Journal of Science Education Vol.23 n◦ 2 2022

Journal of Science Education

ISSN 0124-5481

The evaluation and performance of primary school students ' scientific thinking

Ziwei Liu[a], Yuanyuan Shang[b], Baishun Du[c], Zhibo Jia[d], Jingying Wang[a*]

[a] *Faculty of Education, Beijing Normal University, Beijing 100875, P.R. China*
[b] *Faculty of Education, Beijing Normal University, Beijing 100875, P.R. China*
[c] *Jinzhou Teachers ' Training College of Liaoning Province, Jinzhou 121000, P.R. China*
[d] *Beijing Bayi School, Beijing 100080, P.R. China*

A R T I C L E I N F O

Keywords:
Scientific thinking
evaluation research
primary school science
rasch model competence

A B S T R A C T

Scientific thinking is an important part of core literacy, which cannot be defined only by simple thinking mode stratification or the superposition of various abilities. As a multi-dimensional interactive dynamic system, the executive ability of scientific thinking reflects the students ' thinking process, and the evaluation of the process is more suitable for the students ' scientific thinking level. This study developed a primary school scientific thinking ability assessment tool and evaluated the 6th grade students in Jinzhou primary school. The quality of the test tool was verified by Winstepts software. The results show that the quality of the evaluation questions is good, which can accurately evaluate the students ' scientific higher-order ability level, and preliminarily analyze the overall scientific thinking level, gender differences and regional differences of the evaluation objects. .

1. Research background

With the continuous development of science education, educational policies and educational practice programs formulated for the purpose of cultivating students ' scientific higher-order thinking are constantly emerging. The first set of " National Science Education Standards " promulgated in 1996 in the United States began to emphasize that science courses should train students to use reasoning and scientific evidence and improve the ability of scientific interpretation and models. The " Next Generation Science Standards " more clearly defined the thinking-oriented teaching of science courses, and incorporated eight thinking skills, such as critical thinking, computational thinking, proportional reasoning, qualitative and quantitative thinking, into the curriculum objectives（NRC, 2013）. The British ' Science in the National Curriculum ' (SNC) is divided into four key stages. The teaching objectives of each stage are mainly focused on the ability of students to learn to observe experimental phenomena and use simple scientific language (charts, symbols) to record experimental results, data collection and identification and classification, oral report and written interpretation of the results of inquiry, and the ability to make scientific explanations and understandings using various generalizations and models. Israel 's Ministry of Education has always regarded scientific inquiry as one of the main objectives of thinking teaching in science education since the promulgation of the " teaching horizon-thinking teaching " policy in 2006. In Asian countries such as Singapore and Japan, the key stages of thinking skills teaching have also been set up to encourage primary school students to develop skills such as analysis, formation of personal views, evaluation, investigation and creative problem solving through scientific learning. The teaching of scientific higher-order thinking puts forward common requirements for improving students ' comprehensive ability in scientific reasoning, scientific modeling, written and oral argumentation. China 's latest version of the science curriculum standards also clearly pointed out that the science curriculum is practical, emphasizing the need to focus on cultivating students ' scientific thinking ability, scientific inquiry and practical ability. In China, scientific thinking is considered to be the key trait of top-notch talents in science and technology, which is related to the effectiveness of high-quality science education and the realization of the strategy of reinvigorating China through human resource development. The 20 th National Congress of the Communist Party of China also coordinated the deployment of science and technology, education and talents. It is believed that the cultivation of high-level scientific thinking is the breakthrough strategy for the cultivation of top-notch talents in science and technology in China(Wang J Y, Zhou D H, Yang Y, Ke L & Tian X W，2023).

Corresponding author: Jingying Wang
Email: wangjingying8018@126.com
Accepted 1 October 2022, Available online 10 July 2023

2.Evaluation methods

2.1The connotation and evaluation of scientific thinking

In the initial stage of higher-order thinking research, scholars generally believe that higher-order thinking is a relative concept. Researchers define higher-order thinking by comparing it with lower-order thinking. The distinction between low-order and high-order thinking was first proposed by American cognitive psychologist Bloom. In Bloom 's classification of educational goals, low-order thinking refers to the simple memory and retelling of knowledge, while high-order thinking refers to the psychological process of organizing or reorganizing knowledge to achieve a certain goal（Bloom & Krathwohl, 1956）. Bloom also proposed five different types of thinking modes : comprehension, application, analysis, synthesis, and evaluation. Among them, ' analysis ', ' synthesis ', and ' evaluation ' are attributed to higher-order thinking. With the in-depth study of higher-order thinking, researchers have defined the concept or summarized the elements of higher-order thinking from different perspectives. At this stage, although the definition of higher-order scientific thinking ability is not uniform, almost all researchers agree that higher-order scientific thinking should be a system that includes multiple abilities, including critical thinking, problem solving, decision making and creative thinking（Lewis & Smith，1993）. The OECD regards creative thinking, critical thinking, problem-solving ability and decision-making as the top ten core skills of talents in the 21 st century. In the report of " Chinese students ' development of core literacy, " eighteen basic elements, such as critical questioning, courage to explore, problem solving and innovative spirit, are included in the elements of Chinese students ' development of core literacy. Lai & Hwang (2014) concluded that higher-order thinking ability should include problem solving, critical thinking, teamwork, communication and creative thinking based on the core skills of talents in the 21 st century. Zhong（2004） proposed that higher-order thinking should include critical thinking, problem solving, innovation, acquisition of tacit knowledge, self-management, sustainable development, information literacy, teamwork, compatibility, and decision-making. Among the latest research results in the field of higher-order thinking, Professor Wang Jingying and her team summarized the evolution characteristics of higher-order thinking connotation into four categories : hierarchical advanced theory from the perspective of educational objectives, comprehensive element theory from the perspective of ability evaluation, system structure theory from the perspective of curriculum structure and learning process theory from the perspective of instructional design. In the research results, scientific higher-order thinking is considered to be neither a kind of thinking nor a simple superposition of multiple thinking, but a dynamic interactive comprehensive system. Based on this, a system model of higher-order thinking is constructed. The model takes general higher-order thinking such as problem solving and creative thinking as the basic ' skeleton ', and its essence is reflected in the cognitive process of students ' creative solution to scientific situational problems. Scientific reasoning, scientific argumentation and scientific modeling are the core skills of the field class. As the main ' executors ' of thinking activities, they share the function of thinking practice in the process of scientific learning. At the same time, this process requires the participation of reflection and evaluation behaviors such as metacognition and critical thinking ; in addition, scientific attitude, learning motivation and self-efficacy regulate and affect the operation of scientific higher-order thinking system (Wang et al，2023).

