

# Construction and Practical Outcomes of a Value-Added Evaluation Model for Elementary Science Education

Jingyi Jin, Jun Zhang, Mengna Ye, Bingbing Wang, Wusi Wu

School of Education, Zhejiang Normal University, Jinhua, Zhejiang, China

Email: 2197627340@qq.com

**How to cite this paper:** Jin, J. Y., Zhang, J., Ye, M. N., Wang, B. B., & Wu, W. S.

(2025). Construction and Practical Outcomes of a Value-Added Evaluation Model for Elementary Science Education. *Open Journal of Social Sciences*, 13, 552-570.

<https://doi.org/10.4236/jss.2025.1311033>

**Received:** October 19, 2025

**Accepted:** November 23, 2025

**Published:** November 26, 2025

Copyright © 2025 by author(s) and

Scientific Research Publishing Inc.

This work is licensed under the Creative

Commons Attribution International

License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

To address the challenge that summative assessments in elementary science education fail to adequately reflect student growth and the “net effect” of schools, this study developed a value-added evaluation system grounded in CIPP theory and a multilevel model, proposing a “1 + 3 + X” practical framework. Empirical analysis of 121 elementary schools in Zhejiang Province revealed that classroom participation and parental involvement exert a significant positive influence on science literacy gains, while school environment and teaching processes amplify this effect. Practice demonstrates that this model effectively enhances students’ scientific literacy and promotes high-quality, balanced development of science education across the region.

## Keywords

Elementary Science Education, Value-Added Assessment, Multilevel Models, CIPP Theory, Scientific Literacy

## 1. Introduction

The 20th National Congress of the Communist Party of China positioned education, science and technology, and talent as foundational and strategic pillars for comprehensively building a modern socialist country. In response, the central government issued consecutive documents in 2023 and 2024, explicitly calling for “enhancing science education” and “exploring value-added assessment” to ensure evaluation drives learning and teaching. However, as the starting point for cultivating innovative talent, primary school science education predominantly relies on summative assessments centered on end-of-term tests. This approach falls short in dynamically tracking students’ growth in scientific literacy and fails to isolate

the “net effect” of schools by accounting for student background differences. Therefore, establishing a value-added evaluation system based on multilevel models is not only an urgent requirement for implementing national policies but also a crucial breakthrough for promoting high-quality, balanced development of science education across regions.

Against this backdrop, this study introduces a “1 + 3 + X” model with significant practical implications. While it draws on CIPP theory and educational ecology to construct a theoretical foundation for value-added evaluation in elementary science education, its primary focus lies in offering a practical framework. By validating and optimizing the model, the research provides school administrators with a comprehensive understanding of the actual effectiveness of science education, enabling them to identify key factors influencing students’ scientific literacy and offer scientific evidence for decision-making. Furthermore, deepening the application of this model can drive teaching innovation among educators and optimize science education resources, ultimately supporting the cultivation of innovative talent. This approach not only aligns with national policy goals but also offers a scalable solution for enhancing science education quality across diverse educational contexts.

## 2. Research Status

### 2.1. Research Progress on Scientific Education Evaluation Models

Scientific education evaluation is the process of making value judgments on all aspects of scientific education through scientific methods based on specific rules and value standards (Hu, 2016). Research on elementary science education in China has been on the rise since the promulgation of the 2017 Compulsory Education Elementary Science Curriculum Standards, with its core objective being to cultivate students’ scientific literacy (Pan, 2021). In quantifying students’ scientific literacy, some scholars have utilized software to propose an assessment indicator framework and test item development methodology (Li, 2018). However, due to the low utilization of scientific education resources in schools, such as restricted access to research laboratories (Ren, 2020), relevant assessments are limited. Research on science education in elementary schools abroad began relatively early. In 1995, the United States promulgated the National Science Education Standards, which defined the three core elements of scientific literacy: conceptual understanding, procedural skills, and scientific attitudes (National Research Council, 1999). Subsequently, between 2000 and 2015, both PISA and TIMSS progressively adopted context-oriented approaches. By implementing item response theory (IRT), they achieved cross-year longitudinal equating, thereby establishing the technical foundation for value-added assessment (OECD, 2015; Mullis & Martin, 2019). Singapore released the Primary Science Teaching Framework in 2013, emphasizing the integration of knowledge, process, and attitudes, allowing teachers to teach and assess flexibly (Ministry of Education, Singapore, 2014). Over the past five years, with the rise of learning progression theory, scientific educational assessment mod-

els have incorporated latent variable growth curves and the growth percentile model (SGP). By employing quantile regression to characterize individual nonlinear growth trajectories, these models have mitigated the limitations of traditional single-layer models (Yuan & Zhu, 2016). It is evident that science education assessment is gradually shifting from a single-dimensional knowledge evaluation to a comprehensive, dynamic model that assesses multidimensional core competencies. The domestic science education evaluation system at the school level urgently requires improvement.

## 2.2. Current Status of Value-Added Evaluation in Practice

In the field of educational evaluation, value-added refers to the extent to which school education enhances student quality (Harvey & Green, 1993), its positive impact, and the added value it provides (Nan, 2003). It manifests as the difference between students' quality outputs and inputs across various dimensions (Scheerens et al., 2003). Using "relative progress" as the basis for evaluation and controlling for the "net effect" of contextual factors constitutes its core distinguishing feature from traditional educational assessment (Yang & Xia, 2024). The Coleman Report in 1966 sparked discussions on school effectiveness and gave rise to value-added assessment. In 1992, value-added assessment was incorporated into Tennessee's Education Improvement Act (Sanders & Horn, 1998). In the 1990s, the United Kingdom began advocating value-added assessment, and in 2006, it fully implemented a "multi-dimensional" value-added assessment system (Xin et al., 2009). Currently, the Tennessee Value-Added Assessment System (TVAAS) from the United States (Sanders & Horn, 1998), and the Lancashire Value-Added Performance (LVAP) from the United Kingdom are widely accepted internationally (Xin et al., 2009).

