

A Pilot Study on the Evaluation of the Integrated Science Curriculum in Zhejiang Province Using the CIPP Model

Zeng Xiangfei¹, Nor Hasnida Md Ghazali^{2*}, Huang Dongyuan³, Yao Yao⁴

¹Faculty of Human Development, Universiti Pendidikan Sultan Idris, Malaysia, ²Faculty of Human Development, Universiti Pendidikan Sultan Idris, Malaysia, ³Junior High School Division, Hainan Micro-city Future School, China, ⁴Senior High School Division, Hainan Micro-city Future School, China

Email: 2421645693@qq.com, 1016126810@qq.com, 2199162749@qq.com

*Corresponding Author Email: hasnida@fpm.upsi.edu.my

DOI Link: <http://dx.doi.org/10.6007/IJARPED/v14-i3/26385>

Published Online: 30 September 2025

Abstract

This pilot study examined the psychometric properties of a questionnaire developed to evaluate the implementation of the Integrated Science Curriculum in Zhejiang Province, China, using the CIPP (Context, Input, Process, Product) model as the guiding framework. The instrument initially contained 91 items, reduced to 88 after expert review and content validity testing. Data were collected from 93 lower secondary science teachers across three urban and rural schools. Normality analysis showed skewness and kurtosis values within acceptable ranges, confirming approximate normal distribution. Reliability tests indicated strong internal consistency, with Cronbach's α values exceeding 0.7 for all constructs. Exploratory factor analysis supported the structural validity of the instrument, with KMO values above 0.6, Bartlett's test significant at $p < 0.05$, factor loadings above 0.5, and cumulative explained variance surpassing 60% across all dimensions. The findings demonstrate that the questionnaire possesses robust psychometric quality and is appropriate for large-scale application. This pilot validation ensures the reliability and validity of subsequent confirmatory factor analysis and structural equation modelling, while providing methodological evidence for evaluating integrated science curricula in the Chinese context.

Keywords: Integrated Science Curriculum, CIPP Model, Pilot Study, Psychometric Validation, Zhejiang Province

Introduction

The development of scientific literacy is increasingly recognised as a cornerstone of modern education, enabling individuals to understand scientific concepts, engage in inquiry, and apply scientific reasoning to address real-world challenges (Miller, 1983; Kennedy & Cherry, 2023). In China, scientific literacy has gained prominence as a national goal, reflected in participation

in international assessments such as PISA, where top-performing regions like Zhejiang achieved leading global scores (Ministry of Education of China, 2019). Despite such achievements, significant regional disparities persist due to uneven resource distribution (Guo & Li, 2024).

At the level of lower secondary education, the Integrated Science Curriculum (ISC) has been promoted as a reform strategy to overcome the fragmentation of disciplinary-based teaching. By combining physics, chemistry, biology, and earth sciences, ISC aims to foster cross-disciplinary understanding and enhance student engagement (Fan, 2004; Xiao & Chang, 2013). Zhejiang Province, a pioneer in implementing ISC since 1988, provides a unique context for evaluating its effectiveness. However, challenges remain, including limited interdisciplinary teaching capacity, inconsistent resource development, and insufficient evaluation frameworks (Huang & Chen, 2014; Wang, 2021).

A systematic and evidence-based evaluation of ISC is therefore critical. The CIPP model (Context, Input, Process, Product) offers a comprehensive framework for assessing curriculum implementation and outcomes. However, to apply this model in practice, a validated measurement instrument is required. This study addresses this gap by developing and piloting a questionnaire designed to evaluate ISC implementation in Zhejiang Province. The pilot study focuses on testing the instrument's psychometric properties, including normality, internal consistency reliability, and construct validity, thereby ensuring its suitability for large-scale application in subsequent research.

Literature Review

The integrated science curriculum has been recognised as a vital means of cultivating students' scientific literacy, which encompasses not only scientific knowledge but also inquiry skills, critical thinking, and attitudes toward science (Miller, 1983; Kennedy & Cherry, 2023). Globally, many education systems have emphasised interdisciplinary science teaching as part of broader STEM education reforms (Teo & Choy, 2021; Markula & Aksela, 2022). In China, the ISC has been gradually introduced to address the shortcomings of fragmented disciplinary teaching. Zhejiang Province, as one of the earliest adopters, has accumulated extensive experience yet continues to face structural challenges in implementation (Wang, 2021; Wang, 2024).