At present, there are few studies on the evaluation of higher-order thinking, most of which are based on Bloom 's target classification method, positioning analysis, synthesis and evaluation as higher-order thinking and judging students ' higher-order thinking development level based on this. There are also a few who define higher-order thinking as a collection of multiple abilities, and measure students ' higher-order thinking level by the sum of the abilities of each part. Most of the evaluation problems rely on subject knowledge, such as biological, physical and chemical problem situations, but less on the subject content of primary school science. Today, with the increasing importance of science education and higher-order thinking, it is increasingly important to evaluate the higher-order thinking of adolescents, especially primary school students, and to explore training strategies.

In view of this, based on the dynamic system model of higher-order thinking constructed by Professor Wang Jingying and her team, this study takes the three core competencies of scientific reasoning, scientific argumentation and scientific modeling, the main ' executors ' of thinking activities in the model, as the three sub-skills of scientific higher-order thinking evaluation, and deconstructs the thinking process of the three skills into three dimensions with the explicit and measurable principles, and develops a scientific higher-order thinking evaluation tool including nine dimensions. The evaluation tool takes the six-year scientific knowledge of primary school as the carrier, and integrates all dimensions of higher-order thinking into the same evaluation question, which can better measure the level of higher-order thinking based on the existing cognition of primary school students.

2.2 The sub-dimension of scientific thinking assessment

2.2.1 Connotation and framework of scientific reasoning ability

Scientific reasoning originated from the "formal reasoning" in Piaget's theory of cognitive development, and Inhelder & Piaget（1958） believed that only when they developed into formal operations would children develop scientific thinking. After that, researchers described the concept of scientific reasoning differently from different directions. Reasoning, in logic, refers to one of the basic forms of thinking, the process of deriving new judgments (conclusions) from one or several known judgments (premises). Scientific reasoning is developed on the basis of reasoning, and its essence is considered to be a set of thinking skills or a cognitive process. For example, Lawson（1978） proposed that scientific reasoning is the process by which individuals find reasons for complex phenomena observed, and that individuals support or reject the null hypothesis by finding and evaluating evidence. Kuhn(2002) defined broad scientific reasoning as the deliberate search for knowledge and the coordination of theory and evidence, a process that includes the ability to acquire and transform knowledge "to generate, test, and revise theories and hypotheses, and to reflect on the reasoning process." Zimmerman(2007) argues that scientific reasoning competence involves a range of skills in scientific practice, including or "scientific inquiry, experimentation, evidence evaluation, and reasoning in the service of conceptual change or scientific understanding". Stuessy(2008) believes that scientific reasoning ability is the internal logical thinking form that individuals need to have in the process of scientific inquiry to observe the relationship between things, make hypotheses, design experiments to test hypotheses, predict the probability of results, give logical inferences, evaluate evidence, and finally prove the rationality of specific conclusions .

2.2.2 The connotation and framework of scientific argumentation

Arguments originated in logic, which traditionally believes that arguments are based on the "premise-conclusion" structure of reasoning, which is the external manifestation of reasoning（Xie，2007）. Turmin（2003） believed that argumentation is to use data, reasons, claims and other effective claims to persuade others in the face of complex scientific and social problems, construct their own scientific arguments, and criticize formal and mathematical logical argumentation theories, and put forward a high-level argumentation model containing six elements of "data, support, reason, qualifier, refutation conditions, and claims". Argumentation is considered to be a process aimed at legitimizing and publicizing scientific knowledge by establishing a link between scientists' imagination, speculation, and available evidence（Tang，2016）. For example, Van et al.（2002） see scientific argument as the process of justifying or refuting opposing views; Zohar & Nemet（2002）believe that scientific argument is the process of making a claim or conclusion and defending or justifying it accordingly; Lawson（2003）defines scientific argument as the process of constructing claims in uncertain questions and establishing a connection between claims and evidence . When discussing the nature of scientific argumentation, Hu（2021）defined scientific argument as an activity, believing that scientific argument is based on data, relying on the basis to support one's own claims, refute the views of others, and form a final conclusion through logical reasoning .

Scientific argumentation ability is the ability required to carry out scientific argumentation activities. From the level of the elements of scientific argumentation, Dawson & Carson (2017) proposed that scientific argument ability includes three parts: the ability to put forward arguments, the ability to form counterarguments, and the ability to put forward rebuttals. Kuhn（1992）believes that the scientific argument capability package uses correct and appropriate evidence to support the proposed theory, strengthens its own theory with other theories, refutes the arguments of others, and refutes the arguments of others, and refutes the arguments of others in four parts. Chen & Guo（2017） believe that scientific argumentation ability is the advanced thinking ability of individuals to use scientific knowledge as a mediator, make claims and reasoning based on the collected data, reflect on the inadequacy of their own and others' arguments, and at the same time refute the doubts of others and criticize the advanced thinking ability to defend themselves. Han（2016）defines the structure of the scientific argumentation ability model according to the connotation of scientific argumentation, and proposes that scientific argumentation ability is a high-level scientific thinking ability mediated by scientific concepts, with scientific reasoning and critical thinking ability as the core, and including scientific thinking quality. Zhang & Browne（2023）believe that the scientific argument capability framework should include the identification of scientific arguments, the evaluation of scientific arguments, and the products of scientific arguments, and deconstruct the above parts into four different elements: opinion, evidence, reasoning, and refutation .

2.2.3 Connotation and framework of scientific modeling capability

Early modeling teaching research scholars mainly analyzed the modeling process from the characteristics of scientists ' empirical thinking, and advocated ' analyzing consciousness and experience as many minimum basic elements（Chen & Liu，2017）. For example, Hestenes（1992） believed that scientific modeling is the process of applying certain design principles (scientific theories) to generate object models or natural process models. Modeling itself is a kind of ' process knowledge ', and modeling ability is mainly manifested in the ability to use the four modeling elements of description, conception, derivation and verification. Halloun（1996） further expanded the scientific model horizontally into an organism containing elements, structures, domains and organizations, and considered the selection, construction, verification, analysis and utilization of the model as five important elements in the process of problem solving. Later, some scholars believe that scientific modeling is not a simple process of model representation, but a set of processes of information receiving, processing, extraction and output. Scientific modeling ability is the embodiment of its internal level of mental and behavioral activities. For example, Stratford et al.（1998） proposed a cognitive strategy framework from the specific representation process of modeling, which includes five dimensions: analysis, reasoning, synthesis, testing and interpretation.