Value-Added Assessment in China began in the 1990s and has undergone three distinct phases: introduction (2006-2012), localization (2013-2019), and integration and innovation (2020-present). The introduction phase was marked by Xin Tao and colleagues' systematic elaboration of the TVAAS framework, proposing the replacement of simple difference methods with multi-level models. The localization phase focuses on the compulsory education stage, with Ma Xiaoqiang conducting the first validation of the TVAAS framework's applicability in primary science education at Baoding Primary School in Hebei Province (Xin, 2020). Bian Yufang et al. constructed a four-tiered indicator system—"context, input, process, and output"—to incorporate scientific inquiry interest into the non-cognitive dimension (Bian & Wang, 2013). The integration and innovation phase unfolds along two main threads: First, the integration of intelligent technologies, with Huang Qingpeng proposing a closed-loop approach spanning situational awareness, diagnostic intervention, and outcome tracking (Huang, 2025). Zou Liucong et al. (Zou & Zhang, 2025) developed a Personalized Value-Added Diagnostic Model Using an SVM-LSTM Hybrid Algorithm; Second, model localization adjustments: Zhou Xiangbing et al. modified the SGP model and validated its robustness, incorporat-

ing contextual variables such as household income and parental involvement into the moderation equation for the first time (Zhou et al., 2025). In summary, the scientific education evaluation model is currently in the process of localization, and the application system for value-added assessment remains underdeveloped. Existing value-added evaluations primarily rely on standardized test scores and academic achievement, while non-cognitive factors in elementary education lack large-scale empirical validation. The mechanisms for constructing elementary science education models at the school level have yet to be fully clarified, leaving room for exploration in developing a value-added assessment model for elementary science education based on CIPP and a multi-level framework.

### **3. Theoretical Foundations for Constructing Value-Added Evaluation Models**

#### **3.1. Application Suitability of CIPP Evaluation Theory**

The CIPP evaluation model, proposed by Stufflebeam in the late 1960s, features a four-dimensional framework of “context-input-process-product” that aligns with the comprehensive evaluation needs of educational systems (Qin & Mo, 2022). Transferring this model to elementary science education can establish a logical chain linking macro policies, school resources, classroom implementation, and student development. Domestic empirical research indicates that the CIPP evaluation model demonstrates strong content validity and construct validity in scenarios such as monitoring the quality of elementary STEM after-school programs and science experiment instruction.

#### **3.2. Statistical Support for Multilevel Models**

The core of value-added assessment lies in isolating the “net effect” by eliminating non-educational factors (Yang & Xia, 2024). Hierarchical linear models (HLMs) are internationally recognized as the mainstream approach for addressing bias in nested data. Large-scale assessment systems such as TVAAS and PVAAS employ three-level HLM structures, enabling precise estimation of the true contribution of schools or teachers to students’ value-added gains in science literacy.

#### **3.3. Evaluation Design from an Educational Ecology Perspective**

Bronfenbrenner’s ecological theory posits that student development arises from the interaction of multiple environmental layers: microsystem, mesosystem, exosystem, and macrosystem (Bronfenbrenner, 1979). Placing value-added assessment design within an educational ecosystem perspective ensures that metric design, data collection, and intervention strategies resonate with real-world contexts. If value-added assessments neglect ecosystem variables, they risk producing short-term effects characterized by “high gains but low retention.” Conversely, incorporating ecosystem sustainability metrics into the model significantly enhances the educational improvement value of assessment outcomes. This approach guarantees the sustainable implementation of value-added assessment models across diverse regions and school

types (Kan & Sun, 2025).

## **4. Design and Implementation of a Value-Added Evaluation Model for Elementary Science Education**

### **4.1. Developing a Multi-Level Model for Enhancing Scientific Literacy**

#### **4.1.1. Research Hypothesis**

Based on empirical evidence and scientific theory, the following hypothesis is proposed:

H1 (Student Level): Class participation, parental involvement, and household income are all positively correlated with gains in scientific literacy.

H2 (School Level): The quality of science education implementation at the school level is positively correlated with gains in scientific literacy; modifying the school environment and resource allocation directly impacts gains in scientific literacy among elementary school students.

#### **4.1.2. Research Variables**

Student Layer (First Layer) Variables:

1) XSCY: Student participation in the classroom, measured by the number of roles students hold within the class, is hereafter referred to as “classroom participation.”

2) RJSR: The per capita income of the student’s household is represented by the Richter five-point scale, which indicates the level of household per capita income. The score serves as the household per capita income indicator and is hereafter referred to as “household income.”

3) JZCY: Parental involvement in science education: The parent questionnaire included eight items assessing parental engagement in science education. The mean score of these eight items, calculated using the Likert five-point scale, served as the indicator for parental involvement in science education, hereafter referred to as “parental involvement.”

4) JZRS: Parents’ Perceptions of Science Education The parent questionnaire included nine items assessing parents’ perceptions of science education. The mean score of these nine items, calculated using the five-point Likert scale, served as the indicator for parents’ perceptions of science education. This indicator is referred to as “parent perceptions” in the subsequent text.