Several barriers hinder the effective promotion of ISC in China. From a contextual perspective, the dominance of examination-oriented education and entrenched preferences for subject-based curricula reduce the acceptance of ISC (Fan, 2004; Xiao & Chang, 2013). In terms of input, most science teachers are trained in single disciplines, limiting their ability to adapt to interdisciplinary requirements (Jia, 2013). Additionally, curriculum resources have been criticised for lacking coherence, often impeding cross-disciplinary integration (Fei, 2012; Pan, 2005). In the process of implementation, urban–rural disparities exacerbate inequities, with rural schools relying more heavily on lecture-based teaching that reduces inquiry opportunities (Zhang, 2022; Miao, 2024). From an outcome perspective, assessments remain overly focused on single-subject test scores, neglecting interdisciplinary competencies and broader indicators of scientific literacy (Liu, 2008; Zhang, 2010).

To address these challenges, curriculum evaluation frameworks are essential. Traditional measurement-oriented models focused narrowly on testable outcomes, overlooking contextual and affective dimensions. Objective-based models measured goal attainment but often ignored unintended consequences. Responsive and goal-free models emphasised stakeholder concerns and emergent outcomes but lacked structure and generalisability. Co-constructive models fostered stakeholder participation but were resource-intensive and subjective. Against this backdrop, the CIPP model (Stufflebeam & Shinkfield, 2012) has gained prominence for its systematic and holistic approach. By evaluating contextual needs, resource inputs, instructional processes, and learning outcomes, the CIPP model is particularly suited for complex curricula such as ISC (Wu, 2021; Liang, 2024). Studies across educational levels, from kindergarten (Liang, 2024) to higher education (Wu, 2021), demonstrate its adaptability and effectiveness.

Nevertheless, applying the CIPP model requires reliable and valid instruments. Existing studies have often developed evaluation tools for specific contexts, such as primary science curricula (Lu, 2023) or undergraduate programmes (Yang & Song, 2020), but systematic evaluation of ISC at the lower secondary level remains limited. This gap highlights the necessity of constructing and validating a robust instrument tailored to ISC in Zhejiang Province. The pilot study presented here therefore contributes both methodologically and contextually by establishing the psychometric foundation for large-scale evaluation.

Methodology

This study adopted a quantitative approach to evaluate the psychometric properties of the ISC questionnaire. The instrument was developed based on the CIPP model, with 91 items initially constructed and later reduced to 88 after face validity and content validity testing, as shown in Table 1. Expert review by six science education specialists confirmed strong clarity and relevance, yielding an S-CVI of 0.99.

A pilot study was conducted in February 2025 in three lower secondary schools in Zhejiang Province, covering both urban and rural contexts. Ninety-three science teachers completed the questionnaire. Data were analysed using SPSS 23.0, focusing on normality, internal reliability, and construct validity. Normality was assessed through skewness and kurtosis, with acceptable ranges defined as -1 to $+1$, and up to ± 2 for larger samples (Hair et al., 2019). Reliability was measured using Cronbach's α , with thresholds set at ≥ 0.7 . Construct validity was tested using exploratory factor analysis (EFA), with criteria including $KMO \geq 0.6$, Bartlett's test $p < 0.05$, factor loadings ≥ 0.5 , and cumulative explained variance $\geq 60\%$ (Williams et al., 2010; Hair et al., 2019).

