



Research on the construction of evaluation index system of science inquiry class for primary school students.

Chenhui Guo^{1,2}, Sudan Jin³, Saiqi Tian^{1,*}

¹College of Education, Wenzhou University, Wenzhou, People's Republic of China.

²Ruian Experimental Primary School, Wenzhou, People's Republic of China.

³Guoxi No. 1 Primary School, Wenzhou, People's Republic of China.

ARTICLE'S INFO

Article No.: 031725043

Type: Research

Full Text: [PDF](#), [PHP](#), [EPUB](#), [MP3](#)

DOI: [10.15580/GJETS.2025.1.031725043](https://doi.org/10.15580/GJETS.2025.1.031725043)

Accepted: 20/03/2025

Published: 10/05/2025

Keywords: Evaluation Index system, Science class, Students, Primary school

***Corresponding Author**

Saiqi Tian

E-mail: tiansaiqi@wzu.edu.cn

Article's QR code



ABSTRACT

In this paper, the construction of the evaluation index system of science inquiry class for primary school students is studied in the context of the rapid development of science and technology. The importance of science education is becoming more and more prominent in today's society, so this study selects the inquiry class, which is the most covered in the primary school science curriculum, as the target for the construction of the evaluation index system, and adopts the Delphi method to conduct an in-depth study. In order to ensure the scientificity and effectiveness of the evaluation index system, we clarified the five constructive principles of human nature, scientific, developmental, systematic, and pertinent, and accordingly constructed an evaluation system including four first-level indexes, namely, scientific concept, scientific thinking, inquiry practice, and attitude and responsibility, as well as a number of second-level indexes. Among them, scientific concepts include scientific knowledge, scientific essence and application of explanation; scientific thinking includes analytical ability, comparison and classification, abstraction and generalisation (modelling), induction and association, and innovative thinking, etc.; inquiry practice includes the secondary indicators of posing questions, conjecturing and hypothesising, experimental methods, formulating plans, investigating plans, gathering evidence, analysing evidence, expressing and communicating, and evaluating and reflecting; and attitude and responsibility includes enthusiasm for learning, pragmatism, teamwork and social responsibility. Through the analysis of the results of the two expert consultation questionnaires, the indicator system was revised and improved, and the final draft of the evaluation indicator system of science inquiry teaching for primary school students was finally formed, which can provide a scientific basis for the evaluation of science inquiry teaching for primary school students.

1. Introduction

In the wave of the rapid development of science and technology, science and technology is recognised as the first productive force that brings many opportunities and challenges to people all over the world. Science education plays a key role in this changing era, and "cultivating scientifically literate citizens" has become the ideal goal of science education in all countries of the world, which has an important guiding significance for science education (Huang, 2021). The quality of science simultaneously determines the process of democracy, the strength of competition and the security of a country (Wei, 2016), science inquiry class as a crucial part of cultivating the scientific quality of primary school students, science inquiry class evaluation index system of the construction of research is the measurement of students' science inquiry class learning outcomes, teachers effectively improve the quality of science inquiry class teaching as well as the optimisation of the whole school science inquiry course educational effect, relatively perfect primary school students' science A relatively perfect evaluation index system for science inquiry class is necessary for the development of

science literacy of primary school students.

2. Background

Science is the foundation of the world, the key method to study, understand and master objective things and their laws (Yang, 2002), the scientific and technological revolution has placed higher demands on the training of students in school science education, which is both highly differentiated and highly integrated, and is the integration of science and technology (Ding, 2001). Against the background of the multi-faceted and multi-level demand for new knowledge in today's society, science education not only focuses on teaching young people basic scientific knowledge, but also emphasises the cultivation of young people's scientific literacy, including scientific viewpoints, scientific know-how and scientific behaviour (Chen, 2001) In order to realize the construction of a strong education country, a strong science and technology country and a strong talent country, the new era needs to use more complete science education methods and methods to cultivate talents with the ability to find and solve problems, to adapt to the needs of social development, and to move towards a future society

with more innovation, rationality and sense of responsibility.

Under the background of the reform of science education, the construction of science evaluation system has also received increasing attention in recent years. However, the construction of the current scientific evaluation system is still faced with some dilemmas, the existence of scientific academic evaluation indicators are not comprehensive, the evaluation of the standard is difficult to unify, the evaluation of the indicator system after the construction of the application of front-line teaching lack of operability and other issues (Wang, 2014). In order to solve these problems, scholars have conducted a lot of research and tried to build a more scientific and reasonable evaluation index system.

The literature research method represents the fundamental approach employed in social science research. It entails a comprehensive examination of a vast array of literature, encompassing the processes of search, collection, identification, collation, analysis, and the elucidation of the nature and characteristics of the research object. This method facilitates the formation of a factual and scientific understanding of the subject matter, which in turn enables the author to articulate their own views and opinions on the content of the study (Du, 2013). The Delphi method, also known as the expert supervision procedure research method or expert consultation method, involves requesting information from experts based on their accumulated knowledge and research experience. Experts then evaluate the correspondence question by judging the information in the questionnaire. This method was first proposed by Weaver and W.(1997). Hierarchical analysis (AHP) is a method that decomposes relevant elements into a hierarchical structure and combines qualitative and quantitative decision analysis. The method establishes a judgement matrix through a process of comparison and utilises the vector corresponding to the largest eigenvalue to determine the weight coefficient for each factor. In terms of specific application, the problem is first transformed into a problem of ranking the elements in the hierarchy. Subsequently, a judgement matrix is constructed for the purpose of carrying out a hierarchical single ranking and consistency test (Fang, 2014). The mathematical statistics method is founded

upon conceptual theory and utilises statistical techniques to analyse and examine data, with the objective of collating results in accordance with statistical laws. Primarily, this method is employed to investigate the relationship between the sample and the parent in a random phenomenon, as well as to identify the regularity between factors (Liu, 2021).

A comprehensive analysis of the extant literature reveals that the study of science academic evaluation in primary schools has attracted considerable attention from scholars and offers a promising avenue for further research. However, an analysis of the extant studies reveals several significant limitations. From the perspective of evaluation content, firstly, the diversity in the cognitive and psychological development stages of primary school students gives rise to differentiated requirements for evaluation. However, previous evaluations have not distinguished between different school segments and designed evaluation tools for these segments. Furthermore, the content of primary school science courses is highly comprehensive and complex, and the current evaluation index system lacks a comprehensive scientific evaluation tool to assist the evaluation subject in assessing students' diverse abilities. The current evaluation index system is still lacking in a comprehensive scientific evaluation tool that can adequately cover the diverse course content of students. Additionally, the majority of existing evaluation tools do not assign weights to the evaluation indexes, which hinders the ability to avoid subjectivity and abstraction. Consequently, the subsequent phase of primary school science academic evaluation research should be further developed and expanded in the following areas:

(1) The evaluation of scientific education is conducted in a systematic manner, with a focus on the specific characteristics of the educational institutions and the diverse content of different instructional approaches. In this study, the inquiry class, which represents the most extensively covered class in the primary school science curriculum, has been selected as the target for the construction of the evaluation index system. Subsequently, the objectives of the curriculum have been analysed in order to design targeted evaluation indexes.

