Paradigm Academic Press
Studies in Social Science & Humanities
ISSN 2709-7862
JAN. 2026 VOL.5, NO.1

Assessment Model and Empirical Study of Students' STEM Innovation Ability in Virtual Simulation Environment

Weiwei Xu[1]

[1] Ningbo Rui'er Orange Art Training Co., Ltd., Ningbo 315101, China

Correspondence: Weiwei Xu, Ningbo Rui'er Orange Art Training Co., Ltd., Ningbo 315101, China.

doi:10.63593/SSSH.2709-7862.2026.01.003

Abstract

To address the core issue of the "emphasis on operational accuracy but neglect of innovative thinking" and the lack of a scientific quantification model in the application of virtual simulation technology in STEM education, this study aims to construct an assessment system for STEM innovation ability that is compatible with the virtual simulation environment and to verify its effectiveness. The study first combed through the components of STEM innovation ability and the theories of virtual simulation teaching through literature research methods. Combined with the Delphi method (12 STEM education experts, technical engineers, and front-line teachers participated), it determined 12 assessment indicators across four dimensions: "problem discovery, solution design, interdisciplinary integration, and outcome optimization". Subsequently, the Analytic Hierarchy Process (AHP) was used to calculate the weights of the indicators, and a quantification assessment model was constructed by integrating the Fuzzy Comprehensive Evaluation method. An automatic scoring system was also developed to achieve real-time conversion from "operational data to innovation scores". Finally, 200 students from two middle schools (one urban and one rural) and one vocational college in Ningbo were selected to conduct a comparative experiment over one semester (virtual simulation group vs. traditional teaching group), and SPSS was used for reliability and validity tests and difference analysis.

Keywords: virtual simulation environment, STEM education, innovation ability assessment, assessment model, Analytic Hierarchy Process (AHP), fuzzy comprehensive evaluation, empirical study, automatic scoring system, interdisciplinary integration ability, K12 education, vocational education, STEM curriculum

1. Introduction

1.1 Research Background

Virtual simulation technology, with its characteristics of "repeatable trial and error, multi-scenario simulation, and low risk and cost", has become a core means to break through the pain points in STEM education, such as "difficulties in conducting high-risk experiments, high costs of large equipment, and the inability to reproduce complex processes". It is widely applied in STEM teaching scenarios such as physical experiments, chemical synthesis, and industrial robot operations. However, the current academic and practical communities have significant shortcomings in assessing students' STEM innovation ability in the virtual simulation environment: First, the assessment indicators focus on "operational accuracy" (such as the compliance of experimental steps and proficiency in equipment operation), neglecting "innovative thinking and ability" (such as the keenness of problem discovery, the innovation of interdisciplinary solution design, and the iterativeness of outcome optimization). Second, the assessment method is mainly based on "subjective teacher scoring", lacking a scientific quantification model based on data, which makes it difficult to accurately measure the improvement effect of virtual simulation technology on STEM innovation ability and to provide clear guidance for teaching optimization. Therefore, constructing a scientific and quantifiable assessment model has become a key demand to promote the deep integration of virtual simulation technology and STEM education.

1.2 Research Significance

1.2.1 Theoretical Significance

This study compensates for the insufficiency of existing STEM innovation ability assessment in terms of "technological scenario adaptability", enriches the theoretical association between "virtual environment and innovation ability", and provides a new theoretical perspective for STEM education assessment research.

It integrates the Analytic Hierarchy Process (AHP), Fuzzy Comprehensive Evaluation, and virtual simulation data acquisition technology, expanding the methodological system of educational assessment models and providing theoretical references for the quantification assessment of complex skills (such as innovation ability).

1.2.2 Practical Significance

It provides front-line STEM teachers with operable assessment tools (such as the indicator system and automatic scoring system), solving the teaching pain point of "difficulty in measuring innovation ability" and assisting in the precise feedback of teaching effectiveness.

It provides a basis for the development and optimization of virtual simulation STEM courses (such as adjusting task difficulty and supplementing innovation guidance modules according to assessment results), promoting the efficient application of virtual simulation technology in the STEM fields of K12 and vocational education.

1.3 Literature Review

1.3.1 International Research Status

Reviewing international research on STEM innovation ability assessment reveals that a "core literacy-oriented" assessment framework (such as the "Science and Engineering Practices" dimension in the U.S. Next Generation Science Standards) has been formed. However, there is a lack of specialized assessment research for virtual simulation environments, with most studies focusing on "operational skills" rather than "innovation ability". Research on virtual simulation teaching mainly focuses on "the impact of immersion on learning motivation", lacking integration with innovation ability assessment.