Table 1

Distribution of Questionnaire Items After Validity and Reliability Testing

Section		Measured Constructs	Number of Items	Item Codes
Section A		Demographic information	8	A1–A8
Section B	Context dimension	Science curriculum standards	6	B1–B6
		Student needs	6	B7–B12
		Science curriculum objectives	6	B13–B18
		Total	18	
Section C	Input dimension	Curriculum content	6	C1–C6
		Learning environment	3	C7–C9
		Teaching resources	3	C10–C12
		Science teachers	6	C13–C18
		Total	18	
Section D	Process dimension	Management of teaching activities	7	D1–D7
		Adaptive teaching methods	9	D8–D16
		Student engagement	4	D17–D20
		Reflection	7	D21–D27
		Feedback	6	D28–D33
		Total	33	
Section E	Product dimension	Scientific understanding	6	E1–E6
		Scientific thinking	3	E7–E9
		Scientific inquiry and practice	3	E10–E12
		Scientific attitude and responsibility	10	E13–E19
		Total	19	

Findings and Discussion**Context Dimension**

Table 2 presents the descriptive statistics for skewness and kurtosis of the 18 items in the context dimension. Overall, the skewness and kurtosis values for most items fall within the range of -1 to $+1$, indicating that the data distribution is approximately normal. The minimum kurtosis value is -1.10 , which is below -1 . Among all items, only B11 has a kurtosis value with an absolute value greater than 1. However, as -1.10 still meets the requirements for subsequent analysis, it is considered acceptable.

Table 2

Descriptive Statistics of Skewness and Kurtosis for Measurement Items in the Context Dimension

	N	Range	Minimum	Maximum	Mean	Standard Deviation
Skewness	18	0.37	−0.93	−0.56	−0.73	0.12
Kurtosis	18	1.28	−1.10	0.17	−0.49	0.34

According to Table 3, the Cronbach's α coefficients for the three measurement constructs in the context dimension are all greater than 0.7 and ≥ 0.9 , indicating a high level of internal consistency reliability. The Cronbach's α coefficients for each individual item also exceed 0.7.

Table 3

Cronbach's α Coefficients for Measurement Items in the Context Dimension

Context Evaluation Constructs	Item	Cronbach's α if Item Deleted	Overall Cronbach's α Value
Science Curriculum Standards	B1	0.91	0.92
	B2	0.90	
	B3	0.89	
	B4	0.90	
	B5	0.90	
	B6	0.90	
Student Needs	B7	0.90	0.91
	B8	0.91	
	B9	0.90	
	B10	0.90	
	B11	0.89	
	B12	0.89	
Science Curriculum Objectives	B13	0.88	0.90
	B14	0.88	
	B15	0.88	
	B16	0.88	
	B17	0.88	
	B18	0.88	

The KMO value is 0.864, which is > 0.6 , and the significance level of Bartlett's test of sphericity is < 0.05 , indicating that the data are suitable for factor analysis, as shown in Table 4.

Table 4

KMO and Bartlett's Test for the Context Dimension

Test	Value
KMO Measure of Sampling Adequacy	0.864
Bartlett's Test of Sphericity	Approx. χ^2
	df
	p
	1084.249
	153
	0.000

The results of EFA indicate that three factors were extracted, as shown in Table 5. The cumulative explained variance of the three factors exceeds 60%, and the eigenvalues for each factor are > 1 . The factor loadings of the items on their respective factors are all > 0.5 , demonstrating strong correlations. Therefore, the findings of the pilot study support the expected theoretical assumptions.

Table 5

Pattern Matrix of the Three-Factor Solution Showing Factor Loadings of Each Item

Item	Component 1	Component 2	Component 3
B1			-0.770
B2			-0.875
B3			-0.904
B4			-0.849
B5			-0.794
B6			-0.789
B7	0.829		
B8	0.790		
B9	0.852		
B10	0.799		
B11	0.871		
B12	0.803		
B13		0.807	
B14		0.821	
B15		0.825	
B16		0.828	
B17		0.822	
B18		0.770	
Eigenvalue	6.568	3.570	2.355
Variance Explained (%)	36.488	19.832	13.081
Cumulative Variance (%)	36.488	56.320	69.401

Input Dimension

The descriptive statistics of skewness and kurtosis for the 18 items under the input dimension are displayed in Table 6. The results indicate that most items fall within the acceptable range of -1 to $+1$, suggesting that the data distribution approximates normality. The lowest skewness value was observed for item C18 (-1.11), which slightly exceeds the -1 threshold but remains acceptable for subsequent SEM analysis. Furthermore, several items, including C3, C10, C14, and C16, also showed skewness values below -1 , yet these still satisfy the general criteria for normality assessment.