(2) It is necessary to enhance the objectivity of the evaluation index weights. The weights of the indicators

are contingent upon the scientific and rational integrity of the entire evaluation system. Consequently, the potential for subjectivity to influence the resulting evaluation outcomes must be mitigated by ensuring the objective assignment of indicator weights.

3. Methods

Development of the evaluation indicator system for elementary science inquiry class is occurred in three stages: (1) Generate the first draft of the indicator item; (2) Delphi correspondence to determine the indicator system; (3) Reliability validation through weight analysis(Weber, 2021).

3.1 Design

A review of the literature on the academic evaluation of science in primary schools, both domestic and international, was conducted in order to construct an initial evaluation index system for science inquiry classes in primary schools, applicable to students of different age groups. The Delphi method was employed to enhance the content validity of the preliminary evaluation index system, which was then compiled into an expert consultation questionnaire. Experts were invited to propose modifications to the evaluation index system, and several rounds of inquiries were conducted until unanimous agreement was reached. This was done with the aim of establishing an evaluation index system for primary school students' science inquiry classes in the context of the new curriculum standards. Subsequently, the hierarchical analysis method was used to establish the hierarchical structure and compute the weights of primary school students' science academic evaluation indexes in the context of the new curriculum standards.

3.2 Generation of the Initial Draft

The present study was informed by an investigation into the existing practices of science academic evaluation in primary schools. This investigation involved the incorporation of the invaluable insights of those directly involved in the teaching of science, namely, the teachers themselves. These insights were employed to elucidate the

fundamental principles that underpin the construction of the index system that forms the basis of this study. On this basis, the evaluation indicators were initially constructed and subsequently revised by two rounds of experts using the Delphi method, in order to further improve the indicator system and finally form the evaluation indicator system for primary school students' inquiry class.

3.2.1. Principles for the Construction of the Evaluation Indicator System

In order to guarantee the scientific rigour and efficacy of the evaluation index system for the science inquiry class for primary school students, the following five constructive principles are elucidated: (1) The principle of human nature emphasises students as the main body, paying attention to students' cognitive level and individual differences, and adopting a diversified evaluation method, including teacher evaluation, peer mutual evaluation, and students' self-assessment, etc., to ensure that the evaluation serves the development of the students (Zhou, 2011). (2) The principle of scientificity: The evaluation index system should be based on scientific theories and methods, and any subjectivity should be avoided. The new curriculum should be taken as the basis for the system, and the weights of the indicators should be allocated in a reasonable manner. The evaluation standards should be accurate and clear, and the data collection and analysis process should be scientific and reliable (Xin, 2006). (3) The developmental principle entails not only reflecting the current ability level of students, but also paying more attention to their future development. It stimulates students' potential, encourages the pursuit of progress, and dynamically adjusts the indicator system in accordance with the students' development. (4) The principle of systematicity: a comprehensive indicator system should be constructed, comprising primary and secondary indicators, with clear logical relationships among the indicators. The system should be independent of each other and constitute an organic whole, with the objective of comprehensively evaluating the scientific academic level of students' inquiry class(Zhang et al., 2007). (3) The principle of operability: It is essential to guarantee that the indicator system is practical for implementation, transparent and

unambiguous in its content, straightforward to observe and measure, and effective in supporting the evaluation process (Hu & Huang, 2024).

3.2.2. Initial Draft

The evaluation index system for science inquiry classes for primary school students is constructed based on the principles of human nature, science, development, systematicity and operability. The evaluation index system is divided into four first-level indicators: scientific concept, scientific thinking, inquiry practice and attitude and responsibility. Each first-level indicator is comprised of a number of second-level indicators. The term "scientific conception" is used to describe the overarching understanding of objective phenomena that students develop as a result of their comprehension of scientific concepts, laws, and principles. In the construction of the evaluation indicators, the concept of science encompasses specific concepts from the fields of science, technology and engineering, which are collectively represented by the second-level indicator "scientific knowledge." The concept of science also includes an understanding of the nature of science and the relationship between science, technology, and society. This is summarised as the second-level indicator 'the nature of science'. The concept of science includes the understanding of the nature of science and the relationship between science, technology, society and environment, which is summarised as the second-level indicator 'the nature of science'. The understanding of the nature of science and the relationship between science, technology, society and the environment are summarised as the secondary indicator 'scientific essence'. The ability to apply scientific concepts in explaining natural phenomena and solving practical problems is summarised as the secondary indicator 'application and expansion'. Scientific thinking is defined as the way of understanding the essential attributes, internal laws and interrelationships of objective things from a scientific perspective. This mainly includes model construction, reasoning and argumentation, and innovative thinking. In the construction of evaluation indicators, the ability to analyse phenomena and data in a scientific manner is summarised as the secondary indicator 'Analytical ability'. This encompasses the

capacity to compare and contrast essential features of phenomena and to classify them in a logical and systematic manner. The secondary indicator 'Comparison and classification' encapsulates this ability. Scientific thinking also encompasses the capacity to summarise scientific phenomena or to explain their causes using scientific concepts and logical thinking. This is summarised as the secondary indicator 'Abstraction and generalisation'. The capacity to make generalisations and associations based on observations and experimental results is an essential component of scientific thinking. This is encapsulated in the secondary indicator 'Induction and association'. The capacity to formulate generalisations and associations based on observations and experimental outcomes is encapsulated in the secondary indicator "Inductive and associative thinking." Furthermore, creative thinking in science encourages the consideration of diverse perspectives and the pursuit of novel ideas and solutions, which is encapsulated in the secondary indicator "Creative thinking." The term 'inquiry practice' is primarily concerned with the capacity to engage in scientific enquiry, encompassing an understanding and exploration of the natural world, the acquisition of scientific knowledge and the resolution of scientific problems. It also encompasses the ability to learn independently. The secondary indicators under the primary indicator of inquiry practice in the inquiry practice class are as follows: posing questions, formulating conjectures and hypotheses, selecting an appropriate inquiry method, devising an appropriate inquiry programme, data collection, data processing, expression and communication, evaluation and feedback. The formation of a scientific attitude and social responsibility is contingent upon an understanding of the nature and laws of science, as well as the interrelationship between science, technology, society, and the environment. In the construction of the evaluation indicators, the secondary indicator "Enthusiasm for Learning" is derived from the attitudes of responsibility, curiosity, enthusiasm for inquiry, and willingness to explore and practice. The secondary indicator "Factualism" is derived from the abilities to record and report information from experiments truthfully and to express opinions based on facts. The secondary indicator "Teamwork" is derived from the

attitudes of responsibility, willingness to listen and share, ability to analyse and discuss, and capacity to question others' opinions based on facts. The capacity to engage in constructive dialogue, to analyse and discuss, and to challenge others' perspectives in a well-informed manner is encapsulated by the secondary indicator "Teamwork." The attitude of responsibility underscores the interdependence of human lifestyles and modes of production with science and technology, and the necessity to conserve resources and safeguard the environment is encapsulated by the secondary indicator "Social Responsibility." The ability to document and report findings from experiments and to express opinions based on facts is encapsulated by the secondary indicator "Pragmatism."