School Level (Second Level) Variables:

5) HJJC: The environmental foundation for science education in schools is assessed through a teacher/administrator questionnaire comprising four items. The mean score of these four items, calculated using the five-point Likert scale, serves as the school’s environmental foundation indicator. Hereafter referred to as “environmental foundation.”

6) ZYPZ: The school’s allocation of resources for science education: The teacher/administrator questionnaire included 8 items reflecting the school’s re-

source allocation for science education. The mean score of these 8 items, calculated using the Likert five-point scale, serves as the indicator for the school's resource allocation for science education. Hereafter referred to as "resource allocation."

7) GCSS: The implementation of science education processes in schools was assessed through eight items in the teacher/administrator questionnaire. The mean score of these eight items, calculated using the Likert five-point scale, served as the indicator for the implementation of science education processes in schools. Hereafter referred to as "process implementation."

#### 4.1.3. Model Establishment

Model establishment follows the standard procedure for multilevel linear models, divided into three steps:

##### Step 1: Null Model

Decompose the equation into between-group and within-group differences, without including predictor variables at the first and second levels, that is:

First Level:

$$Y = \beta_0 + \varepsilon$$

Second Level:

$$\beta_0 = \gamma_{00} + \mu_0$$

Using the null model, calculate the interclass correlation coefficient (ICC) across hierarchical levels, representing the proportion of total variance attributable to differences in school-level variables. The ICC is defined as the ratio of group variance to total variance:

$$ICC = \text{Var}(\mu_0) / (\text{Var}(\mu_0) + \text{Var}(\varepsilon))$$

Using the common 0.05 benchmark, if ICC exceeds 0.05, the Y values exhibit significant differences at the second level. The results of the zero-model analysis are shown in **Table 1**.

**Table 1.** Variance component estimation.

Random effects	Standard deviation	Variance
$\mu_0$	4.72086**	22.28648**
$\varepsilon$	7.86312**	61.82868**

$ICC = 0.265 > 0.05$  indicates that there are significant differences in students' science literacy gains at the school level, thus requiring the use of a multilevel linear model for statistical analysis.

##### Step 2: Random Effects Model

This stage analyzes only student-level variables to determine whether they are significant at the school level. The model is as follows:

Student Level:

$$Y = \beta_0 + \beta_1 * XSCY + \beta_2 * JTSR + \beta_3 * JZCY + \beta_4 * JZRS + \varepsilon$$

School Level:

$$\beta_0 = \gamma_{00} + \mu_0$$

$$\beta_1 = \gamma_{10} + \mu_1$$

$$\beta_2 = \gamma_{20} + \mu_2$$

$$\beta_3 = \gamma_{30} + \mu_3$$

$$\beta_4 = \gamma_{40} + \mu_4$$

As shown in **Table 2**, class participation ( $\beta = 0.3680, P < 0.05$ ) and parental involvement ( $\beta = 0.2128, P < 0.05$ ) exerted significant positive predictive effects on the dependent variable scientific literacy value-added ( $Y$ ). Family income demonstrated a significant negative predictive effect ( $\beta = -0.3810, P < 0.05$ ). The positive predictive effect of parental cognition on science literacy value-added was not significant ( $\beta = 0.2121, P > 0.05$ ).

**Table 2.** Results of the random effects model analysis.

	Regression Coefficients and Significance Tests			Variance Components and Significance Tests	
	Regression coefficient	Standard deviation	t-test	Variance component	$\chi^2$ Inspection
Scientific literacy value-added ( $Y$ ) as the dependent variable					
Class Participation	0.3680	0.2036	1.807*	0.3561	39.4312*
Household income	-0.3810	0.1455	-2.619*	0.1765	36.8394*
Parental Involvement	0.2181	0.1455	1.948*	0.1110	38.5161*
Parental Awareness	0.2121	1.1120	1.725	0.0455	33.8850

**Table 2** provides information on the variation in regression effects for each independent variable across different schools. The results of the  $\chi^2$  test indicate that the variance components of the regression coefficients (slopes) for class participation, household income, and parental involvement reached a significant level. Therefore, the regression coefficients for these three independent variables exhibit significant differences across different schools and vary depending on the specific school.

The regression coefficient for parental perceptions shows no significant variation across schools ( $\chi^2 = 33.8850, P > 0.05$ ). Therefore, this study will exclude the regression coefficient for the value-added effect of science literacy based on parents' perceptions when constructing the school-level model.

**Step 3: Complete Model**

Using the regression coefficients with significant  $\chi^2$  tests in **Table 2** as dependent variables, we constructed school-level regression equations with school-related variables as independent variables to explain the variation in science literacy value-added across schools.

Student Level:

$$Y = \beta_0 + \beta_1 * XSCY + \beta_2 * JTSR + \beta_3 * JZCY + \beta_4 * JZRS + \varepsilon$$

School Level:

$$\beta_0 = \gamma_{00} + \mu_0$$

$$\beta_1 = \gamma_{10} + \gamma_{11} * HJJC + \gamma_{12} * ZYPZ + \gamma_{13} * GCSS + \mu_1$$

$$\beta_2 = \gamma_{20} + \gamma_{21} * HJJC + \gamma_{22} * ZYPZ + \gamma_{23} * GCSS + \mu_2$$

$$\beta_3 = \gamma_{30} + \gamma_{31} * HJJC + \gamma_{32} * ZYPZ + \gamma_{33} * GCSS + \mu_3$$

$$\beta_4 = \gamma_{40}$$

As shown in **Table 3**, the environmental foundation significantly strengthens ( $\beta = 0.4922$ ,  $P < 0.05$ ) the positive association between class participation, parental involvement ( $\beta = 0.2901$ ,  $P < 0.01$ ), and science literacy value-added.