A Taiwan scholar first constructed a complete scientific modeling ability model from the three dimensions of ontology, epistemology and methodology, and divided the scientific modeling ability of the methodological dimension into problem solving ability, ability to understand and observe phenomena, and ability to connect and develop ideas（Qiu，2008）. Based on the process of the model, Hung & Lin(2009) divided it into five abilities : model selection, construction, verification, analysis and application, and 19 sub-dimensions. The scientific modeling ability model proposed by Schwarz （2005） divides the scientific modeling ability into two parts : modeling practice ability and modeling meta-knowledge. The modeling practice ability includes four sub-dimensions : model creation, model application, model comparison and model correction. Meta-knowledge consists of metacognitive knowledge and meta-modeling knowledge in the modeling process. Modeling meta-knowledge includes model essence, modeling essence, modeling purpose, modeling utility and modeling evaluation. Enright et al. (2020) divides the structure of scientific modeling ability into two parts : the ability to generate models and the ability to change models. The ability to generate models focuses on the text and characteristics of information and the characteristics and commonalities of models. The ability to change models focuses on the rationalization of models and the application of models.

3.Development of assessment tools

By sorting out and summarizing the framework of scientific reasoning ability, scientific argumentation ability and scientific modeling ability, this study divides scientific reasoning ability into three sub-dimensions : ' raising questions and making hypothesis design, experiment and generating data, explaining data and drawing conclusions '. The scientific argumentation ability is divided into three sub-dimensions : ' viewpoint, fact and theoretical basis, reasoning and refutation '. The scientific modeling ability is divided into three sub-dimensions : ' model construction and use, model verification and correction, modeling metacognition and meta-modeling knowledge ' (see figure 1).

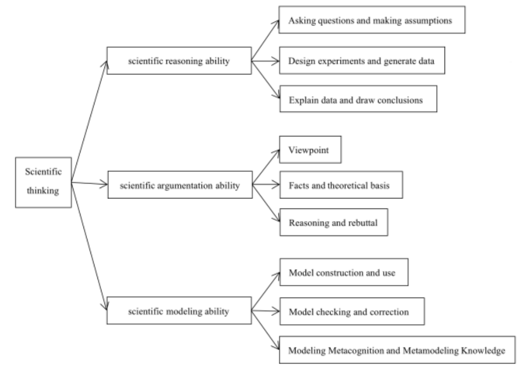

Figure. 1 Evaluation framework of scientific thinking ability

The problems in the evaluation rely on the scientific knowledge of primary schools as the carrier. After the evaluation framework is established. Based on the official websites of the four major examination bureaus in the United Kingdom : AQA (The Assessment and Qualifications Alliance), OCR (Oxford Cambridge and RSA Examinations), CIE (Cambridge International Examinations) and Edexcel (Edexcel Pearson- London Examinations), as well as the official websites of education in 50 states in the United States, the final examination and periodic examination papers of grades 4-6 in the past four years can be collected. The collected test papers are classified and summarized according to the knowledge points investigated by the questions, and the contents of the more frequent college entrance examination are selected to compare with the contents of the six-year science curriculum standards of domestic primary schools, and the overlapping questions of knowledge points are screened. After four rounds of discussion among the research group, the expert group and the first-line teachers, the topic of the questionnaire test is ' the movement of the earth '. Then, two questions with situational creation and the best ability to examine students ' higher-order thinking ability were selected from the examination questions of ' The Movement of the Earth ', which were the 21st question of the 2018 final test volume of Texas 5th grade science and the 26 th question of the 2022 final test volume.The first question in the scientific test part of the questionnaire was adapted from the 21 st question of the 2018 test volume, and the second and third questions were adapted from the 26 th question of the 2022 test volume.

Scientific reasoning, argumentation and modeling are the domain attribute skills of scientific higher-order thinking, and they are also the main ' executors ' of scientific higher-order thinking practice. These three functions are described as the means of logical reasoning, the tools of teaching reform and the components of practical activities. In the process of scientific learning, they are not independent of each other and do not interfere with each other, but a circular and interactive whole (Wang et al.,2023). The teaching practice of scientific thinking originated from the exploration of scientific reasoning. Accordingly, the evaluation tools developed in this study are based on the idea of comprehensively mobilizing students ' scientific reasoning, scientific argumentation and scientific modeling, and integrate the evaluation of the three abilities into the same test questions. And for the three sub-skills and nine sub-dimensions of each question, the gauge is set according to the three-level division standard.

2.4 Research object

This evaluation combines cluster sampling and random sampling to select primary school students in grade 6 in Jinzhou City, Liaoning Province. They are from Beizhen City, High-tech Zone (Songshan New District), Guta District, Heishan County, Linghai City, Linghe District, Taihe District, Yixian eight experimental areas. Table 1 details the total number of samples for each experimental area. After removing the invalid questionnaire, there were 4415 primary school students, including 2171 boys, accounting for 49.2 % of the total sample size ; a total of 2243 girls, accounting for 50.8 % of the total sample size, the proportion of male and female students.

Table 1 Composition of primary school science higher-order thinking evaluation samples

School location	Beizhen	Gaoxin	Guta	Heishan	Linghai	Linghe	Taihe	Yixian	total
Effective questionnaire	707	226	408	949	843	531	236	515	4415
Percentage	16.0%	5.1%	9.2%	21.5%	19.1%	12.0%	5.3%	11.7%	100.0%

3.Research results

3.1 Tool quality analysis

3.1.1 Overall quality

In order to verify the quality of the test questions, this study is based on the existing research of the Rasch model. Winsteps3.72.0 software is used to analyze the quality of the five aspects of the overall quality, one-dimensionality, item fitting and rating structure of the test tools. The results are shown in Table 1. In addition, through SPSS software, the Cronbach 's Alpha value of this test tool is 0.793, which is between 0.60 and 0.80, and the reliability coefficient is good, indicating that the statistical results have good credibility and the statistical results are effective.