The second-level indicators facilitate the categorisation of students into low, middle and high levels in accordance with their physical and mental development patterns. This is accompanied by the delineation of specific evaluation criteria for each category. In evaluating primary school students' science inquiry class using this evaluation index system, it is essential to refer to the specific scoring criteria and adopt the Richter 5-point scale as a scoring tool to quantify the performance of primary school students at the current stage of their learning in the science inquiry class. A score of 5 indicates that the student has performed well in accordance with the evaluation criteria for the indicator. A score of 4 indicates that the student has performed well on the evaluation indicator. A score of 3 indicates that the student has neither highlighted nor obvious deficiencies in this indicator. A score of 2 indicates that the student fails to meet the criteria of the evaluation indicator and needs to be strengthened in terms of competence. A score of 1 indicates that the student has obvious deficiencies in this evaluation indicator.

3.3 Delphi Process

The initial evaluation index system for science inquiry classes in primary schools, as derived from the literature, still requires significant improvement. To this end, the Delphi method has been employed to enhance the scientificity and rationality of the evaluation indexes. Through several rounds of expert consultation on the evaluation indexes of primary school science, the indexes can be further improved with the help of the experts' professional knowledge and practical experience. Furthermore, the Delphi method can reduce the potential influence of subjective factors on the construction of the evaluation indexes, thus making the index system more objective. Furthermore, the Delphi method serves to minimise the potential influence of subjective factors on the construction of evaluation indicators, thereby ensuring greater objectivity in the indicator system.

3.4.1. Confirmation of Expert Panel and Inclusion Criteria

The study selected experts and scholars in the field of science education research from colleges and universities, as well as teachers who have been engaged in primary school science education for more than five years with excellent performance, as the target of this expert consultation (Fan & Yang, 2023), by absorbing and drawing on the opinions of experts, we will construct a humanistic, scientific, developmental, systematic and operable evaluation index system of science inquiry class for primary school students under the synthesis of theory and practice. The basic information of the experts is shown in Table 3-1.

Table 3-1 Basic Information of the Experts

Respondent	Gender	Age	Years of Teaching	Professional Title	Academic Degree
A	Male	33	10	First-grade Teacher	Bachelor's
B	Female	33	11	First-grade Teacher	Bachelor's
C	Female	30	5	Lecturer	Doctoral
D	Male	33	10	First-grade Teacher	Bachelor's
E	Male	36	7	First-grade Teacher	Bachelor's
F	Female	39	15	Senior Teacher	Bachelor's
G	Female	48	27	First-grade Teacher	Bachelor's
H	Female	39	19	Senior Teacher	Bachelor's
I	Female	40	18	Senior Teacher	Bachelor's
J	Female	32	9	First-grade Teacher	Bachelor's
K	Female	35	12	First-grade Teacher	Bachelor's

3.4.2. Questionnaire Development

The questionnaire is divided into two parts, the first part is the preliminary constructed evaluation index system of science inquiry class for primary school students (Appendix), which contains the first and second level indexes and their corresponding evaluation standards for each class type, and the second part is formed on the basis of the preliminary constructed evaluation index system of science inquiry class for primary school students, which aims to invite experts to quantitatively evaluate the indexes (Su, 2000), the degree of importance of the indicators is assessed on a scale of '1-5' (from 'not at all important' to 'very important'), and experts are asked to assess the basis of judgement and the degree of familiarity of each indicator. At the same time, the experts were asked to assess the authority of the experts on the basis of the judgement and familiarity of each indicator, and if there is a need for modification in the first and second level indicators or the scoring criteria, the experts were asked to annotate the modifications in the evaluation index system of science inquiry class for primary school students originally constructed in the first part of the questionnaire.

3.4.3. Consultation with Experts Process

In the course of this study, two rounds of expert consultation activities were planned and subsequently conducted, resulting in the distribution of two expert

consultation questionnaires. The initial evaluation index system for science inquiry classes for primary school students was then subjected to a process of preparation by the author, resulting in the formulation of the first expert consultation questionnaire. Following a detailed analysis of the feedback obtained from the initial expert consultation questionnaire, the opinions expressed by the experts and the relevant literature were synthesised in order to facilitate the revision of the primary school science academic evaluation indicators and their associated scoring criteria.

The second expert consultation questionnaire was similarly structured into two parts. The initial questionnaire presented the revised indicators and scoring criteria, as well as statistical information on the experts' choices from the first consultation, including the plurality, the mean score, and the standard deviation. The second part of the questionnaire invited experts to rate the importance of the revised indicators and scoring criteria on a scale of 1 to 5. The second part of the questionnaire invited experts to evaluate the importance of the revised indicators and scoring criteria on a scale of 1 to 5 (with 1 indicating very unimportant and 5 indicating very important). Additionally, experts were encouraged to provide suggestions for modifications in the "Modification Suggestions" section if they felt that modifications were necessary.

3.4.4. Data analysis

The Coefficient of Reliability (Cr) is based on the self-evaluation of the expert and is determined by two factors, one of which is the basis of judgement made by the expert on the programme, i.e., the Coefficient of Authority (Ca), and the other is the degree of familiarity of the expert with the problem, i.e., the Familiarity with the Subject (Cs) (Wang & Siqin, 2011).

3.4.4.1 The Coefficient of Authority (Ca)

The Coefficient of Authority is typically evaluated based on four main criteria: practical experience, theoretical basis, reference to domestic and foreign literature, and intuitive feelings. These criteria are then used to assess the strength of the influence, which is classified as strong, medium, or weak. The quantitative assessment of the Coefficient of Authority is presented in Table 3-2.

Table 3-2: The Quantitative Assessment of the Coefficient of Authority

Basis of Judgment	Level of Influence on Expert Judgment		
	High	Moderate	Low
Practical Experience	0.5	0.4	0.3
Theoretical Basis	0.3	0.2	0.1
Literature Review (Domestic & International)	0.1	0.1	0.1
Intuitive Feelings	0.1	0.1	0.1

The value of the coefficient of influence of judgement (Ca) is equal to the frequency of each basis of judgement and the score of the basis of judgement multiplied by the total number of experts after the sequential addition, according to the calculation of the

sum of the coefficient of influence of judgement is 0.909, the frequency of the basis of judgement of each expert, the frequency of the basis of judgement and the coefficient of influence of judgement of the basis of judgement is shown in Table 3-3 below.

Table 3-3 Frequency, Proportion, and Influence Coefficients of Expert Judgment Criteria

Basis of Judgment	Level of Influence on Expert Judgment						Influence Coefficient
	High		Moderate		Low		
	Freq.	%	Freq.	%	Freq.	%	
Practical Experience	9	81.81	2	18.18	0	0.00	0.4818
Theoretical Basis	3	27.27	7	63.63	2	18.18	0.2272
Literature Review (Domestic & International)	1	9.09	5	45.45	5	45.45	0.1000
Intuitive Feelings	6	54.54	4	36.36	1	9.09	0.1000

3.4.4.2. The Familiarity with the Subject (Cs)

At this juncture of the study, in order to accurately gauge the experts' familiarity with the indicators, they were meticulously categorised into five levels, namely "very familiar", "familiar", "average", "unfamiliar" and "very unfamiliar". The final two categories were "unfamiliar" and "very unfamiliar." In order to facilitate

subsequent statistical analyses and calculations, specific quantitative values were assigned to each of the five levels, as follows: 1 point for 'very familiar', 0.8 points for 'familiar', 0.6 points for 'general', 0.5 points for 'very unfamiliar', and 0.6 points for 'very unfamiliar'. A value of 0.6 was assigned to the category "very familiar," 0.8 to "familiar," 0.6 to "average," 0.4 to "unfamiliar," and 0.2 to "very unfamiliar." In light of the

feedback from the Expert Group, which indicated that six experts self-assessed their familiarity with the indicator as 'familiar' and the remaining five experts considered their familiarity with the indicator to be at the 'average' level, the level of familiarity of the Expert Group with the indicator was calculated on the basis of the quantification table as 0.709.