Table 6

Descriptive Statistics of Skewness and Kurtosis for Measurement Items in the Input Dimension

	N	Range	Minimum	Maximum	Mean	Standard Deviation
Skewness	18	0.54	-1.11	-0.57	-0.88	0.14
Kurtosis	18	1.29	-0.68	0.61	-0.06	0.36

According to Table 7, the Cronbach's α coefficients for the four input dimension constructs are all greater than 0.7, and all exceed 0.8, indicating high internal consistency reliability. Most items also have a Cronbach's α above 0.7, except for item C8, which has a Cronbach's α of 0.65. However, the correlation coefficients of C8 with other items are all above 0.3, with the lowest inter-item correlation being 0.57, suggesting that C8 remains justifiable for retention.

To improve clarity, the wording of item C8 was refined by replacing “experimental venue” with “laboratory”, as the original wording may have misled respondents by broadening the concept beyond its intended measurement scope. The validity test further determined whether C8 should be retained—if its factor loadings were low across all factors, it would indicate that its retention was unjustified, and it should be removed.

Table 7

Cronbach's α Coefficients for Measurement Items in the Input Dimension

Input Evaluation Constructs	Item	Cronbach's α if Item Deleted	Overall Cronbach's α Value
Curriculum Content	C1	0.87	0.89
	C2	0.87	
	C3	0.89	
	C4	0.86	
	C5	0.87	
	C6	0.86	
Learning Environment	C7	0.78	0.80
	C8	0.65	
	C9	0.72	
Teaching Resources	C10	0.80	0.84
	C11	0.78	
	C12	0.75	
Science Teachers	C13	0.88	0.90
	C14	0.88	
	C15	0.88	
	C16	0.87	
	C17	0.89	
	C18	0.88	

As shown in Table 8, the KMO value is 0.830, which is greater than 0.6, and the significance level of Bartlett's test of sphericity is less than 0.05, indicating that the data are suitable for factor analysis.

Table 8

KMO and Bartlett's Test for the Input Dimension

Test	Value
KMO Measure of Sampling Adequacy	0.830
Bartlett's Test of Sphericity	Approx. χ^2
	df
	p
	902.269
	153
	0.000

The results of EFA indicate that four factors were extracted, as shown in Table 9. The cumulative explained variance of the four factors is 69.575%, exceeding 60%, and the eigenvalues for each factor are all greater than 1. The factor loadings of the items on their respective factors are all greater than 0.5, demonstrating strong correlations. The correlation of C8 with Component 3 reached 0.883, confirming that C8 can be retained.

Table 9

Pattern Matrix of the Four-Factor Solution Showing Factor Loadings of Each Item

Item	Component 1	Component 2	Component 3	Component 4
C1		0.721		
C2		0.789		
C3		0.719		
C4		0.856		
C5		0.771		
C6		0.888		
C7			0.777	
C8			0.883	
C9			0.855	
C10				0.858
C11				0.884
C12				0.781
C13	0.744			
C14	0.791			
C15	0.764			
C16	0.742			
C17	0.838			
C18	0.874			
Eigenvalue	6.103	2.625	2.120	1.675
Variance Explained (%)	33.904	14.584	11.780	9.307
Cumulative Variance (%)	33.904	48.488	60.268	69.575

Process Dimension

The descriptive statistics for skewness and kurtosis of the 33 items within the process dimension are summarised in Table 10. The results show that the majority of items fall within the acceptable range of -1 to $+1$, suggesting approximate normality of the data distribution. The lowest skewness value was recorded for item D7 (-1.17), slightly below the threshold but still acceptable for subsequent SEM analysis. Similarly, items D2, D5, D6, and D11 also exhibited skewness values below -1 , though these do not compromise the validity of later analyses. Regarding kurtosis, the lowest value was observed for item D25 (-1.13), while the remaining items were all within the acceptable range of -1 to $+1$.