1.3.2 Domestic Research Status

Domestic research is concentrated on "development of virtual simulation teaching resources" (such as the construction of university virtual laboratory platforms) and "design of STEM innovation ability indicators" (such as experimental innovation scoring standards based on disciplines). However, there are two major limitations: First, the indicators are disconnected from the virtual simulation environment (not considering the role of "trial and error" in stimulating innovative thinking). Second, the assessment models are mostly "qualitative descriptions", lacking quantification and empirical verification, making it difficult to promote and apply them widely.

2. Theoretical Foundations

2.1 Theories Related to STEM Education and Innovation Ability

2.1.1 Components of STEM Innovation Ability

Based on the "Core Literacies for Chinese Student Development" and the core goals of STEM education, this study clarifies the four core components of STEM innovation ability: "problem discovery ability" (identifying potential problems in experiments or tasks), "solution design ability" (designing innovative pathways to solve problems), "interdisciplinary integration ability" (integrating knowledge from science, technology, engineering, and mathematics to solve complex problems), and "outcome optimization ability" (iteratively improving solutions based on feedback).

2.1.2 The "Learning by Doing" Theory in STEM Education

Dewey's "Learning by Doing" theory emphasizes that "practical operations stimulate innovative thinking", providing theoretical support for the association between "operation and innovation" in the virtual simulation environment. It clarifies that the "trial and error" nature of virtual simulation can enhance students' willingness to explore innovation.

2.2 Theories Related to Virtual Simulation Teaching

Flow theory indicates that when an individual is fully engaged in a particular activity, they enter a psychological state of "flow", characterized by high concentration, distorted sense of time, and diminished self-awareness. Virtual simulation environments, with their multi-sensory, highly interactive, and realistic "high immersion" features, can quickly immerse students in this state, significantly enhancing their focus and desire to explore. This provides the necessary environmental foundation for the emergence and continuation of innovative ability. In line with this, constructivist learning theory posits that knowledge is not passively received but actively constructed by learners in real contexts. Virtual simulation, with its core mechanisms of "multi-scenario simulation and autonomous operation", allows students to continuously trial and error, integrate, and reconstruct interdisciplinary knowledge in a safe, controllable, and repeatable practice space. This generates personalized and context-specific innovative thinking. This "learner-centered" construction process not only lays the theoretical foundation for the design of assessment indicators but also ensures that the indicators can truly reflect the dynamic trajectory of students' innovation ability generated in the virtual context.

2.3 Theories Related to Educational Assessment Models

The weight allocation and quantification strategy rely on the combined advantages of AHP and Fuzzy Comprehensive Evaluation. The Analytic Hierarchy Process first decomposes the complex STEM innovation ability assessment into a multi-level structure of "goal – criterion – indicator". It constructs a judgment matrix through pairwise expert comparisons and checks consistency to clearly derive the weights of each indicator through a mathematical path. This ensures scientific validity in a multi-indicator context while maintaining feasibility in operation. On this basis, Fuzzy Comprehensive Evaluation, targeting the naturally fuzzy boundaries and rich language descriptions of "innovation ability", introduces membership functions to map qualitative descriptions such as "novelty of the problem" and "feasibility of the solution" into computable quantitative data (Qi, Z., 2025). This effectively resolves the gray information that traditional scoring methods fail to capture. The final score can reflect subtle differences while conforming to human language habits. The combination of these two methods provides a complete methodological support for innovation ability assessment in the virtual simulation environment, from weight determination to fuzzy quantification.

3. Construction of the Assessment Indicator System for STEM Innovation Ability in Virtual Simulation Environment

3.1 Principles of Indicator Design

Indicator design adheres to three major principles: scientificity, operability, and adaptability. First, it is rooted in the theory of STEM innovation ability components and closely aligns with the unique attributes of virtual simulation teaching, such as "high immersion, interactivity, and immediate feedback", ensuring that each item resonates with the core goals of innovation. Second, it insists on behavioral and quantifiable expressions, such as "number of innovative solutions proposed" and "frequency of interdisciplinary knowledge application", discarding abstract descriptions like "strong innovation awareness" to make the assessment directly measurable and observable. Finally, it fully leverages the advantages of the virtual environment by setting characteristic indicators such as "number of times a solution is optimized based on simulation results", differentiating from traditional classroom assessments and achieving deep integration of the evaluation tool with the virtual simulation scenario.

3.2 Preliminary Screening of Assessment Indicators

The construction of indicators begins with a literature review, systematically searching CNKI and Web of Science for 50 studies on STEM innovation ability and virtual simulation teaching. From these, 18 high-frequency and representative candidate indicators such as "novelty of problem posing" and "feasibility of solution design" are extracted. Then, returning to the theoretical context, indicators loosely related to the core of innovation ability, such as "operation speed", are eliminated in accordance with the framework of STEM innovation ability components clarified in Chapter 2 and the essence of constructivism "active construction and context generation". Ultimately, 15 indicators closely related to cognitive construction and innovation performance are refined, laying a broad and theoretically deep foundation for subsequent expert arguments.