3.4.4.3. Specialist authority coefficient (Cr)

The degree of authority attributed to the experts is determined through a process of combining their assessment of the indicators with their familiarity with said indicators. This is achieved through the application of a specific formula: The calculation of the degree of authority of the experts is achieved through the combination of the basis of the experts' judgement on the indicators and the experts' familiarity with the indicators, with the specific formula: A degree of authority exceeding 0.7, as indicated by Cr, is typically regarded as high. This signifies that the experts' judgement is characterised by a considerable degree of credibility and professionalism.

In accordance with the aforementioned formula, the coefficient of the degree of authority of the aforementioned panel of experts is calculated to be 0.809, which is greater than 0.7. It can thus be concluded that the coefficient of expert authority of this questionnaire is high, and that the questionnaire is therefore a reliable source of data.

4. Results

4.1 Analysis of Inaugural Expert Consultation Questionnaire

The expert consultation employed a combination of face-to-face interviews between the author and experts and written questionnaires. A total of 11 expert questionnaires were distributed, and the response rate was 100%. The following section presents an analysis of the responses obtained from the initial expert consultation questionnaire. In light of the feedback received from the initial expert consultation questionnaire, a series of significant amendments were made to the evaluation index system for the science inquiry class for primary school students.

The initial evaluation index system for science inquiry classes in primary schools is divided into four first-level indexes: scientific conception, scientific thinking, inquiry practice, and attitude and responsibility. The final category is that of 'attitude and responsibility'. Second-level indicators are included under each first-level indicator, resulting in a total of 20 second-level indicators. These second-level indicators are divided into low, middle, and high levels based on the characteristics of students' physical and mental development patterns. Specific evaluation criteria, totalling 60 items, have been developed for these second-level indicators.

The feedback was primarily concentrated on two key areas. Firstly, the design of the evaluation criteria for the evaluation indicators remains problematic, particularly in terms of compatibility with the new curriculum standards. In particular, the evaluation criteria that can be achieved by students in the lower grades are overly ambitious and require adjustment in line with the new curriculum standards and the physical and mental development characteristics of primary school students. Secondly, the preliminary indicators are perceived as being overly academic, which may impede comprehension among the evaluation subjects, including front-line teachers, students and parents. Secondly, the experts considered the presentation of the preliminary constructed indicators to be overly academic, which might impede comprehension among the evaluation subjects, including frontline teachers, students and parents. Consequently, the language of the indicators was revised to facilitate more accurate grasp of the key evaluation points by the aforementioned subjects. Following the collation of the experts' opinions, the author proceeded to delete four evaluation criteria, amend 34 evaluation criteria and add three new evaluation criteria.

Following the collation of expert suggestions, the indicator should be amended by replacing the secondary indicator, "Application and extension," under the primary indicator, "Scientific concepts," with "Application of explanations." Additionally, the evaluation of "scientific nature" in the lower band should be deleted, and the secondary indicator, "Abstraction and generalisation," under the primary indicator, "Scientific thinking," should be replaced with "Abstraction and generalisation (modelling)." The

evaluation of the 'nature of science' in the lower band should be deleted; the second indicator, 'abstraction and generalisation', under the first level indicator, 'scientific thinking', should be replaced by 'abstraction and generalisation (modelling)'; and the second indicator, 'scientific concepts', should be replaced by 'application and interpretation'. Furthermore, "creative thinking" should be replaced with "innovative thinking," "data collection" with "data collection," and "scientific thinking" with "scientific thinking." The terms "data collection" and "data processing" should be replaced with "collection of evidence" and "analysing evidence," respectively. The term "evaluation and feedback" should be replaced with "evaluation and reflection," and the indicator "methods of inquiry" should be deleted and replaced with a new indicator, "planning." The indicator "Methods of inquiry" should be deleted and a new secondary indicator, "Developing a plan," should be added. Furthermore, the secondary indicator "Sense of social responsibility" under the primary indicator "Attitudinal responsibility" should be replaced with "Social responsibility."

4.2 Analysis of Second Expert Consultation Questionnaire

This expert consultation employed a combination of face-to-face interviews between the author and

experts, in addition to written questionnaires. Eleven expert questionnaires were distributed, and a 100% recovery rate was achieved. The following section presents an analysis of the questionnaire administered during the second expert consultation. In response to the feedback provided in the questionnaire distributed during the second expert consultation, the following amendments were made to the evaluation index system for the science inquiry class for primary school students.

At the second expert consultation, the experts reached consensus on the primary and secondary indicators and did not propose any amendments. In the second round of comments, the experts provided feedback on the detailed presentation of individual indicators in practical application. One evaluation criterion was amended according to the experts' comments, and no deletion or addition was made to the evaluation criteria.

4.3 Final draft of the evaluation index system of science inquiry class for primary school students

The following final draft (Table 4-1) of the evaluation index system of science inquiry class for primary school students was formed after two rounds of expert consultation.

Table 4-1 Scientific Literacy Evaluation Framework for Primary Students

Primary Dimension	Secondary Dimension	Grade Level	Assessment Criteria
Scientific Concepts	Scientific Knowledge	Lower	Recognize basic external characteristics of common objects and provide simple descriptions; understand meanings of scientific vocabulary.
		Middle	Understand fundamental knowledge in physics, biology, geography, and astronomy; describe scientific vocabulary meanings.
		Upper	Comprehend and apply various scientific concepts; connect scientific knowledge with daily life.
	Nature of Science	Middle	Design and conduct experiments to understand reproducibility of results; recognize provisional nature of scientific theories.
		Upper	Appreciate principles of simplicity and logical coherence in scientific theories; understand

Primary Dimension	Secondary Dimension	Grade Level	Assessment Criteria
			repeatable verification.
	Application & Interpretation	Lower	Describe natural phenomena; explain human-environment interdependence.
		Middle	Use scientific knowledge to explain common phenomena; recognize importance of resource conservation.
		Upper	Solve scientific problems using knowledge; understand environmental impact mitigation strategies.
	Analytical Skills	Lower	Observe and summarize phenomena with guidance; demonstrate evidence awareness.
		Middle	Analyze component relationships; connect facts with viewpoints.
		Upper	Establish evidence-hypothesis linkages; conduct controlled variable analysis.
	Comparison & Classification	Lower	Compare superficial features; classify under guidance.
		Middle	Compare essential characteristics; develop functional analogies.
		Upper	Conduct comprehensive classification with independent criteria.
Scientific Thinking	Abstraction & Modeling	Lower	Describe phenomena simply; distinguish facts from opinions.
		Middle	Represent 3D objects in 2D; use models to explain phenomena.
		Upper	Construct dynamic models to interpret scientific processes.
	Induction & Association	Lower	Make basic inferences from observations.
		Middle	Formulate scientific explanations through systematic induction.
		Upper	Apply deductive reasoning to real-world problems.
Inquiry Practice	Innovative Thinking	Lower	Generate ideas from multiple perspectives.
		Middle	Develop novel solutions using creative thinking methods.
		Upper	Propose science-based innovations with logical analysis.
	Questioning	Lower	Generate diverse ideas about object functions.
		Middle	Formulate investigable questions using scientific vocabulary.
		Upper	Propose structured research questions across dimensions.
	Hypothesis	Lower	Make basic assumptions.