Table 10

Descriptive Statistics of Skewness and Kurtosis for Measurement Items in the Process Dimension

	N	Range	Minimum	Maximum	Mean	Standard Deviation
Skewness	33	0.64	-1.17	-0.54	-0.84	0.17
Kurtosis	33	1.84	-1.13	0.71	-0.14	0.50

According to Table 11, the Cronbach's α coefficients for the five process dimension constructs are all greater than 0.7, with the lowest value being 0.87, indicating high internal consistency reliability. Additionally, the Cronbach's α coefficients for all individual items exceed 0.7, further confirming strong internal consistency reliability.

Table 11

Cronbach's α Coefficients for Measurement Items in the Process Dimension

Process Evaluation Constructs	Item	Cronbach's α if Item Deleted	Overall Value
Management of Teaching Activities	D1	0.89	0.90
	D2	0.89	
	D3	0.89	
	D4	0.89	
	D5	0.89	
	D6	0.89	
	D7	0.89	
Adaptive Teaching Methods	D8	0.89	0.90
	D9	0.89	
	D10	0.90	
	D11	0.89	
	D12	0.90	
	D13	0.88	
	D14	0.89	
	D15	0.89	
Student Engagement	D16	0.89	0.87
	D17	0.82	
	D18	0.86	
	D19	0.82	
Reflection	D20	0.81	0.92
	D21	0.91	
	D22	0.90	
	D23	0.90	
	D24	0.90	
	D25	0.90	
	D26	0.91	
Feedback	D27	0.91	0.90
	D28	0.89	
	D29	0.88	
	D30	0.90	
	D31	0.89	
	D32	0.88	
	D33	0.89	

As shown in Table 12, the KMO value is 0.831, which is greater than 0.6, and the significance level of Bartlett's test of sphericity is less than 0.05, indicating that the data are suitable for factor analysis.

Table 12

KMO and Bartlett's Test for the Process Dimension

Test	Value
KMO Measure of Sampling Adequacy	0.831
Bartlett's Test of Sphericity	Approx. χ^2
	1954.680
	<i>df</i>
	528
	<i>p</i>
	0.000

The results of EFA indicate that five factors were extracted, as shown in Table 13. The cumulative explained variance ratio of the five factors is 65.842%, exceeding 60%, and the eigenvalues for each factor are all greater than 1. The factor loadings of the items on their respective factors are all greater than 0.5, demonstrating strong correlations.

Table 13

Pattern Matrix of the Five-Factor Solution Showing Factor Loadings of Each Item

Item	Component 1	Component 2	Component 3	Component 4	Component 5
D1				0.780	
D2				0.772	
D3				0.726	
D4				0.783	
D5				0.851	
D6				0.763	
D7				0.814	
D8	0.694				
D9	0.754				
D10	0.620				
D11	0.756				
D12	0.708				
D13	0.805				
D14	0.797				
D15	0.653				
D16	0.791				
D17					-0.742
D18					-0.784
D19					-0.769
D20					-0.636
D21			-0.799		
D22			-0.751		

Item	Component 1	Component 2	Component 3	Component 4	Component 5
D23			-0.865		
D24			-0.823		
D25			-0.835		
D26			-0.760		
D27			-0.741		
D28		0.806			
D29		0.794			
D30		0.739			
D31		0.790			
D32		0.850			
D33		0.841			
Eigenvalue	9.567	3.947	3.375	3.103	1.736
Variance Explained (%)	28.990	11.960	10.228	9.402	5.262
Cumulative Variance (%)	28.990	40.950	51.178	60.580	65.842

Product Dimension

Table 14 presents the descriptive statistics for skewness and kurtosis of the 19 items in the product dimension. Overall, the skewness and kurtosis values for most items fall within the range of -1 to +1, indicating that the data distribution is approximately normal. The minimum skewness value is -1.25, which is below -1, and corresponds to item F19. However, -1.25 still meets the requirements for subsequent SEM analysis. Additionally, the skewness values for items F10, F12, F14, F15, and F16 are also below -1, but they do not affect the validity of the subsequent SEM analysis. The minimum kurtosis value is -1.01, corresponding to item F1, while the kurtosis values for the remaining items fall within the range of -1 to +1.