Process implementation significantly strengthens ( $\beta = 1.3101$ ) the positive relationship between classroom participation and science literacy gains ( $P < 0.05$ ). Process implementation also significantly strengthens ( $\beta = 0.1088$ ) the positive relationship between parental involvement and science literacy gains ( $P < 0.05$ ).

Resource allocation does not significantly moderate any of these relationships ( $P > 0.05$ ).

School-level variables did not significantly influence the negative relationship between household income and science literacy gains ( $\beta = -0.3866$ ,  $P < 0.05$ ).

**Table 3.** Prediction results of school variables on students' level regression coefficients.

	Regression coefficient	Standard deviation	t-test
Student Science Literacy Value-Added ( <i>Y</i> ) as the independent variable			
Class Participation—Slope*	0.3696		
Environmental Foundation	0.4922	0.2276	2.163*
Resource allocation	0.2126	0.4049	0.525
Process Implementation	1.3101	0.5359	2.445***
Household Income—Slope	-0.3866		
Environmental Foundation	0.0427	0.1990	0.214
Resource allocation	-0.2019	0.4199	-0.481
Process Implementation	0.0959	0.4988	0.192
Parent Involvement—Slope	0.2199		
Environmental Foundation	0.2901	0.1128	2.572**
Resource allocation	0.4533	0.2219	2.043
Process Implementation	0.1088	0.2628	0.414**
Parent Perception Slope	The differences between schools were not significant, and no school-level equation was constructed.		

\*Student Science Literacy Value-Added Slope, same below.

To explain the proportion of variance in student-level variables attributable to

the value added in scientific literacy resulting from school effects, it is necessary to calculate the ratio of conditional variance to total variance.

Based on the model analysis results, substituting the school-level equation into the student-level equation yields the overall equation for the complete model:

$$Y = 0.3866 + 0.3696 * XSCY + 0.4922 * HJJC * XSCY + 1.3101 * GCSS * XSCY - 0.3866 * JTSR + 0.2199 * JZCY + 0.2901 * HJJC * JZCY + 0.1088 * GCSS * JZCY + 0.0241 * JZRS$$

**Table 4** shows the proportion of variance in regression coefficients at the student level explained by school variables. Specifically, 12.55%, 3.62%, and 9.82% of the variance in class participation, household income, parental involvement, and science literacy gains across different schools, respectively, is explained by school-level variables. The variance in parental awareness across schools was not significant; no school-level equation was established, meaning no portion of this variance is explained by school-level variables.

**Table 4.** The variance and proportion explained by school variables.

	Original variance	Conditional variance	Explained variance
Scientific literacy value-added ( <i>Y</i> ) as the dependent variable			
Class Participation	0.3561	0.3114	12.55
Household income	0.1765	0.1701	3.62
Parental Involvement	0.1110	0.1001	9.82
Parent Awareness	0.0455	0.0461	0.00

**Table 5.** Reliability analysis of survey instruments (N = 13,253).

Level	Variable	Number of Items	Cronbach's $\alpha$
Student Level			
	Class Participation	5	0.87
	Parental Involvement	4	0.85
	Parental Cognition	3	0.82
School Level			
	Environmental Foundation	4	0.89
	Resource Allocation	3	0.88
	Process Implementation	5	0.90
Outcome Variable			
	Scientific Literacy (Total)	20	0.92
	Scientific Concepts	4	0.84
	Scientific Competence	4	0.86
	Scientific Engagement	4	0.83
	Scientific Spirit & Character	4	0.85
	Scientific Attitude	4	0.81

**Table 6.** Confirmatory factor analysis: model fit indices and convergent validity.

Model/Latent Variable	$\chi^2/df$	CFI	TLI	RMSEA [90% CI]	SRMR	Composite Reliability	Average Variance Extracted
Seven-Factor Model (All Student & School Variables)	2.95	0.95	0.94	0.06 [0.055, 0.065]	0.04		
Five-Factor Model (Scientific Literacy Subdimensions)	2.81	0.96	0.95	0.05 [0.048, 0.062]	0.04		
Class Participation						0.87	0.58
Parental Involvement						0.85	0.55
Parental Cognition						0.82	0.53
Environmental Foundation						0.89	0.62
Resource Allocation						0.88	0.60
Process Implementation						0.90	0.64
Scientific Literacy						0.92	0.65

To examine the reliability and validity of all measurement instruments, we first conducted Cronbach's alpha coefficient analyses. As shown in **Table 5**, the alpha coefficients for all multi-item scales ranged from 0.81 to 0.92, indicating excellent internal consistency reliability across all scales. Subsequently, we employed Mplus 8.3 to conduct confirmatory factor analysis for structural validity. As shown in **Table 6**, we validated both a model encompassing all seven core constructs and a five-dimensional model of scientific literacy. **Table 6** results indicate that all fit indices for both models met good standards (e.g., CFI > 0.95, RMSEA < 0.06), confirming strong data-model fit. Furthermore, composite reliability exceeded 0.8 for all latent variables, and average variance extracted surpassed 0.5, further demonstrating the scale's outstanding convergent validity.

#### 4.1.4. Conclusion Analysis

Through multilevel modeling, several research hypotheses were validated, yielding the following core conclusions:

At the individual level, classroom participation significantly and positively influences the value-added growth of students' scientific literacy. At the family level, parental involvement significantly and positively impacts the value-added growth of students' scientific literacy, while family income significantly and negatively affects this growth. At the school level, environmental foundations and process implementation strengthen the positive relationship between classroom participation and students' scientific literacy value-added growth, as well as the positive relationship between parental involvement and students' scientific literacy.