In the Rasch model, the average difficulty estimate of the Item is set to 0, so the estimate of the Person Measure is actually the average ability value (literacy level) of the subject. As shown in the data in Table 1, the scientific ability of the subjects in this study was -0.58, lower than the difficulty value of the item, indicating that the assessment tool was generally difficult for the subjects, but the gap was not huge. It shows that the evaluation items fit the literacy level of the subjects well and are suitable for the evaluation of the theoretical samples. Error represents the difference between the theoretical model and the actual observed value. The errors of subjects and items are 0.36 and 0.03, respectively, which are close to 0, indicating that the observed value obtained by the preliminary test tool can more truly reflect the scientific ability of subjects. Infit and Outfit represent the fit degree between theoretical model and actual observed values, including MNSQ and ZSTD. According to the ideal parameter values, the ideal value of MNSQ is close to 1, and the ideal value of ZSTD is close to 0. The data in Table 1 shows that the MNSQ and ZSTD of subjects and projects are both very ideal. It shows that the observed values of the evaluation tool are in good agreement with the theoretical model of Rasch. Separation refers to the degree to which the evaluation tool distinguishes the literacy level of the subjects. The ideal separation degree should be greater than 2, and the greater the value, the better. In this study, the separation degree of the subjects (3.31) and the items (13.41) are much higher than 2, indicating that the evaluation items can meet the literacy level of different subjects. The higher the Reliability, the lower the error value of the measurement. The reliability of the subject and the reliability of the item are 0.92 and 0.99 respectively, and the data results are good. The results of the above parameters show that the overall characteristics of the evaluation tool are good.

Table 2 Overall quality analysis results of assessment tools

	Measure	Error	Infit		Outfit		Separation	Reliability
			MNSQ	ZSTD	MNSQ	ZSTD		
Person	-0.58	0.36	0.99	0.0	1.05	0.0	3.31	0.92
Item	0.00	0.03	1.00	-1.8	1.05	-.2.2	13.41	0.99

3.1.2 Unidimensional analysis of test questions

dimensionality is one of the basic assumptions of the Rasch model, that is, each evaluation item in the evaluation measures the same underlying quality. The Rasch model uses principal component analysis to test the unidimensionality of test items, and Standardized Residual Contrast Plot presents the relationship between the Loading coefficient of evaluated items and the estimated Item measure of difficulty. As shown in Figure 3, the upper and lower case letters in the figure represent 27 evaluation items respectively, the horizontalcoordinate represents the project difficulty, and the vertical coordinate shows the project load coefficient, the ideal value is between -0.4 and +0.4, if it exceeds the range, it is considered that it does not meet the one-dimensional requirements. It can be seen from the figure that f (subject BG) and c (subject BB) in this test question slightly deviate from the ideal range. The load coefficient of most projects falls between -0.4 and +0.4, which can be considered to meet the requirements of unidimensional. Therefore, on the whole, most of the questions in this assessment tool are measuring students' scientific thinking, and this

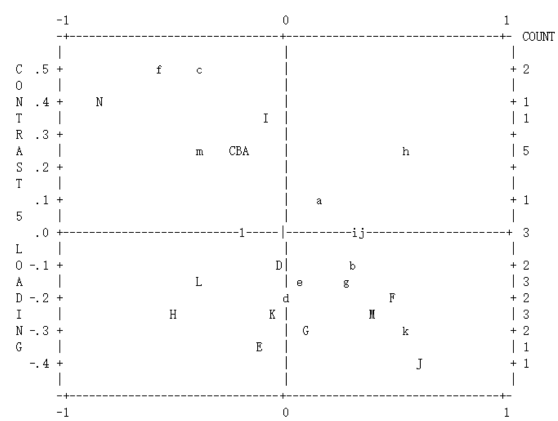

Figure 2 Unidimensional test questions

3.2.3 Test question grading structure detection

In order to test the integrity of primary school students' scientific thinking, this study developed three sub-skills of scientific reasoning ability, scientific argumentation ability and scientific modeling ability for each essay question, and each sub-skill was evaluated from three different dimensions.This study formulates the rules corresponding to the three levels from three questions and a total of 27 sub-dimensions. The students' answers are corresponding to the corresponding dimension rules, and the students' sub-dimension levels are judged as level 0, level 1 and level 2 respectively. Table 3 presents an overview of the grades for one of the essay questions.

Table 3 Scoring structure of scientific thinking assessment tool

Scientific thinking	Sub-skill dimension	grade		
Scientific reasoning	Asking questions and making assumptions	Level 0	Level 1	Level 2
	Design experiments and generate data	Level 0	Level 1	Level 2
	Explain data and draw conclusions	Level 0	Level 1	Level 2
Scientific Argumentation	Viewpoint	Level 0	Level 1	Level 2
	Facts and theoretical basis	Level 0	Level 1	Level 2
	Reasoning and rebuttal	Level 0	Level 1	Level 2
Scientific modeling	Model construction and use	Level 0	Level 1	Level 2
	Model checking and correction	Level 0	Level 1	Level 2
	Modeling Metacognition and Metamodeling Knowledge	Level 0	Level 1	Level 2

In order to test the gauge developed in this study, Winsteps3.72.0 software was used to analyze the item grading structure of the test questions, and the results were shown in Figure 3.

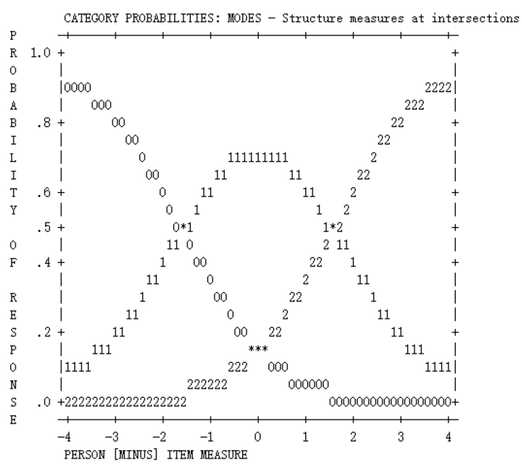

Figure 3 Grading structure of test questions

The horizontal coordinate (PERSON[MINUS] Item MEASURE) in the figure represents the difference between the student's ability and the difficulty value OF the item, and the vertical coordinate (PROBABILITY OF RESPONSE) represents the probability of the student getting 0, 1 and 2 points. The Threshold position, that is, the place where the curves cross in the graph, corresponds to the same ordinate, which means that students have the same probability of getting two scores. As can be seen from the result graph of 27 sub-dimensions, the rating grade category curve of basically each dimension has obvious peaks and is straight, and covers a certain range in the horizontal coordinate, performing well.