Primary Dimension	Secondary Dimension	Grade Level	Assessment Criteria
Attitudinal Responsibility	Formation	Middle	Develop hypotheses based on conditions and processes.
		Upper	Create multidimensional hypotheses.
	Experimental Methods	Lower	Conduct guided observations.
		Middle	Apply systematic inquiry methods.
		Upper	Implement advanced investigative approaches.
		Lower	Recognize planning importance.
	Planning	Middle	Design simple controlled experiments.
		Upper	Develop comprehensive safety-conscious plans.
	Evidence Collection	Lower	Record basic observations.
		Middle	Document data using scientific terminology.
		Upper	Design independent data collection systems.
	Data Analysis	Lower	Perform basic data sorting.
		Middle	Apply elementary statistical methods.
	Communication	Upper	Conduct advanced quantitative analysis.
		Lower	Describe findings simply.
		Middle	Present structured reports.
	Evaluation & Reflection	Upper	Produce formal scientific papers.
		Lower	Accept basic feedback.
		Middle	Implement improvement strategies.
	Learning Engagement	Upper	Conduct comprehensive meta-analyses.
		Lower	Show curiosity about natural phenomena.
		Middle	Demonstrate experimental enthusiasm.
	Intellectual Integrity	Upper	Exhibit investigative persistence.
		Lower	Record observations faithfully.
		Middle	Maintain factual accuracy.
		Upper	Exercise independent judgment.
	Collaboration	Lower	Share ideas receptively.
		Middle	Engage in constructive debates.
Upper		Lead evidence-based discussions.	
Social Responsibility	Lower	Recognize technology's dual nature.	
	Middle	Understand human-environment interactions.	
		Upper	Advocate sustainable practices.

5. Discussion

5.1 First-Round Expert Consultation

In the first expert consultation, with the help of the mean (M), we can judge the concentration trend of the experts' ideas on the importance degree of the indicator, the larger the mean value is, the higher the

concentration of the experts' recognition of the importance degree of the indicator is, and vice versa, the lower the recognition, that is to say, the lower the importance degree of the indicator is; and through the standard deviation (SD), we can reflect the dispersion of the experts' views on the importance degree of the indicators, the smaller the SD is, the more the experts'

opinions are concentrated, and vice versa, it means the less the experts' opinions are concentrated. The smaller the standard deviation (SD) is, the more concentrated the experts' opinions are, and vice versa, the less concentrated the experts' opinions are. By collating and analysing the data from the expert consultation questionnaire, the mean and standard deviation were calculated and the data were analysed as follows.

As illustrated in Table 5-1, the mean scores of the indicators for each lesson type exhibited a mean value exceeding 4.00 during the initial expert consultation. Additionally, the mean value of the standard deviation was observed to be consistently below 1.00. However,

it is noteworthy that a few indicators demonstrated a mean value below 4.00, accompanied by a mean value of the standard deviation exceeding 1.00. This indicates that the experts' opinions on the general indicators of this questionnaire are more generalised. A detailed examination of the data reveals that experts hold disparate views on the significance of the evaluation indicators in the lower section (grades 1 to 2). In light of this, the present round of expert consultation questionnaire focuses on the evaluation indicators and evaluation criteria for the lower section, in conjunction with the recommendations of the experts, with a view to making the necessary adjustments and corrections.

Table 5-1 Means and Standard Deviations of Science Literacy Evaluation Indicators (First-Round Expert Consultation)

Dimension	Grade Level	Scientific Concepts		Scientific Thinking		Inquiry Practice		Attitudinal Responsibility		
		M	SD	M	SD	M	SD	M	SD	
Grade-Specific	Lower	3.69	1.08	3.91	1.01	3.89	0.99	4.62	0.47	
	Middle	4.32	0.66	0.66	4.39	0.60	4.42	0.56	4.84	0.33
	Upper	4.65	0.66	4.83	0.34	4.71	0.44	4.91	0.18	
Dimension Average		4.22	0.74	4.38	0.65	4.34	0.66	4.79	0.32	
Grand Summary				Overall Mean = 4.43						
				SD/M Ratio = 0.13						

5.1.1. Expert Opinion Concentration Analysis

The absolute value of the difference between the plural (MO) and the mean (M) in the results of this round of the expert consultation questionnaire indicates the degree of consistency of the experts' opinions. If $|M_0 - M| \leq 1.00$, it indicates a high degree of consistency of experts' opinions, and on the contrary, if its value > 1 , it indicates a high degree of divergence of experts'

opinions on this questionnaire. As can be seen from Table 3-10, the mean values of the indicators of the first expert consultation questionnaire of this study $|M_0 - M|$ are < 1.00 , but some of the mean values are at the edge of the critical value of 1.00, indicating that the experts' opinions on this questionnaire are more consistent in general, so it is necessary to further adjust the evaluation indexes and their evaluation criteria in the subsequent study, and to correct the indexes based on the opinions of the experts.

Table 5-2: Evaluation Indicators for Science Academic Performance | M0-M | Average Value (First Expert Consultation)

	Scientific Concepts				Scientific Thinking				Inquiry Practice				Attitudinal Responsibility				Total Average	M0-M /M
	Lower	Middle	Upper	Average	Lower	Middle	Upper	Average	Lower	Middle	Upper	Average	Lower	Middle	Upper	Average		
M ₀ -M	0.93	0.53	0.26	0.57	0.95	0.41	0.17	0.51	0.73	0.42	0.29	0.48	0.18	0.16	0.09	0.14	0.43	0.10

5.2 Statistical Analysis of Expert Opinions from the Second Consultation

5.2.1. Concentration of Expert Opinions

By collating and analysing the opinions of the second round of expert consultation questionnaires, the mean score and standard deviation of the questionnaires were obtained. From Table 3-16, it can

be seen that in this expert consultation questionnaire, the average mean score of each index is higher than 4.00, and the average standard deviation is lower than 1.00, and the ratio of the average standard deviation to the average mean score (SD/M) is 0.14 after retaining two decimal places, which can be seen that the experts' opinions on the questionnaire are highly concentrated.

Table 5-3: Average Values and Standard Deviations of Evaluation Indicators for Science Academic Performance (Second Expert Consultation)

	Scientific Concepts				Scientific Thinking				Inquiry Practice				Attitudinal Responsibility				Total Average	SD/M
	Lower	Middle	Upper	Average	Lower	Middle	Upper	Average	Lower	Middle	Upper	Average	Lower	Middle	Upper	Average		
M	4.09	4.33	4.61	4.34	3.62	0.57	4.83	3.01	4.41	4.52	4.63	4.52	4.55	4.77	3.93	4.42	4.07	0.14
SD	0.99	0.76	0.67	0.80	0.77	0.57	0.35	0.56	0.62	0.61	0.49	0.57	0.53	0.43	0.22	0.39	0.58	

5.2.2. Consistency of Expert Opinions

The absolute value of the difference between the multitude (M0) and the mean (M) is indicative of the degree of consistency observed in a set of data. A value of |M0-M| ≤ 1 indicates a high degree of consistency among experts' opinions. Conversely, a value >1 indicates a high degree of disagreement among experts' opinions on this questionnaire. the

overall mean value of |M0-M| for each level 1 indicator in the questionnaire of the second expert consultation in this study is 0.42, and the mean value of |M0-M| is <1, which indicates that there is a high degree of consistency in the experts' opinions about this questionnaire. In comparison to the initial expert consultation, the ratio of |M0-M|/M is 0.09 less than 0.10, which suggests that the experts' opinions are more consistent.