## 4.2. Case Study of Model Implementation

### 4.2.1. Sampling Procedure and Inclusion Criteria

To ensure sample representativeness and minimize selection bias, this study em-

ployed a stratified cluster sampling method, randomly selecting 121 elementary schools from regions across Zhejiang Province with varying levels of economic development (high, medium, and low). School inclusion criteria were: (1) offering foundational science courses; (2) possessing basic information technology teaching facilities; (3) agreeing to participate in pre- and post-test data collection. From each school, 1 - 2 classes were randomly selected from grades 3 to 6, involving a total of 13,253 students in the survey. Concurrently, questionnaire data were collected from 12,356 parents and 1215 science teachers to construct a multilevel analysis model.

#### 4.2.2. Comparison of Applications between Elite Schools and Ordinary Schools

To assess the impact of schools on science education, Zhejiang Province’s R Primary School and S Primary School were selected for analysis. The former, designated as a Ministry of Education science education pilot school, possesses a comprehensive science education system. The latter is a regular school that offers relevant courses but lacks a structured science education program.

**Table 7.** Overview of science education at schools R and S.

	Project	Mean	Standard Deviation
School R, Zhejiang Province	Resource allocation	3.7664	0.37493
	Process Implementation	3.9206	0.39844
	Environmental Foundation	4.0658	0.57484
School L, Zhejiang Province	Resource allocation	3.2800	0.33516
	Process implementation	3.0039	0.29331
	Environmental foundation	3.0654	0.32482

**Table 8.** Overview of science education at schools R and S.

	Scientific Concept Mean	Scientific Competence Mean	Scientific Participation Mean	Mean Value of Scientific Spirit and Character	Average Science Attitude Score	Overall mean
R School	4.3255	4.5619	3.8366	4.2472	3.7124	4.1367
S School	3.8789	3.5741	3.8114	3.6024	3.6816	3.7097
R School	0.2756	0.3425	0.2600	0.2500	0.4975	0.3251
S School	0.1156	0.3700	0.1250	0.105	0.2725	0.1976

An analysis of the pre-test scores and value-added gains in scientific literacy among students from the two schools reveals that, as shown in **Table 7**, students from R Primary School demonstrated higher pre-test scores and greater value-added gains in scientific literacy than those from S Primary School. Interviews indicate that the former school has long prioritized science education with a comprehensive system, resulting in higher scientific literacy and value-added gains

among its students. In contrast, the latter school implemented science education for only one year. Although well-received by students, its value-added gains remained lower than those of the former school due to resource constraints (Table 8).

#### 4.2.3. Analysis of Application Differences between Urban and Rural Schools

As shown in Table 9, two primary schools in Zhejiang Province were selected: H Primary School, located in an urban center without specialized science courses, and L Primary School, a rural institution equipped with science education facilities that regularly organizes student science activities.

**Table 9.** Overview of science education at schools H and L.

	Project	Mean	Standard Deviation
School H, Zhejiang Province	Resource allocation	2.2830	0.34269
	Process Implementation	2.0029	0.28651
	Environmental Foundation	2.0345	0.32674
School L, Zhejiang Province	Resource allocation	3.6913	0.39315
	Process Implementation	3.8583	0.36661
	Environmental Foundation	3.7542	0.48039

Analysis of the pre-test scores and value-added gains in scientific literacy for students at both schools reveals that, as shown in Table 10, the pre-test scores and value-added gains in scientific literacy for students at Zhejiang Province's H Primary School were lower than those at L Primary School. Interviews and Table 7 indicate that the former school faces geographical constraints and a scarcity of activities, while the latter offers a systematic curriculum and abundant resources. Consequently, urban children demonstrated lower value-added gains than their rural counterparts.

**Table 10.** Pre-test and value-added results for students at H and L schools.

	Scientific Concept Mean	Scientific Competence Mean	Scientific Participation Mean	Mean Value of Scientific Spirit and Character	Average Science Attitude Score	Overall mean
H School	3.2956	3.1309	3.1309	3.1309	3.1309	3.16384
L School	3.4340	3.5843	3.5843	3.5843	3.5843	3.55424
H School	0.2339	0.0851	0.1583	0.1131	0.1731	0.1527
L School	0.3707	0.1312	0.2538	0.2116	0.2817	0.2498

### 4.3. Building the "1 + 3 + X" Model Framework

#### 4.3.1. "1" Theoretical Model

First, construct a value-added evaluation model based on the CIPP framework as the core theoretical foundation for the entire program.

Contextual Evaluation (C) diagnoses the gap between student baseline and program objectives, providing evidence for policy formulation. Input Evaluation (I) assesses the feasibility of faculty, facilities, and curriculum to ensure precise resource allocation. Process Evaluation (P1) collects interactive data from teachers, students, and administrators, providing timely feedback on instructional inputs and value-added outcomes. Finally, Outcomes Evaluation (P2) conducts a comprehensive assessment of teaching results, incorporating feedback mechanisms for dynamic adjustments.

#### **4.3.2. “3” Digital Platform**

To implement value-added assessment, we are building a three-tier digital platform covering the entire process of “monitoring-diagnosis-intervention.”

The front-end serves school administrators, providing efficient, intelligent visualization interfaces and management tools. The platform supports curriculum resource management, assessment feedback, faculty development, and distinctive school-based development, enhancing teaching resource utilization and school-specific growth.