2.3.4 White Chart

The White Chart provides information on how the distribution of project difficulty matches students' ability levels, listing 4,415 students and 27 project locations in a generic scale. The first column is the logit scale, and columns 2 and 3 graphically describe the student's location and 27 items, respectively. The White chart converts student scores and item scores on a universal interval scale in logit units. As shown in Figure 4, the student and item logit scales from -5 to +3logit. From top to bottom, the ability of students and the difficulty of questions are reduced in turn. Each "#" represents 20 students, each "." Representing 1 to 19 students. Content adequacy and content validity can be assessed from the distribution and ranking of items in the White chart. The difficulty distribution range of the test questions is about 6 logits, the most difficult item BE, BH, CI is 0.57logit, the simplest item AA is -0.86logit, The most difficult item be, BH, CI is 0.57logit, the simplest item AA is -0.86logit, there is no significant gap between the items, indicating that the difficulty distribution of the test questions is good, and it has good content adequacy and effectiveness.

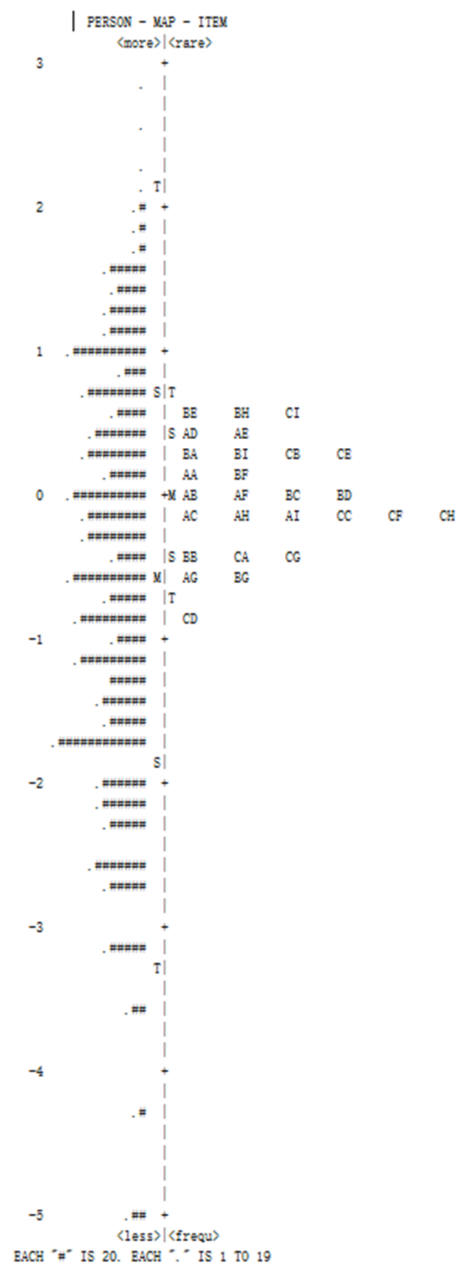

Figure 4 White Diagram

3.2 Scientific higher-order thinking ability and performance

3.2.1 Overall performance of scientific thinking ability

In the three dimensions of the assessment of scientific thinking ability, as shown in Figure 5, the overall average score of primary school students' scientific thinking ability is 0.82, and the comparison between primary school students' scientific thinking ability and their three sub-abilities can be visually displayed on this axis. The score of scientific argument is the lowest, only 0.78, which is much lower than the average score of scientific thinking. The average score of scientific reasoning ability is slightly lower than the average score of scientific thinking, which is 0.81 points; The score of scientific modeling ability was higher than that of scientific thinking ability (0.84).

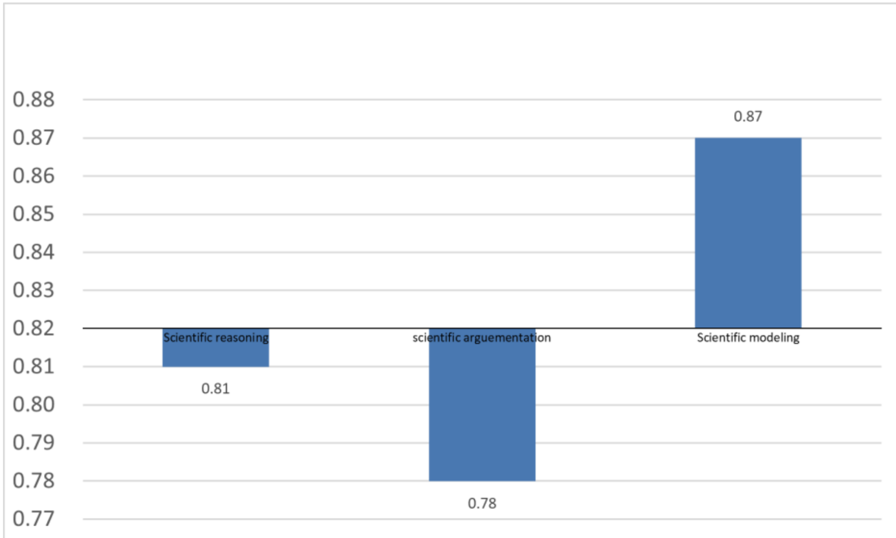

Figure 5 Overall score and sub-dimension score of scientific thinking ability of primary school students

Table 4 Descriptive statistics of scientific thinking ability

	Dimension	mean value	standard deviation	bias angle	kurtosis
Title number	Question 1	0.82	0.4611	0.267	-0.721
	Question 2	0.81	0.47881	0.184	-1.115
	Question 3	0.83	0.67288	0.157	-1.496
Subskills	Scientific argumentation ability	0.78	0.4945	0.16	-1.168
	Scientific reasoning ability	0.81	0.51506	0.127	-1.244
	Scientific modeling ability	0.87	0.51658	0.503	-0.621
Inference sub-dimension	Asking questions and making assumptions	0.80	0.62422	0.169	-1.285
	Design experiments and generate data	0.82	0.59262	0.186	-1.149
	Explain data and draw conclusions	0.81	0.7533	0.285	-1.624
Argumentation sub-dimension	Viewpoint	0.83	0.5321	0.245	-0.663
	Facts and theoretical basis	0.69	0.53452	0.274	-1.057
	Reasoning and rebuttal	0.82	0.61485	0.163	-1.303
Modeling sub-dimensions	Model construction and use	0.97	0.60939	-0.022	-1.128
	Inspection and correction	0.80	0.47703	0.446	-0.48
	Modeling metacognition and modeling meta-knowledge	0.74	0.51067	0.461	-0.51
Overall level	Scientific thinking	0.82	0.46916	0.191	-1.079