5.3. Weight analysis

A review of current scientific academic evaluation tools revealed that the majority of index systems have not yet assigned weights to each evaluation index, which may compromise the objectivity of the evaluation to some extent. In light of the aforementioned shortcomings, this paper introduces the Analytic Hierarchy Process (AHP), solicits expert input on the relative importance of each indicator, and determines the weights of each indicator with the aim of enhancing the accuracy and scientific rigour of the evaluation process.

In the context of multi-objective decision-making, particularly in complex systems involving multiple indicators, such as the evaluation of primary school science, it is of paramount importance to ascertain the weight distribution among the various indicators (Liu, 2021). A study was conducted on the construction of a teaching evaluation index system for basketball specialised courses in Jiangsu colleges and universities based on core literacy. In 2021, the objectivity and reasonableness of the weights assigned to each indicator will have a significant impact on the final evaluation results (Guo et al., 2014). In light of the distinctive context in which the evaluation index system for science inquiry classes in primary schools is deployed, namely with students as the primary subjects of evaluation and teachers, students and parents as the primary evaluators, the system must accommodate subjectivity and reasonableness. Consequently, this paper opts for the hierarchical analysis method as the weight calculation method. In the selection of experts, the previous expert group, which was involved in the construction of the indicator system, offers a greater number of modifications. The evaluation indicator system is also more familiar to the experts who are invited again to evaluate the importance of the indicators and the degree of the evaluation of two to two comparisons. This is done in order to arrive at the evaluation of the indicator weights.

Ultimately, the weights of the primary and

secondary indicators were calculated, and the comprehensive weights were calculated by multiplying them. The results of the calculation of the comprehensive weights of the primary school science academic evaluation indexes for each type of lesson will be presented in the following section.

5.3.1. Analysis of Weight Results for Lower Grade Levels

The results of the scoring exercise were discussed and summarised in order to ascertain the relative importance of the various indicators. Once the two-by-two judgement matrix had been obtained, it was possible to calculate the random consistency ratio (CR) of the weights assigned to the primary indicators, which was found to be 0.0847. Similarly, the CR of the weights assigned to the secondary indicators under the heading of 'scientific thinking' was calculated to be 0.0324, while the CR of the weights assigned to the secondary indicators under the heading of 'inquiry and practice' was calculated to be 0.0396. The CR of the weights of the secondary indicators under the heading of 'scientific thinking' is 0.0324, under the heading of 'inquiry and practice' is 0.0396, and under the heading of 'attitude and responsibility' is 0.0535. All of these values are lower than 0.1, indicating satisfactory consistency in the results of the hierarchical analyses. This is further supported by the reasonable distribution of the weights. The distribution of coefficients is highly reasonable. The final step was to multiply the weights of the primary indicators by those of the secondary indicators in order to calculate the comprehensive weights of the 'evaluation indicators of the inquiry experiment class' (Table 5-4). This revealed that the evaluation of the lower section of the inquiry experiment class should favour the indicators of scientific thinking and practice of inquiry, particularly within the 'scientific concept' under 'scientific concepts' in the secondary indicators. It is notable that the secondary indicator 'scientific knowledge' under 'scientific concepts' accounts for a larger proportion of the combined weight.

Table 5-4: Comprehensive Weight Results for Evaluation Indicators of Inquiry Experiment Classes for Lower Grade Levels

Primary Dimension	Weight	Secondary Dimension	Secondary Weight	Composite Weight
Scientific Concepts	0.2008	Scientific Knowledge	0.6734	0.1236
		Applied Interpretation	0.3266	0.0600
Scientific Thinking	0.3160	Analytical Ability	0.2665	0.0842
		Comparison & Classification	0.3121	0.0986
		Abstraction & Generalization (Modeling)	0.1349	0.0426
		Induction & Association	0.1454	0.0459
		Innovative Thinking	0.1411	0.0446
Inquiry Practice	0.3420	Question Formulation	0.1537	0.0526
		Hypothesis Development	0.1381	0.0472
		Experimental Methods	0.1271	0.0435
		Planning	0.0924	0.0316
		Investigation Design	0.0895	0.0306
		Evidence Collection	0.0782	0.0267
		Evidence Analysis	0.0816	0.0279
		Communication & Presentation	0.0784	0.0268
Attitudinal Responsibility	0.1412	Evaluation & Reflection	0.0905	0.0310
		Learning Enthusiasm	0.3176	0.0448
		Intellectual Integrity	0.3634	0.0513
		Team Collaboration	0.1265	0.0179
		Social Responsibility	0.1925	0.0272

5.3.2 Analysis of Weight Results for Middle Grade Levels

The stochastic consistency ratio (CR) of the weights of the first-level indicators in the middle section of the document entitled "Evaluation indicators for inquiry laboratory lessons" is 0.0419. The CR of the weights of the second-level indicators under the heading "Scientific concepts" is 0.0414. The CR of the weights of the second-level indicators under the heading "Scientific thinking" is 0.0150. The CR of the weights of the second-level indicators under the heading "Practice of inquiry" is also 0.0150. The random consistency ratio (CR) for the weights of the secondary indicator under the heading of 'Scientific thinking' is 0.0150, for the weights of the secondary indicator under the heading of 'Practice of inquiry' is 0.0230, and for the weights of the secondary indicator

under the heading of 'Attitude and responsibility' is 0.0419, and for the weights of the secondary indicator under the heading of 'Science perception' is 0.0414. The CR of the weights of the secondary indicators under 'inquiry and practice' is 0.0150, the CR of the weights of the secondary indicators under 'attitude and responsibility' is 0.0230, and the CR of the weights of the secondary indicators under 'attitude and responsibility' is 0.0189. All of these values are less than 0.1, indicating satisfactory consistency in the results of the hierarchical analyses. This implies that the distribution of weighting factors is reasonable. The final step was to multiply the results of the weights of the primary indicators and the weights of the secondary indicators in order to calculate the combined weights of the 'evaluation indicators of the inquiry experiment class' in the lower section (Table 5-5). From the table, we can see that in the same way, the middle section of

the evaluation in the inquiry experiment class still favours scientific thinking and inquiry practice indicators, and in the secondary indicators, it should be

particularly pointed out that 'analytical ability' under the indicator of 'scientific thinking' accounts for the largest proportion of the overall weight.