The middle tier leverages university resources to construct a scientific literacy cultivation program repository, encompassing diverse scenarios such as experimental inquiry and interdisciplinary integration. Based on diagnostic results, it automatically matches 3 - 5 tailored programs while linking them to textbook knowledge points and curriculum standards.

The backend employs a multimodal SVM fusion model that dynamically adapts RBF or multiphase kernel functions. Simultaneously, it constructs an error correction model to dynamically correlate SVM coding outcomes, generating actionable feedback. Through SHAP explainability analysis, it optimizes feature weights and maps student growth trajectories.

#### **4.3.3. “X” Unit Reserve Plan**

The “X” Unit Program Reserve Module aims to establish an integrated system featuring multiple stakeholders, pathways, and domains to support the implementation of value-added assessment.

This system emphasizes tripartite collaboration among schools, families, and communities, establishing a four-pronged mechanism encompassing learning, teaching, assessment, and research. Schools are responsible for designing and executing curricula and instructional plans; families provide essential support and guidance; and the involvement of social resources—such as partnerships with enterprises, research institutions, and community organizations—further expands students’ practical platforms, fostering multidimensional growth.

The multi-path implementation system flexibly configures activities through diverse educational formats—including traditional classroom instruction, after-school extended services, and project-based learning—to address individual student needs. Leveraging assessment tools like scales, students gain clear insight into their strengths and areas for growth during the learning process, enabling teachers to develop more precise personalized teaching plans.

Multi-setting learning designs emphasize interdisciplinary inquiry-based activities—such as laboratory research and museum visits—to broaden students’ learning horizons. These diverse practical environments expand their perspectives while deepening their understanding and application of acquired knowledge.

#### **4.4. Five-Dimensional Dynamic Evaluation Mechanism**

##### **4.4.1. Innovation in Course Formats**

Through scenario simulations, PBL, and interdisciplinary projects, we enhance process feedback, broaden assessment dimensions, and support students’ holistic development.

##### **4.4.2. Evaluation Subject Collaboration**

Establish an evaluation mechanism involving collaborative participation from schools, teachers, students, and parents to facilitate a shift from single-result assessments to formative evaluations. Student assessments will emphasize interest, practical application, innovation, and interdisciplinary competencies. Teacher evaluations will be grounded in fostering moral integrity and cultivating talent, highlighting scientific literacy and teaching effectiveness, while incorporating developmental assessments to promote professional growth.

##### **4.4.3. Building Platforms for Competency Demonstration**

Leveraging online courses, blended learning, and virtual simulation experiments to create a “panoramic learning” assessment environment; utilizing big data technology to establish an intelligent “panoramic information” analysis system; and designing a “panoramic practice” inquiry environment based on a student-centered philosophy. The synergy of these three domains enables a visual, intelligent, and multidimensional evaluation of students’ scientific inquiry skills, problem-solving abilities, and critical thinking.

##### **4.4.4. Evaluation Strategy Transition**

Transcending the limitations of traditional outcome-oriented approaches, this methodology integrates assessment throughout the entire teaching process: - Pre-class foundational evaluations establish baseline starting points; - In-class process and inquiry-based assessments provide real-time feedback; - Post-class value-added, summative, and tracking evaluations ensure continuous follow-up, forming a complete closed-loop system centered on value-added assessment. Through diverse methods including knowledge assessments, classroom observations, and project-based research, it systematically tracks the dynamic development of students’ scientific literacy.

##### **4.4.5. Value-Added Efficiency Activation**

The evaluation process is guided by a longitudinal value-added logic, while the horizontal dimension utilizes placement and diagnostic assessments to anchor baseline levels of student development. The longitudinal dimension relies on formative and value-added assessments to capture learning gains in real time and chart growth trajectories. The multidimensional perspective employs a “content-time-

subject” coordinate system to precisely map the magnitude and direction of individual progress, achieving visual representation of growth.

## 5. Analysis of Model Implementation Outcomes

### 5.1. Improving the Quality of Science Education in Schools

#### 5.1.1. Overall Effectiveness of the Region

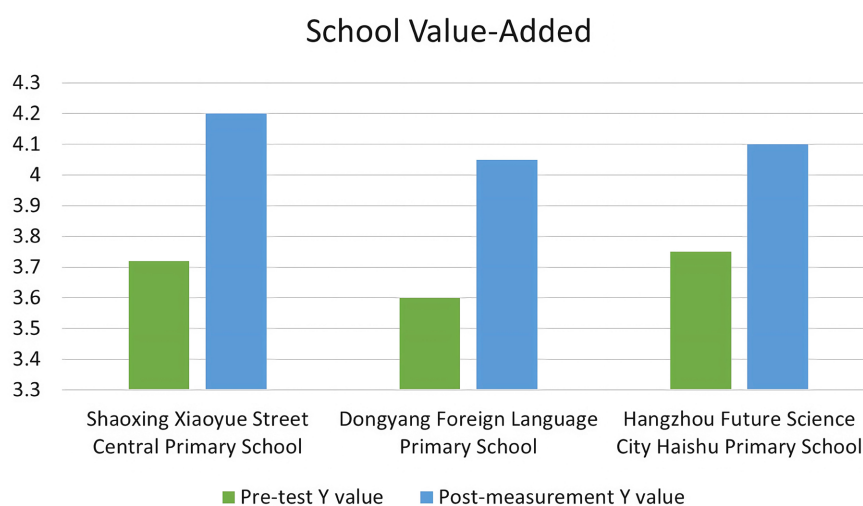
In 2024, Jinhua City passed the review by the Ministry of Education’s expert panel and became one of the first national pilot zones for science education in primary and secondary schools. Among its initiatives, the “1 + 3 + X” value-added assessment model was designated as a core innovative measure for the pilot zone. Focusing on 10 elementary schools within Jinhua’s first cohort of pilot science education schools, the systematic implementation of value-added assessment and improvement plans yielded significant multidimensional enhancements in students’ scientific literacy. As shown in **Table 11**, students achieved value-added gains exceeding 3.2 in areas including scientific knowledge mastery, experimental skills, and innovative thinking.