Table 4 shows the descriptive statistical analysis results of each sub-dimension and sub-skill of students' scientific thinking ability in different questions, which intuitively shows the score levels of each sub-dimension and sub-skill of primary school students' scientific thinking ability in different questions. The scores of three different questions were close, ranging from 0.81 to 0.83. According to the sub-dimensions of scientific thinking, the average score of scientific reasoning ability is between 0.80 and 0.83, the average score of scientific argumentation ability is between 0.69 and 0.83, and the average score of scientific modeling ability is between 0.74 and 0.97. Among them, the dimension of "model construction and use" in scientific modeling ability has the highest score (0.97), and the dimension of "fact and theoretical basis" in scientific argumentation ability has the lowest score (0.69). The absolute value of skewness of each dimension is less than 1, indicating that the score of the subjects is relatively symmetrical. The skewness of the dimension "model construction and use" is less than 0, indicating that more people score lower than the average in this dimension. The kurtosis of each sub-dimension of scientific thinking is less than 0, indicating that the overall score of students' scientific thinking level is in the middle level, and very low or very high score is less. From the above data analysis, in the process of developing students' scientific thinking, it is most urgent to improve the sub-dimension of "facts and theoretical basis" in students' scientific argumentation ability.

In order to further characterize the score of each dimension of scientific thinking ability, this study calculated the score rate of each sub-dimension and drew the score rate distribution map. From the map, it can be seen that the score distribution of each sub-dimension of scientific thinking of primary school students in grade 6 in Jinzhou is relatively average. Except for the sub-dimension of model construction and use ' with the highest score rate and the three sub-dimensions of ' fact and theoretical basis ' and ' modeling metacognition and modeling meta-knowledge ' with lower score rate, the score rates of the other six sub-dimensions are between 0.40 and 0.42, all of which are in the lower score rate range.

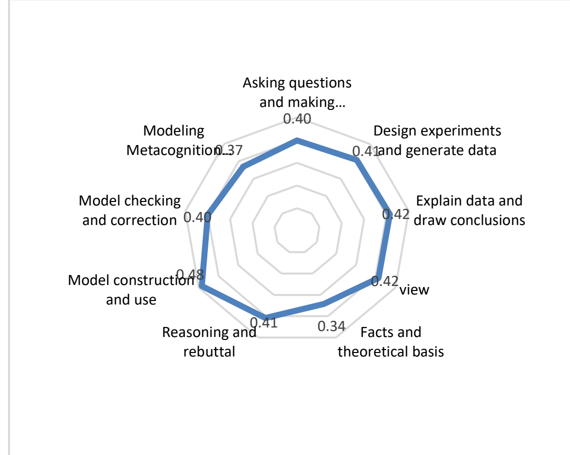

Figure. 6 The distribution map of the scoring rate of sub-dimensions of scientific higher-order thinking ability

3.2.2 Research on gender differences

From the overall results of the scientific thinking ability level of this evaluation, the overall scientific thinking level of girls (0.84) is slightly higher than that of boys (0.80). From the three sub-skills and nine sub-dimensions of scientific thinking, the scores of girls in grade 6 primary school students in Jinzhou are higher than those of boys. In the table 5, the scores of boys and girls are shown in detail. Among them, the gender difference in the sub-dimension of 'explanation data and conclusion ' under scientific reasoning ability is the most significant, which is 0.06 points. The gender difference in the ' model test and correction ' dimension under the scientific modeling ability is the least significant, only 0.02 points.

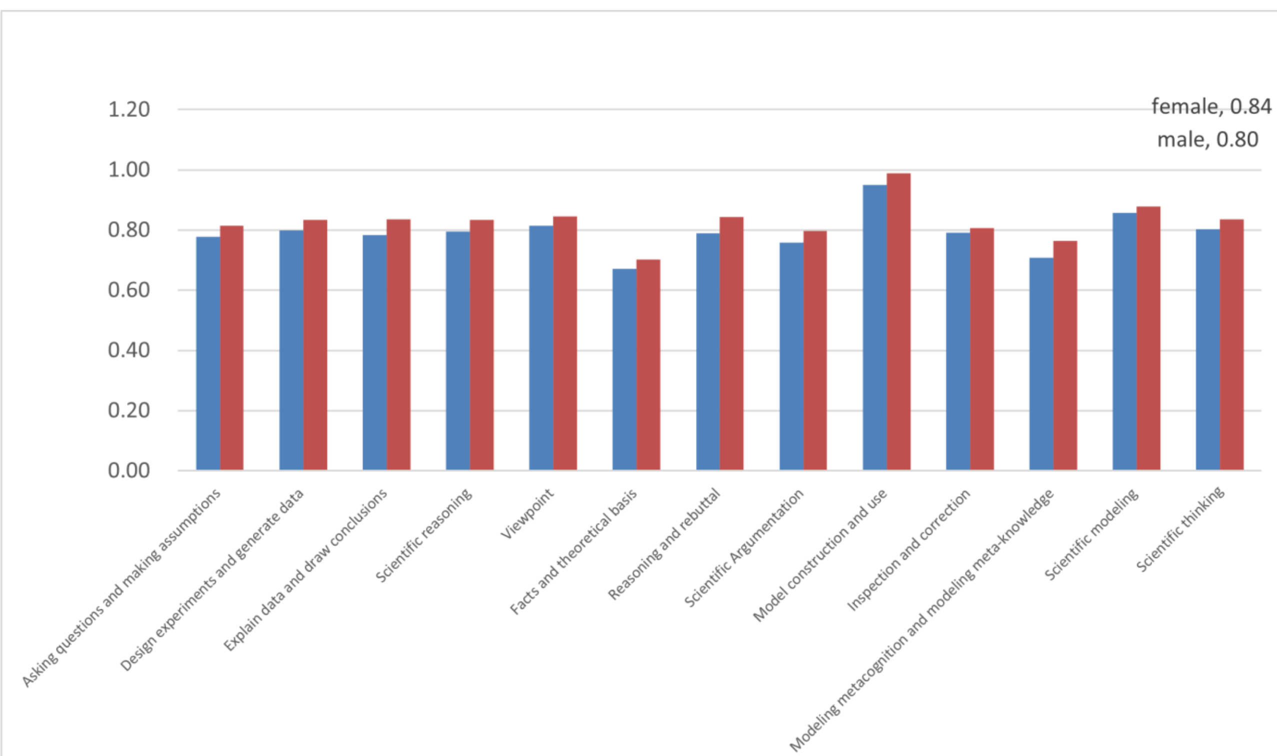

Figure.6 Comparison of male and female students ' scientific thinking ability sub-dimension level