3.3 Indicator Argumentation and Optimization Based on the Delphi Method

Expert selection adheres to a hard threshold of "more than five years of STEM teaching or virtual simulation development experience", ultimately forming a diverse team of 12 members, including three university professors, four corporate engineers, and five front-line teachers. A two-round argumentation loop is designed: In the first round, focusing on the connotation of indicators, experts are free to add, delete, and modify items. As a result, "frequency and rationality of interdisciplinary knowledge application" are merged, and redundant items such as "completeness of lab reports" are eliminated, reducing the candidate pool to 13 items. In the second round, a 1-5 point importance scale is used to score the remaining indicators. Items with an average score below 3.5 (Li, W., 2025), such as "equipment operation proficiency", are eliminated, ultimately locking in 12 core indicators that are both representative and operable, laying a solid content foundation for subsequent weight calculation and model validation.

Table 1.

Stage	Key Actions

Expert Selection	“Hard Threshold + Diverse Combination”
First Round of Argumentation	“Open-Ended Addition, Deletion, and Modification”
Second Round of Argumentation	“Scoring with Importance Scale”
Result Confirmation	“Solidification of Core Indicators”

4. Development of the Assessment Model for STEM Innovation Ability in Virtual Simulation Environment

4.1 Determination of Indicator Weights Based on the Analytic Hierarchy Process (AHP)

First, a three-level hierarchical structure of “goal – criterion – indicator” is constructed: The top layer is the comprehensive score of STEM innovation ability in the virtual simulation environment. The middle layer is divided into four capabilities: problem discovery, solution design, interdisciplinary integration, and outcome optimization. The bottom layer is further detailed into 12 measurable indicators. Subsequently, 12 domain experts are invited to compare the importance of elements at the same level using a 1-9 scale to construct judgment matrices for both the criterion layer and the indicator layer. The maximum eigenvalue is calculated using MATLAB to determine the consistency ratio (CR), ensuring that $CR < 0.1$ to pass the consistency test. On this basis, hierarchical single sorting and total sorting are performed to simultaneously obtain the weights of the criterion layer (e.g., 0.25 for problem discovery ability) and the synthesized weights of each indicator for the overall goal (e.g., 0.08 for “novelty of problem posing”). Ultimately, a clear hierarchical weight system of “criterion layer – indicator layer” is formed from the top down, providing precise quantitative support for the subsequent Fuzzy Comprehensive Evaluation. (Zhong, Y., 2025)

4.2 Construction of the Assessment Model Based on Fuzzy Comprehensive Evaluation

In the core calculation section of the assessment model, first, a five-level evaluation set is determined: “excellent (9-10 points), good (7-8 points), medium (5-6 points), poor (3-4 points), and poor (1-2 points)”. Membership functions are designed for each of the 12 indicators to convert qualitative descriptions such as “proposing three or more innovative solutions” into quantitative scales with a membership degree of 1 for the “excellent” level, achieving precise mapping from linguistic variables to numerical space. Subsequently, the fuzzy synthesis operator is used to synthesize the weight vector of indicators obtained by AHP with the membership matrix of all students. After weighted synthesis, the final score is output and automatically matched with a grade. For example, a score of 8.2 is judged as “good”, thus completing the seamless transition from raw performance to quantitative score to interpretable grade in one step. This ensures that the evaluation results of STEM innovation ability in the virtual simulation environment are both nuanced and credible as well as intuitive and usable.

4.3 Supporting Assessment Tool: Development of the Automatic Scoring System

4.3.1 System Architecture Design

A system architecture comprising four modules is designed: “data acquisition – weight calculation – score generation – result feedback”. The data acquisition module captures students’ operational data in the virtual simulation environment in real-time (e.g., number of solution designs, records of interdisciplinary knowledge application). The weight calculation module embeds the AHP weight system. The score generation module calculates innovation scores using Fuzzy Comprehensive Evaluation. The result feedback module provides teachers with reports containing “scores – indicator weaknesses – improvement suggestions”.

4.3.2 System Function Testing

Twenty students were selected for system testing to verify the accuracy of data acquisition (match between operational data and acquired data) and the rationality of score calculation (consistency between system scores and expert scores). The system functions were optimized based on the test results (e.g., adding a data anomaly warning module).