Table 14

Descriptive Statistics of Skewness and Kurtosis for Measurement Items in the Product Dimension

	N	Range	Minimum	Maximum	Mean	Standard Deviation
Skewness	19	0.75	-1.25	-0.50	-0.83	0.25
Kurtosis	19	1.76	-1.01	0.75	-0.20	0.56

According to Table 15, the Cronbach's α coefficients for the four product dimension constructs are all greater than 0.7, with the lowest value being 0.81, indicating high internal consistency reliability. Most items also have a Cronbach's α above 0.7, further confirming strong internal consistency reliability. Only item F12 has a Cronbach's α of 0.65, but its lowest correlation coefficient with other items is 0.622, demonstrating relatively high correlation. Therefore, F12 was temporarily retained, subject to further validation in factor analysis. If F12

shows low loadings across all factors, it would indicate that its retention is unjustified, and it should be removed.

Table 15
Cronbach's α Coefficients for Measurement Items in the Product Dimension

Product Evaluation Constructs	Item	Cronbach's α if Item Deleted	Overall Value	Cronbach's α
Scientific Understanding	F1	0.87	0.90	
	F2	0.87		
	F3	0.89		
	F4	0.88		
	F5	0.88		
	F6	0.88		
Scientific Thinking	F7	0.76	0.83	
	F8	0.74		
	F9	0.78		
Scientific Inquiry and Practice	F10	0.77	0.81	
	F11	0.80		
	F12	0.65		
Scientific Attitude and Responsibility	F13	0.89	0.91	
	F14	0.89		
	F15	0.90		
	F16	0.88		
	F17	0.89		
	F18	0.89		
	F19	0.89		

As shown in Table 16, the KMO value is 0.818, which is greater than 0.6, and the significance level of Bartlett's test of sphericity is less than 0.05, indicating that the data are suitable for factor analysis.

Table 16
KMO and Bartlett's Test for the Product Dimension

Test	Value
KMO Measure of Sampling Adequacy	0.818
Bartlett's Test of Sphericity	Approx. χ^2
	998.954
	df
	171
	p
	0.000

The results of EFA indicate that four factors were extracted, as shown in Table 17. The cumulative explained variance ratio of the four factors is 69.275%, exceeding 60%, and the eigenvalues for each factor are all greater than 1. The factor loadings of the items on their respective factors are all greater than 0.5, demonstrating strong correlations.

Table 17

Pattern Matrix of the Four-Factor Solution Showing Factor Loadings of Each Item

Item	Component 1	Component 2	Component 3	Component 4
E1	0.864			
E2	0.840			
E3	0.777			
E4	0.812			
E5	0.780			
E6	0.730			
E7			0.871	
E8			0.773	
E9			0.790	
E10				0.823
E11				0.718
E12				0.930
E13		-0.821		
E14		-0.774		
E15		-0.762		
E16		-0.853		
E17		-0.776		
E18		-0.800		
E19		-0.752		
Eigenvalue	6.065	3.405	2.296	1.396
Variance Explained (%)	31.921	17.921	12.087	7.346
Cumulative Variance (%)	31.921	49.843	61.929	69.275

Conclusion

This pilot study validated the questionnaire developed to evaluate the implementation of the ISC in Zhejiang Province, using the CIPP framework as the guiding model. The results confirmed that the instrument demonstrates strong psychometric properties. Normality tests showed that data distributions across all four dimensions were within acceptable thresholds, supporting the suitability of the dataset for further statistical analysis. Reliability analysis indicated that the constructs exhibited high internal consistency, with Cronbach's α values consistently exceeding the recommended threshold. Minor revisions were made to improve clarity in a small number of items, thereby enhancing measurement precision. Exploratory factor analysis further confirmed the structural validity of the questionnaire, with satisfactory factor loadings, eigenvalues greater than one, and cumulative explained variance exceeding 60% in each dimension.

Overall, the findings establish that the instrument is both reliable and valid for assessing ISC implementation and is ready for subsequent confirmatory factor analysis and structural equation modelling. Beyond methodological validation, this pilot study underscores the importance of systematically evaluating curriculum implementation using a multidimensional framework such as CIPP. The results provide a sound foundation for large-scale application of the questionnaire and contribute to the broader effort of promoting evidence-based curriculum reform in science education.