Table 5 Descriptive statistics of scientific thinking ability of students of different genders

boy	girl	boy	girl	boy	girl	boy	girl

Scientific reasoning		Asking questions and making assumptions		Design experiments and generate data		Explain data and draw conclusions	
0.80	0.83	0.78	0.81	0.80	0.83	0.78	0.84
Scientific Argumentation		Viewpoint		Facts and theoretical basis		Reasoning and rebuttal	
0.76	0.80	0.82	0.85	0.67	0.70	0.79	0.84
Scientific modeling		Model construction and use		Inspection and correction		Modeling metacognition and modeling meta-knowledge	
0.86	0.88	0.95	0.99	0.79	0.81	0.71	0.76
Scientific thinking							
0.80	0.84						

3.2.3 Regional differences

In the analysis of regional differences, this study divides students with an average score of 0 in the scientific thinking level test into students with low scientific thinking level, students with an average score of 0-1 in the scientific thinking level test into students with medium scientific thinking level, and students with an average score of 1-2 in the scientific thinking level test into students with high scientific thinking level.

Figure 7 shows the scores of the eight experimental areas selected in this test in terms of scientific thinking ability. From the perspective of the percentage of scores at the three levels, Heishan County has the largest number of students with high scientific thinking level, with 79.70 % of students getting 1-2 points in the test of scientific thinking ability. At the same time, it is also the area with the least students with low scientific thinking level, with only 0.20 % of students having 0 points in scientific thinking ability. The second is the high-tech zone, where students with high scientific thinking level account for 67.00 %, and students with low scientific thinking level account for only 0.90 %. Taihe District also has more students with high scientific thinking level. The proportion of students with high scientific thinking level in this area is 60.30 %, and students with low scientific thinking level account for 0.80 %. In addition, the proportion of students with high scientific thinking level in Beizhen City is the least, only 15.00 %, and the proportion of students with low scientific thinking level is the most, reaching 2.10 %. There are also Guta District where students with high scientific thinking level account for a relatively small proportion. Students with high scientific thinking level in this area only account for 19.90 %, and students with low scientific thinking level account for 1.50 %. The proportion of students with high scientific thinking level in other regions is above 20.00 %. The proportion of students with high scientific thinking level in Linghai City, Yixian County and Linghe District increases in turn, which is 20.60 %, 21.50 % and 26.50 % respectively. The proportion of students with low scientific thinking level is 1.50 %, 0.40 % and 1.10 % respectively.On the whole, the students with high scientific thinking level in the eight experimental areas of Jinzhou City only account for 37.70 %, and the students with high scientific thinking level account for a relatively small proportion, and most of the students are at the level of scientific thinking (60.70 %).

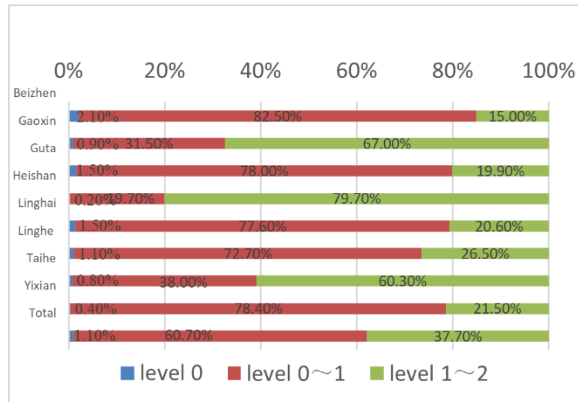

Figure.7 Regional differences in scientific thinking ability of primary school students in Jinzhou

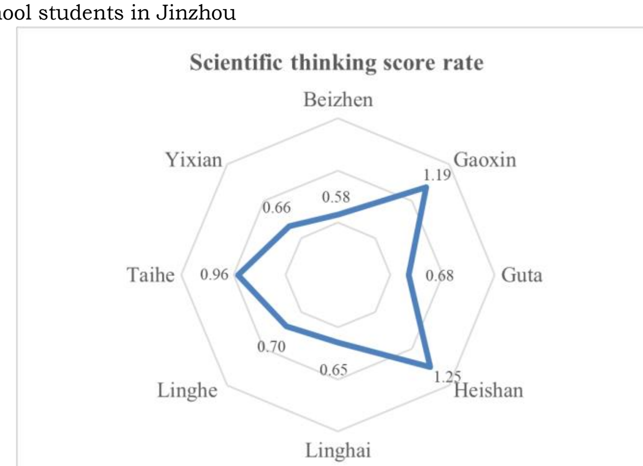

Figure.8 Average score radar chart of scientific thinking level test in eight experimental areas of Jinzhou City

From the regional difference radar chart of the average score of the scientific thinking level test, it can be seen that the level of scientific thinking in various regions of Jinzhou City is quite different. There are only two regions with an average score of more than 1 point, among which the sixth grade students in Heishan County have the highest level of scientific thinking, which is 1.25 points. The second is the high-tech zone (1.19 points) ; among the areas with an average score of less than 1 point, the average score of Taihe District was 0.96 points, and the rest from high to low were Linghe District (0.70 points), Guta District (0.68 points), Yixian County (0.66 points), Linghai City (0.65 points), and Beizhen City (0.58 points).

4. Summary and discussion

The evaluation tool performs well in terms of reliability, validity and discrimination. The measured results fit well with the model. The test question rating scale structure is good, which can effectively distinguish the levels of different subjects in different dimensions. In general, the quality of the high-order scientific thinking assessment tool for grade 6 primary school students in Jinzhou is ideal, and it can accurately evaluate the level of students ' high-order scientific thinking ability.

The research shows that the overall level of scientific higher-order thinking of grade 6 students in Jinzhou primary school is low, among which the scientific modeling ability is relatively good, and the dimension of ' facts and theoretical basis ' in scientific argumentation skills is poor. The scientific higher-order thinking ability of girls in Jinzhou is slightly higher than that of boys. There are great differences in the level of higher-order thinking among sixth-grade students in different regions of Jinzhou City.

4.1 Focus on evidence-based learning

The results show that students perform poorly in the scientific argumentation sub-skills, in which the ability of the factual and theoretical basis dimension needs to be developed by students in scientific practice. Teachers need to guide students to carry out exploratory experiments. In the process of experimental exploration, students are required to collect relevant data by themselves and use scientific methods to improve their ability to analyze and process data. Evidence-based teaching enables students to acquire new knowledge, to construct and evaluate hypotheses and ideas, and to seek answers to questions on their own. The process of evidence-based thinking is a process that promotes understanding of the nature of science as well as knowledge learning(Deng & Wang, 2014). Therefore, in the teaching process, students should pay attention to the learning of the facts and theoretical basis for drawing conclusions.