Table 5-5: Comprehensive Weight Results for Evaluation Indicators of Inquiry Experiment Classes for Middle Grade Levels

Primary Dimension	Weight	Secondary Dimension	Secondary Weight	Composite Weight
Scientific Concepts	0.1836	Scientific Knowledge	0.4129	0.0758
		Nature of Science	0.3763	0.0691
		Application & Interpretation	0.2108	0.0387
Scientific Thinking	0.3566	Analytical Ability	0.2858	0.1019
		Comparison & Classification	0.2281	0.0813
		Abstraction & Generalization (Modeling)	0.1685	0.0601
		Induction & Association	0.1563	0.0557
		Innovative Thinking	0.1613	0.0575
Inquiry Practice	0.2945	Question Formulation	0.1108	0.0326
		Hypothesis Development	0.1018	0.0300
		Experimental Methods	0.1144	0.0337
		Planning	0.1109	0.0327
		Investigation Design	0.0953	0.0281
		Evidence Collection	0.0924	0.0272
		Evidence Analysis	0.1014	0.0299
		Communication & Presentation	0.0753	0.0222
Attitudinal Responsibility	0.1653	Evaluation & Reflection	0.1091	0.0321
		Learning Enthusiasm	0.2628	0.0434
		Intellectual Integrity	0.3025	0.0500
		Team Collaboration	0.2578	0.0426
		Social Responsibility	0.1769	0.0292

5.3.3. Analysis of Weight Results for High Grade Levels

The stochastic consistency ratio (CR) of the weights of the first-level indicators in the upper middle level of the 'Evaluation indicators for inquiry laboratory lessons' is 0.0061. The CR of the weights of the second-level indicators under 'Scientific concepts' is 0.0206. The CR of the weights of the second-level indicators under 'Scientific thinking' is 0.0102. The CR of the weights of the second-level indicators under 'Inquiry practices' is 0.0206. The random consistency ratio (CR) of the weights of the secondary indicator

under the heading of 'Scientific thinking' is 0.0102, that of the weights of the secondary indicator under the heading of 'Practice of inquiry' is 0.0188, and that of the weights of the secondary indicator under the heading of 'Attitude and responsibility' is 0.0061. The random consistency ratio (CR) of the weights of the secondary indicators under 'inquiry and practice' is 0.0188, while the random consistency ratio (CR) of the weights of the secondary indicators under 'attitude and responsibility' is 0.0115. In both cases, the values are lower than 0.1, indicating satisfactory consistency in the results of the hierarchical analyses. This implies that the distribution of the weight coefficients is reasonable. The final step

was to multiply the results of the weights of the primary indicators and the weights of the secondary indicators in order to calculate the comprehensive weights of the 'evaluation indicators of the inquiry experiment class' in the lower section (Table 5-6). From the table, we can see that the distribution of weights in the high and middle sections of the evaluation in the inquiry experiment class is similar, not only in the primary

indicators favouring scientific thinking and inquiry practice, but also in the secondary indicators, 'analytical ability' under the indicator of 'scientific thinking' is still the most important indicator in the overall weight. The largest share of the combined weights is still given to 'analytical skills' under 'scientific thinking' in the secondary indicators.

Table 5-6: Comprehensive Weight Results for Evaluation Indicators of Inquiry Experiment Classes for High Grade Levels

Primary Dimension	Weight	Secondary Dimension	Secondary Weight	Composite Weight
Scientific Concepts	0.2093	Scientific Knowledge	0.2607	0.0546
		Nature of Science	0.4422	0.0926
		Application & Interpretation	0.2971	0.0622
Scientific Thinking	0.3704	Analytical Ability	0.2831	0.1049
		Comparison & Classification	0.1503	0.0557
		Abstraction & Generalization (Modeling)	0.2082	0.0771
		Induction & Association	0.1935	0.0717
		Innovative Thinking	0.1649	0.0611
Inquiry Practice	0.2573	Question Formulation	0.0881	0.0227
		Hypothesis Development	0.0880	0.0226
		Experimental Methods	0.1117	0.0287
		Planning	0.1152	0.0296
		Investigation Design	0.0907	0.0233
		Evidence Collection	0.0941	0.0242
		Evidence Analysis	0.1084	0.0279
		Communication & Presentation	0.1062	0.0273
Attitudinal Responsibility	0.1631	Evaluation & Reflection	0.1002	0.0258
		Learning Enthusiasm	0.2088	0.0341
		Intellectual Integrity	0.2551	0.0416
		Team Collaboration	0.332	0.0541
		Social Responsibility	0.2042	0.0333

6. Limitations

It must be acknowledged that the study is not without shortcomings, which are a consequence of the limitations of human, material and time resources, as well as the researcher's own level. In order to guarantee the scientific and operational efficacy of the evaluation index system for science inquiry classes in primary schools in the context of the new curriculum

standard, two rounds of the Delphi method were employed to determine the evaluation index system. Furthermore, the hierarchical analysis method was utilised to ascertain the index weights based on the results of the hierarchical analysis of consistency. Nevertheless, the evaluated index system has yet to be practically implemented in primary schools, and further issues may emerge during the actual application process.

7. Conclusion

This paper presents an initial refinement of the evaluation index system for primary school students' inquiry classes, based on an analysis of the current status quo of evaluation practices. Furthermore, the Delphi method was employed to enhance the content validity of the indicators within the system. This involved the preparation of two rounds of questionnaires, the objective of which was to solicit the opinions of experts on the importance and operability of the indicators, and to determine the evaluation indicator system for science inquiry classes in primary schools, taking into account the new standards. Subsequently, the hierarchical analysis method was employed once more to construct a judgement matrix, which was then issued to the panel of experts. This was followed by the distribution of questionnaires to determine the weight of the indicators within the indicator system, as well as a consistency test. The results demonstrated satisfactory consistency in the hierarchical analysis, which in turn provided a solid foundation for the evaluation of the evaluation indicators of each type of class, with each indicator being clearly weighted. The construction of the evaluation index system and the determination of its constituent indicators and weights provide a theoretical basis for the reform of primary science education. This will facilitate the optimisation of the educational environment and assist educators in more accurately comprehending the objectives of training in accordance with the revised science standards, thereby enabling the development of corresponding teaching strategies and evaluation methods.

Funding

This work was supported by the Zhejiang Social Science Federation Subject (2023N079), Department of Education of Zhejiang Province(Y202351624), and Chinese Society of Educational Development Strategy (CEE202308).

References

- Huang, X. (2021). Aims for cultivating students' key competencies based on artificial intelligence education in China. *Education and Information Technologies*, 1-21.
- Weber, R. (2021). CONSTRUCTS AND INDICATORS: AN ONTOLOGICAL ANALYSIS. *Mis Quarterly*, 45(4), 1644-1678. <https://doi.org/10.25300/misq/2021/15999>
- Chen, X. S. (2001). Exploring scientific attitudes and their evaluation through secondary physics education (Part 1). *Subject Education*, (6), 47-49. https://kns-cnki-net-s.webvpn.wzu.edu.cn/kcms2/article/abstract?v=xpM8-w1VMS86k-h1iIlMdCPMD8xqjsri_3cbH5uI20JfkLISp6rWN_K97uAkG5g6XCsySY9k9JA2RGVPuOK03nc0KmYk2jm928Xp3sLciTiL0AgD6Cqh7LzF9aNzI8-o7IDBCMYTQQoCFL_t1wDtcqclIKDh5bli72oDNgKZVS3Cet2jAwgZ76FNTvpVs6La&uniplatform=NZKPT&language=CHS
- Ding, B. P. (2001). Constructivism and science education reform for the 21st century. *Comparative Education Review*, (8), 6-10. https://kns.cnki.net/kcms2/article/abstract?v=xpM8-w1VMS81dOICcjPqpG37cfa4OEoOLz46YTCBKr1DxbPTmpHpzt0ZHF-FLmSox4ccacjNALJOXu1aY56FAr1Zy5ToRHXzX_wKeFz-Ult0nSL44YSUAdXFZK74nfB8HlrR0V5Hnv6UylgSOL3oAFJX0oBQjvKWwyKpslGg9ycQ2_vyt6rCk3kO_5nNJbgW&uniplatform=NZKPT&language=CHS
- Du, X. L. (2013). The vibrant methodology of literature research. *Shanghai Journal of Educational Research*, (10), 1. <https://doi.org/10.16194/j.cnki.31-1059/g4.2013.10.002>
- Fang, J. H. (2014). Research on the evaluation and standard system construction of vocational core competencies for secondary vocational students [Doctoral dissertation, Nanjing Normal University]. CNKI. <https://kns.cnki.net/kcms2/article/abstract?v=sxrP1m9hSI8-K9kITdBQxGMJQ1c-7BZasdBaPLz9iUixy7oFe79y9sxxgQ0wPN2vgZr4>