References

- Fan, X. Y. (2004). *Analysis of factors influencing the implementation of integrated science curriculum* (Master's thesis, Northeast Normal University).
- Fei, M. (2012). Strategies for strengthening conceptual integration in the implementation of integrated science curriculum. *Education and Teaching Forum*, 8, 81–82.
- Guo, Y., & Li, X. (2024). Regional inequality in China's educational development: An urban-rural comparison. *Heliyon*, 10(4), e26249. <https://doi.org/10.1016/j.heliyon.2024.e26249>
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2019). Multivariate data analysis. Cengage learning. *Hampshire, United Kingdom*, 633.
- Huang, X., & Chen, W. H. (2014). Problems and reflections on the promotion of the integrated science curriculum in Zhejiang Province: An empirical study based on the implementation status of Zhejiang's integrated science curriculum. *Teacher Education Research*, 26(2), 8.
- Jia, L. (2013). *Research in the installation of integrated science curriculum in junior middle school* (Master's thesis, Shandong Normal University).
- Kennedy, T. J., & Cherry, A. R. (2023). Sustainable Development Goals and Science and Technology Education. In B. Akpan, B. Cavas, & T. Kennedy (Eds.), *Contemporary Issues in Science and Technology Education* (pp. 131–149). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24259-5_10
- Liang, Y. Q. (2024). Evaluation practice of kindergarten science curriculum based on the CIPP model. *Educational Science Forum*, 16(6), 31–35.
- Liu, J. (2008). Analyzing students' mastery of scientific knowledge in integrated and disciplinary science curricula. *Educational Measurement and Evaluation: Theoretical Edition*, 11, 4.
- Lu, F. (2023). Development and construction of a primary science curriculum evaluation system based on core competencies. *Primary Science Education Research*, 9, 12–17.
- Markula, A., & Aksela, M. (2022). The key characteristics of project-based learning: How teachers implement projects in K-12 science education. *Disciplinary and Interdisciplinary Science Education Research*, 4(1), 2. <https://doi.org/10.1186/s43031-021-00042-x>
- Miao, Q. (2024). High-quality development of science education in rural schools: Value implications, existing challenges, and potential solutions. *Modern Distance Education*, 214, 29–39.
- Miller, J. D. (1983). Scientific Literacy: A Conceptual and Empirical Review. *Daedalus*, 112(2), 29–48. JSTOR.
- Ministry of Education of the People's Republic of China. (2019, December 10). *Chinese students excel in reading, math, science: OECD PISA results*. Retrieved from http://en.moe.gov.cn/news/press_releases/201912/t20191210_411536.html

- Pan, S. (2005). Analysis of the organizational structure of integrated science curriculum content. *Comparative Education Research*, 26(5), 49–54.
- Stufflebeam, D. L., & Shinkfield, A. J. (2012). *Systematic evaluation: A self-instructional guide to theory and practice* (Vol. 8). Springer Science & Business Media.
- Teo, T. W., & Choy, B. H. (2021). STEM education in Singapore. *Singapore Math and Science Education Innovation: Beyond PISA*, 43–59.
- Wang, Y. (2024). Practical exploration of integrated science curriculum development in junior high schools. *Primary and Secondary School Science Education*, (2), 31–35.
- Wang, Y. C. (2021). Cultivating scientific literacy: The exploration of Zhejiang Province's junior high school integrated science curriculum. *Global Education Outlook*, 50(12), 14.
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8, 1–13.
- Wu, X. (2021). Construction and application of a CIPP-based curriculum evaluation system for undergraduate surgery courses. *Chinese Medical Education Technology*, 35(3), 289–293.
- Xiao, H., & Changyun, M. (2013). *The Integrated Science Curriculum in Mainland China* (pp. 189–215). https://doi.org/10.1007/978-94-6209-359-1_10
- Yang, H., & Song, Y. (2020). Construction of a new evaluation model for materials science and engineering curriculum under engineering accreditation. *Teaching and Education Forum*, (39), 232–233.
- Zhang, B. (2010). A new reflection on science curriculum evaluation: From the perspective of value. *Educational Measurement and Evaluation*, (6), 39–40.
- Zhang, J. (2022). Strategies for experimental teaching and management of rural primary school science curricula. *Tianjin Education*, 9, 162–163.