**Table 11.** Pre- and Post-test results of students’ scientific literacy across different schools.

School	Scientific Concept Mean	Average Science Proficiency	Scientific Participation Average	Mean Value of Scientific Spirit and Character	Average Science Attitude Score	Overall mean
1	Pre-test	3.8789	3.5741	3.8114	3.6024	3.7097
	Post-test	4.2078	3.8066	4.4514	4.1649	4.1170
2	Pre-test	3.7965	4.0203	4.0386	4.3167	4.1078
	Post-test	4.1165	4.2353	4.4236	4.8017	4.4948
3	Pre-test	3.7645	3.7124	3.8453	4.2764	3.9748
	Post-test	4.1112	3.8399	4.2153	4.7914	4.3551
4	Pre-test	3.2322	3.8862	3.0393	3.9573	3.6338
	Post-test	3.5433	4.2061	3.3193	4.4198	3.9945
5	Pre-test	3.7974	3.5022	3.5076	4.4957	3.9239
	Post-test	4.1333	3.8772	4.0876	5.0207	4.3915
6	Pre-test	4.2174	4.0132	3.1265	4.3507	3.9968
	Post-test	4.4731	4.3032	3.3815	4.5707	4.2429
7	Pre-test	4.2174	4.0131	3.1265	4.3507	3.9968
	Post-test	4.1814	4.1364	4.0456	4.4079	4.1811
8	Pre-test	4.2174	4.0132	3.1265	4.3507	3.9968
	Post-test	4.4011	5.1044	4.0966	4.7447	4.4618
9	Pre-test	4.2174	4.0133	3.1265	4.3507	3.9968
	Post-test	4.0464	4.1914	4.0842	3.9874	4.1651
10	Pre-test	4.2174	4.0134	3.1265	4.3507	3.9968
	Post-test	3.5882	4.4397	4.7558	5.4185	4.3833

### 5.1.2. Typical School Case Studies

As shown in **Figure 1**, the three pilot schools demonstrated significant value-added gains. Shaoxing Xiaoyue Subdistrict Central Primary School's Y-value increased from 3.72 to 4.20, representing a 12.8% rise; Dongyang Foreign Language Primary School's score rose from 3.60 to 4.05, reflecting a 12.5% increase; and Hangzhou Future Science City Haishu Primary School's score climbed from 3.75 to 4.10, marking a 9.33% improvement. The results indicate that the value-added assessment model has a replicable positive effect on students' scientific literacy.



**Figure 1.** School value-added.

### 5.2. Social-Level Impact Diffusion

The research findings were disseminated through official platforms such as the Ministry of Education's "Science Education Plus Empowerment" Youth Science Education Symposium. The case study was selected for inclusion in the Collection of Outstanding Practices from Science Education Pilot Zones, providing a referenceable template for improvement. The model has gained recognition from the Future Education Research Institute and the China Collaborative Innovation Center for Basic Education Quality. The model has been adopted by three pilot elementary schools in Zhejiang Province, establishing long-term collaborations. It transforms value-added data into shared resources for school-based teaching research and parent workshops, attracting nearby schools to actively observe and exchange insights. This has fostered a regional collaborative improvement framework expanding from individual schools to broader areas, laying the groundwork for driving the coordinated advancement of science education by diverse societal stakeholders.

## 6. Conclusion and Outlook

### 6.1. Research Findings

The study employed a mixed-methods approach combining quantitative and qualitative data. Through literature review, expert consultation, and in-depth in-

interviews, three sets of questionnaires were developed. Two large-scale surveys were conducted across 121 schools in Zhejiang Province, involving 13,253 students, 12,356 parents, and 1215 teachers. Data were analyzed using multiple techniques.

First, at the family level, both students' classroom participation and parental involvement in science education significantly and positively influenced students' science literacy gains, with more active participation yielding greater gains. Conversely, family income showed a significant negative correlation, indicating that higher material resources in affluent families did not translate into educational advantages. The investment in educational resources may exhibit diminishing marginal returns. Compared to high-income families, middle- and low-income households, constrained by relatively limited resources, may allocate their scarce resources more precisely to their children's education, thereby generating a stronger "compensatory effect." This dynamic could potentially lead to an unexpected negative coefficient for household income. Although our model controls for key variables like parental involvement, the emergence of this negative coefficient suggests the potential presence of unobserved omitted variables. For instance, the model may focus solely on the quantity of parental time spent with children, neglecting the quality of that time and the structure of household educational investments. Parental perceptions had no significant impact on these gains.

Second, at the school level, the foundational school environment significantly amplified the positive associations between classroom participation, parental involvement, and scientific literacy gains. A robust science education environment magnified the positive effects of the former two factors. The quality of science education implementation significantly moderated the value-added effects of classroom participation and parental involvement; high-quality implementation effectively integrated educational resources. The direct impact of school resource allocation on scientific literacy gains was insignificant, indicating that hardware investments require effective utilization mechanisms to yield results.