- GFv77RnaJaYAM5-
WCHafIjwzUW9LQLiAdnxU5VzoB0DezZqnfPdZd
ZsbJW5II2gC1_sY-
lkjl2kVJdcotjCZGWJU9qLPrdWgpciXN7KfCxxHQ
8chdpUmSz9TPEIPfuRYwTsc_Zul=&uniplatform
=NZKPT&language=CHS
- Guo, J. W., Pu, X. Q., Gao, X., & Zhang, Y. A. (2014). An improved calculation method for multi-objective decision-making indicator weights. *Journal of Xidian University*, 41(6), 118-125. <https://doi.org/10.13884/j.1003-3807hxjy.2023060168>
- Hu, J. H., & Huang, D. F. (2024). Designing unit teaching objectives oriented by core competencies based on chemistry curriculum standards. *Chinese Journal of Chemical Education*, 45(13), 1-7. <https://doi.org/10.13884/j.1003-3807hxjy.2023060168>
- Liu, X. C. (2021). Research on the construction of a teaching evaluation index system for basketball courses in Jiangsu universities based on core competencies [Master's thesis, Nanjing Sports Institute]. CNKI. <https://link-cnki-net-s.webvpn.wzu.edu.cn/doi/10.27245/d.cnki.gnjjsu.2021.003257>
- Su, W. H. (2000). Research on theoretical and methodological issues in multi-indicator comprehensive evaluation [Doctoral dissertation, Zhejiang Gongshang University]. CNKI. <https://kns.cnki.net/kcms2/article/abstract?v=8pq0kR8SZyVeyi6SFQUJMjq27zfoR0EPSHFauUHmNC2wlhEU--EpgdD1oXusEDE-qZG7yaQH2Bd0InmvKFWfH6EGMUtOdwgErH8vB66RNnE8c-0GvLbGMltz37oDsXcfyQndV5WVufsqDu6IAV64BfdBLD9PWvlv432lvTlaJRqp5N5TkZoPepO8HzoHuW51r7sSaGXWiKc=&uniplatform=NZKPT&language=CHS>
- Wang, C. Z., & Siqin. (2011). Statistical processing methods in Delphi technique and their application research. *Journal of Inner Mongolia University of Finance and Economics (Comprehensive Edition)*, 9(4), 92-96. <https://doi.org/10.13895/j.cnki.jimufe.2011.04.019>
- Wang, Q. Z. (2014). Scientific evaluation for holistic development: A research report on "standards-based student academic assessment". *Basic Education Forum*, (12), 38-41. https://kns-cnki-net-s.webvpn.wzu.edu.cn/kcms2/article/abstract?v=691tpyMQYm0IIEH2eRon46jhL6OWPqNj9E3DyCkTetbQWvvoBSQwpl1m0BN5rCZM971nOzEYQ6D4ULbYAIvchQgZ7dlx0tjH9W84UZa2ZUBrlm-L9lzE6SaKydrR9ABpjiJ_Sc64KOHmjXnUFG3pkgvuUASiWqf-5PyR2oWhrlyyAJ7BoHIZKr33o3Fe9MXgbiwry0iknw=&uniplatform=NZKPT&language=CHS
- Wei, Y. (2016). Implementing science education through the lens of big ideas. *People's Education*, (1), 41-45. <https://kns.cnki.net/kcms2/article/abstract?v=xpM8-w1VMS9Juxd2mCUQXcsLYlulycTU7BdmtHI47S3vqqj3ohQNElwzb5zhO3ZyoVv2R9TbIBmlRxxW5Zwk2gRx6yOtNQStj8F4FI77QdN1OzLYQ-A3V20DNbmY6FzsE3f4Yzako0ljkNi38OsQLSWnfQWe-r4vstNIYdlCPjFssGJEOCc0ADepubG5AtqlfJT9JtbRAt0=&uniplatform=NZKPT&language=CHS>
- Wiersma, W., & Jurs, S. G. (1997). *Educational research methods: An introduction (7th ed.)*. Allyn & Bacon.
- Xin, T. (2006). The value of measurement theory in academic assessment under the new curriculum. *Journal of Beijing Normal University (Social Sciences)*, (1), 56-61. https://kns.cnki.net/kcms2/article/abstract?v=yQB21MkjwM_JJ4MA9QSVvk93ZI-KEtxCmFRhSKRdZeWjZ1iKO7zesULSm2GN-C1tXnSp5tgSYGPp8kinB0CLQ_VoOuOrgNLIHjsYUDFeni8SbOJdhrOk3uCeKrv5hnjguKUfmlvIXCkwwTW0EKmLxJLJA98oPFGeZXOyvj8kA3byXBbYdXWjVM2E5pt9okEp&uniplatform=NZKPT&language=CHS
- Yang, S. Z. (2002). Green education: The integration of scientific and humanistic education. *Educational Research*, (11), 12-16. https://kns.cnki.net/kcms2/article/abstract?v=xpM8-w1VMS9ryj3LW39C7JPQzbHx62HkUsfE9jSAPBvBdCDMA4hDBYvqDxy9t4wC6AkfXOrJyzgKnY_E295QSoU5OX1vPpcpyFxxzrkHMRfjUf7lylxCn861AE8RR6OV1nYhkmizvB6OoMkq8MY-

- apsSezb6KRmE50JiH24ZKPIltoZEqessGort9x
F1rm&uniplatform=NZKPT&language=CHS
Zhang, D., Wu, H. X., & Zhang, D. (2007). Research on the comprehensive evaluation index system of information literacy for Chinese university students. *Information Studies: Theory & Application*, (1), 56-60. <https://doi.org/10.16353/j.cnki.1000-7490.2007.01.016>
- Zhou, L. L. (2011). A brief discussion on teaching evaluation of information technology under the new curriculum concept. *Science and Technology Innovation Herald*, (30), 202. <https://doi.org/10.16660/j.cnki.1674-098x.2011.30.149>

Cite this Article: Chenhui, G; Sudan, J; Saiqi, T (2025). Research on the construction of evaluation index system of science inquiry class for primary school students. *Greener Journal of Education and Training Studies*, 8(1), 1-21, <https://doi.org/10.15580/gjets.2025.1.031725043>.