4.2 Correctly view gender differences in scientific thinking ability

The results of the gender difference test show that the level of scientific thinking of girls is slightly higher than that of boys. In the teaching process, it is necessary to eliminate the stereotyped concept of gender difference between men and women, treat every student fairly in teaching, respect the characteristics and ability level of individual development and teach students according to their aptitude.

4.3 Eliminate regional disparities

The level of scientific thinking varies greatly among different regions in Jinzhou, and areas with low high-order thinking level in the assessment results can increase investment in education, and high-quality educational resources can be introduced to promote local thinking education in the era of high sharing of digital education resources.

In this study, the higher-order thinking level of some primary school sixth-grade students in various areas of Jinzhou City was only evaluated by the developed assessment tools, and the samples were all from the same city, which was not sufficiently representative. In future research, we can focus on exploring the factors affecting higher-order thinking ability and effective teaching strategies and paths to improve the level of higher-order thinking ability, so as to provide an effective basis for the development of scientific thinking of Chinese students.

Acknowledge

The authors acknowledge the support provided by The Teacher Work Department of the Ministry of Education of China commissioned the project "Theoretical and Practical Research on the Professional Literacy of Primary School Science Teachers" (JSSKT2022009).

References

National Research Council. (2013). Next generation science standards: For states, by states.

Wang J Y, Zhou D H, Yang Y, Ke L & Tian X W.(2023). Scientific higher order thinking: Connotation value, structure function and practice Approach. Modern distance education (02), 11 to 18, doi: 10.13927 / j.carol carroll nki. Yuan. 20230522.002.

Bloom, B. S., & Krathwohl, D. R. （1956）. Taxonomy of Educational Objectives: The Classification of Educational Goals Handbook I: Cognitive Domain. New York: Longmans.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. Theory into practice, 32(3), 131-137.

Lai, C. L., & Hwang, G. J. (2014). Effects of mobile learning time on students' conception of collaboration, communication, complex problem – solving, meta – cognitive awareness and creativity. International Journal of Mobile Learning and Organisation, 8(3-4), 276-291.

Zhong Zhixian. (2004). The purpose of instructional design : to promote the development of learners ' higher-order abilities. Audio-visual education research (11), 13-19. doi : 10.13811 / j.cnki.eer.2004.11.003.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures. New York: Basic Books. https://doi.org/10.1037/10034-000

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. Journal of Research in Science Teaching, 15(1), 11 – 24.

Kuhn, D. (2010). What is scientific thinking and how does it develop?. The Wiley - Blackwell handbook of childhood cognitive development, 497-523.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. Developmental Review, 27(2), 172 – 223. https://doi.org/10.1016/j.dr.2006.12.001

Stuessy, C. L. (1989). Path analysis: A model for the development of scientific reasoning abilities in adolescents. Journal of Research in Science Teaching, 26(1), 41 – 53.

Xie Yun. (2007) • Argumentation, Argumentation, Argumentation-Dual-Dimensional Interpretation of the Concept of Argumentation in the Perspective of Contemporary Argumentation Theory.Research on Dialectics of Nature, 23 (4), 26-30.

Toulmin, S. E. (2003). The uses of argument. Cambridge university press.

Tang Xiaojun. (2016). Research on the teaching process model and its application strategy based on ' argumentation '. Contemporary educational science, (7), 51-54.

Van Eemeren, F. H., Grootendorst, R., & Henkemans, A. F. S. (2002). Argumentation: Analysis, evaluation, presentation. Routledge.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 39(1), 35-62.

Lawson, A. (2003). The nature and development of hypothetico - predictive argumentation with implications for science teaching. International journal of science education, 25(11), 1387-1408.

Hu Hongxing. (2021). The essence, value and teaching strategy of scientific argumentation. Contemporary Education and Culture (02), 42-49.doi : 10.13749 / j.cnki.cn62-1202 / g4.2021.02.007.

Dawson, V., & Carson, K. (2017). Using climate change scenarios to assess high school students’ argumentation skills. Research in Science & Technological Education, 35(1), 1-16.

Kuhn, D. (1992). Thinking as argument. Harvard educational review, 62(2), 155-179.

Chen Ying & Guo Yuying. (2017). The construction of the evaluation framework for the performance of high school physics scientific argumentation ability. Middle school physics teaching reference (21), 12-15.

Han Kwai Kwai. (2016). Scientific argumentation ability of middle school students (doctoral dissertation, Shaanxi Normal University).https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFDLAST2017&filename=1017043764.nh

Zhang, J., & Browne, W. J. (2023). Exploring Chinese high school students' perform ance and perceptions of scientific argumentation by understanding it as a three - compo nent progression of competencies. Journal of Research in Science Teaching, 60(4), 847-8 84.

Hestenes, D. (1992). Modeling games in the Newtonian world. American Journal of p hysics, 60(8), 732-748.

Halloun, I. (1996). Schematic modeling for meaningful learning of physics. Journal o f Research in Science Teaching: The Official Journal of the National Association for Resea rch in Science Teaching, 33(9), 1019-1041.

Stratford, S. J., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic mod eling processes: Analyzing, reasoning about, synthesizing, and testing models of stream e cosystems. Journal of Science Education and Technology, 7, 215-234.

Lin Jingwen, & Qiu Meihong. (2008).Explore the knowledge of high school students ' modeling and modeling process from the perspective of cognition / methodology. Scien ce Education Monthly, (307), 9-14.

Hung, J. F., & Lin, J. C. (2009). The development of the simulation modeling system and modeling ability evaluation. International Journal of u-and e-Service, Science and Te chnology, 2(4), 1-16.

Schwarz,C.,&White, B. (2005). Metamodeling knowledge: developing students’ unde rstanding of scientific modeling. Cognition and Instruction, 23(2), 165 205.

Enright, R. D., Wang Xu, J., Rapp, H., Evans, M., & Song, J. Y. (2022). The philosop hy and social science of agape love. Journal of theoretical and philosophical psychology.

Deng Yang & Wang houXiong. (2014). Necessity analysis of integrating scientific arg umentation into science education - based on the perspective of science ontology, scienc e learning and international comparison. Foreign primary and secondary education (03), 60-65 + 54.