, vol. 52, no. 3, pp. 277–283, 2004.

[12]   T. R. Johnson, "Violation of the homogeneity of regression slopes assumption in ANCOVA for two-group pre-post designs: Tutorial on a modified Johnson-Neyman procedure," Quant Method Psychol, vol. 12, no. 3, pp. 253–263, 2016.

[13]   P. S. Jones, J. W. Lee, L. R. Phillips, X. E. Zhang, and K. B. Jaceldo, "An adaptation of Brislin's translation model for cross-cultural research," Nurs Res, vol. 50, no. 5, pp. 300–304, 2001.

[14]   Y. H. Kim and N. Choi, "Career decision self-efficacy of Asian American students: The role of curiosity and ethnic identity," Career Dev Q, vol. 67, no. 1, pp. 32–46, 2019.

[15]   M. M. Nauta, J. H. Kahn, J. W. Angell, and E. A. Cantarelli, "Identifying the antecedent in the relation between career interests and self-efficacy: Is it one, the other, or both?," J Couns Psychol, vol. 49, no. 3, p. 290, 2002.

[16]   S. V Owen and R. D. Froman, "Development of a College Academic Self-Efficacy Scale.," 1988.

[17]   K. Schweizer, C. DiStefano, and S. Reiss, "Structural validity of the OSA figures scale for the online self-assessment of fluid reasoning," European Journal of Psychological Assessment, 2016.

[18]   S. Steinberg, "An Introduction to Communication Studies", Juta and Company Ltd, 2007, pp. 52–3. ISBN 978-0-7021-7261-8.

[19]   A. Susanto, Y. Chang, and Y. Ha, "Determinants of continuance intention to use the smartphone banking services: An extension to the expectation-confirmation model," Industrial Management & Data Systems, vol. 116, no. 3, pp. 508–525, 2016.

[20]   M. Taylor, "Values education in Europe: a comparative overview of a survey of 26 countries in 1993", 1993, pp. 1–66. Dundee: Scottish Consultative Council on the Curriculum.

[21]   A. Yada, "Different processes towards inclusion: A cross-cultural investigation of teachers' self-efficacy in Japan and Finland," JYU dissertations, 2020.

[22]   X. Yao, L. Yu, Y. Shen, Z. Kang, and X. Wang, "The role of self-efficacy in mediating between professional identity and self-reported competence among nursing students in the internship period: A quantitative study," Nurse Educ Pract, vol. 57, p. 103252, 2021