Finally, in terms of practical outcomes, the "1 + 3 + X" value-added evaluation model drives multidimensional improvements in students' scientific literacy at pilot schools, with particularly notable progress in scientific attitudes and practical abilities. Schools of different types can achieve differentiated development through this model: high-performing schools maintain their advantages through robust systems, while rural schools achieve value-added gains and even overtake others by leveraging distinctive curricula. This model has gained recognition at the regional level, contributing to Jinhua City's designation as a national science education pilot zone. Its social impact continues to expand, demonstrating strong potential for broader implementation.

## 6.2. Future Outlook

Although this study achieved significant results in constructing an evaluation system, limitations remain. The current sample is confined to developed eastern re-

gions and relies heavily on quantitative data, lacking analysis of individual differences. Emerging technologies enhance personalized learning experiences, while policy funding drives development but also exacerbates educational inequality, posing challenges to teachers' information technology literacy.

Future research should therefore focus on the following areas: First, expand the sample coverage. Second, employ qualitative research methods to explore underlying causes. Third, conduct longitudinal studies to collect more dynamic data. Fourth, strengthen training in teachers' information technology application capabilities. Fifth, optimize family participation mechanisms to reduce the impact of household economic conditions on students' science learning. Continuous research and practice will propel primary science education forward, achieving the dual goals of educational equity and high-quality development, thereby providing a solid talent foundation for China's modernization efforts.

### Acknowledgements

We acknowledge the support from supported by the 2025 China National University Student Innovation & Entrepreneurship Development Program (Grant No. 202510345045: Making Growth Traceable: Construction and Implementation Strategies of the Value-added Evaluation Model for Primary School Science Education—Based on a Survey of 121 Schools in Zhejiang Province).

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- Bian, Y., & Wang, Y. (2013). Value-Added Evaluation: An Effective Approach to Assessing School Quality. *Journal of Educational Studies*, *9*, 43-48.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Harvard University Press.
- Harvey, L., & Green, D. (1993). Defining Quality. *Assessment & Evaluation in Higher Education*, *18*, 9-34. <https://doi.org/10.1080/0260293930180102>
- Hu, Z. (2016). *Educational Evaluation* (3rd ed.). China Renmin University Press.
- Huang, Q. (2025). A Framework and Practice of AI-Enabled Value-Added Evaluation in Ideological and Political Education. *Public Relations World*, *No. 15*, 160-162.
- Kan, W., & Sun, Y. (2025). Theory Borrowing, Indicator Construction and Practical Principles of School Evaluation in China's Basic Education: An Educational-Ecology Perspective. *Fudan Education Forum*, *2*, 23-30.
- Li, X. (2018). *A Study on The Thinking-Oriented Teaching Model of Primary Science Based on Core Literacy*. Master's Thesis, Shaanxi Normal University.
- Ministry of Education, Singapore (2014). *Science Syllabus Primary 2014*. <https://www.moe.gov.sg/-/media/files/primary/science-primary-2014.pdf>
- Mullis, I. V. S., & Martin, M. O. (2019). *TIMSS 2019 Assessment Frameworks*. Boston College, TIMSS & PIRLS International Study Center. <http://timssandpirls.bc.edu/timss2019/frameworks/>

- Nan, J. (2003). Educational Value-Added and the Reconstruction of School Evaluation Models. *Journal of The Chinese Society of Education, No. 7*, 59-61.
- National Research Council (1999). *National Science Education Standards* (pp. 1-23). Science and Technology of China Press.
- OECD (2015). *PISA 2015 Draft Science Framework*.  
<https://www.oecd.org/pisa/pisaproducts/draft-pisa-2015-science-framework.pdf>
- Pan, H. (2021). 110 Years of Primary-School Science Curriculum in China (1912-2021). *Education and Teaching Research, 35*, 45-61.
- Qin, F., & Mo, K. (2022). Quality Evaluation of Industry-Education Integration in Vocational Education Based on The CIPP Model. *Journal of Southwest University (Social Sciences Edition), 48*, 194-203.
- Ren, Y. (2020). A Brief Discussion on the Current Situation of Science Education in Chinese Primary Education and Its Implications. *Education Teaching Forum, No. 7*, 56-57.
- Sanders, W. L., & Horn, S. P. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education, 12*, 247-256.  
<https://doi.org/10.1023/a:1008067210518>
- Scheerens, J., Glas, C., & Thomas, S. M. (2003). *Educational Evaluation, Assessment, and Monitoring: A Systemic Approach*. Swets & Zeitlinger Publishers.
- Xin, T. (2020). Key Issues in “Exploring Value-Added Evaluation.” *Primary and Secondary School Management, No. 10*, 1.
- Xin, T., Zhang, W., & Li, X. (2009). Value-Added Assessment: Review and Prospect. *Journal of The Chinese Society of Education, No. 4*, 40-43.
- Yang, L., & Xia, Z. (2024). Fifty Years of Educational Value-Added Evaluation: Evolution, Challenges and Future Paths. *Higher Education Development and Evaluation, 40*, 1-18+119.
- Yuan, Y., & Zhu, B. (2016). An Analysis of “Learning Progressions” Research in Foreign Science Education. *Global Education, No. 7*, 59-64.
- Zhou, X., Zhang, M., Liu, Y. et al. (2025). Revision and Validation of The Student Growth Percentile Rank Model in Value-Added Evaluation: A Case Study of the 2024 Cohort of Changde Zhilan Experimental School. *Educational Measurement and Evaluation, No. 4*, 1-13.
- Zou, L., & Zhang, J. (2025). Can Learning Data Truly “Diagnose” Science Education? Personalized Diagnosis and Evaluation Empowered by Multimodal Data. *Educational Measurement and Evaluation, 1-20*.