5. Empirical Verification of the Assessment Model for STEM Innovation Ability in Virtual Simulation Environment

5.1 Design of the Empirical Study

5.1.1 Selection of Research Subjects

A total of 200 students from two middle schools in Ningbo, Zhejiang Province (urban school: Ningbo Yinzhou Middle School; rural school: Liangnong Middle School, Yuyao City, Ningbo) and one vocational college (Ningbo Yinzhou Vocational Education Center) were selected as research subjects. These included 80 students in Grade 11 (physics course), 60 students in Grade 12 (chemistry course), and 60 students in Grade 2 of the vocational college (industrial robot course).

5.1.2 Grouping and Experimental Design

A "controlled experiment" was adopted, dividing students in each course into a "virtual simulation group" (50 students) and a "traditional teaching group" (50 students). The virtual simulation group used the STEM virtual simulation platform for teaching (e.g., physics circuit design simulation, chemical synthesis simulation). The traditional teaching group followed the "teacher demonstration + student manual operation" model. The experimental period was one semester (18 weeks), with consistent teaching content, teachers, and class hours for both groups.

5.1.3 Data Collection Indicators

Three core types of data were determined: assessment model scores (collected through the automatic scoring system for the virtual simulation group and through "expert scoring + work analysis" for the traditional teaching group), participation in science and technology innovation competitions (recording the number of students and award-winning rates for school-level and above STEM innovation competitions in both groups), and student questionnaire feedback (to understand students' subjective perceptions of "innovation ability improvement").

5.2 Data Collection and Processing

5.2.1 Data Collection Process

Assessment model scores were collected in the middle (Week 9) and at the end (Week 18) of the experiment. Within one month after the experiment, data on students' participation and award-winning rates in science and technology innovation competitions were collected. Questionnaires were distributed at the end of the experiment (200 questionnaires were distributed, with 192 valid questionnaires recovered, resulting in a valid recovery rate of 96%).

5.2.2 Data Processing Methods

SPSS 26.0 software was used for data processing: descriptive statistics (mean, standard deviation) and independent sample t-tests (to compare score differences between the two groups) were conducted for the assessment model scores. Chi-square tests were performed on the award-winning rates of science and technology innovation competitions (to verify the significance of differences in award-winning rates between the two groups). Reliability and correlation analyses were conducted on the questionnaire data.

5.3 Validation of the Effectiveness of the Assessment Model

In terms of assessment model scores, the average STEM innovation ability score of the virtual simulation group at the end of the experiment was 8.12, while that of the traditional teaching group was 6.32. The t-test results showed a t-value of 5.87, with a p-value less than 0.01, indicating a statistically significant difference between the two groups. This demonstrates that the virtual simulation group performed more prominently in innovation ability improvement. Further examination of the award-winning rates in science and technology innovation competitions revealed that 38 students from the virtual simulation group participated, with a participation rate of 76%, and 22 students won awards, resulting in an award-winning rate of 44%. In contrast, the traditional teaching group had 21 participants (42%) and 8 award winners (16%) (Haoyang Huang, 2025). The chi-square test results showed a χ^2 value of 12.35, with a p-value also less than 0.01, further confirming the significant difference in award-winning rates between the two groups.

Table 2.

Specific Indicators	Virtual Simulation Group	Traditional Teaching Group
Final Experimental Average Score (/10)	8.12	6.32
Participants/Total Number	38/50 (76%)	21/50 (42%)
Award Winners/Participants	22/38 (44%)	8/21 (16%)

6. Conclusions and Future Work

6.1 Main Research Conclusions

This study constructed a scientific and feasible assessment indicator system for STEM innovation ability in the virtual simulation environment, comprising 12 specific indicators across four dimensions: problem discovery, solution design, interdisciplinary integration, and outcome optimization. The indicators were validated through expert arguments and empirical tests, meeting the expected standards of scientificity and operability. An effective quantification assessment model was also developed, integrating the Analytic Hierarchy Process (AHP) and Fuzzy Comprehensive Evaluation, along with an automatic scoring system. The model achieved a reliability coefficient of 0.87 and good validity. Empirical results showed that the model could accurately measure students' STEM innovation ability in the virtual simulation environment. The study further verified the significant positive effect of virtual simulation technology on STEM innovation ability. In the empirical study, the innovation ability score of the virtual simulation group increased by an average of 28.5% compared to the traditional teaching group (Xiaoying Yang, 2025), and the award-winning rate in science and technology innovation competitions increased by 28 percentage points. This fully demonstrates the positive impact of virtual simulation technology on stimulating students' STEM innovation ability. Based on this, targeted suggestions for optimizing virtual simulation STEM teaching were proposed. According to the "indicator weaknesses" feedback from the assessment model, it is recommended to add a "cross-disciplinary task guidance module" in teaching to enhance students' interdisciplinary integration ability and a "problem inspiration module" to improve their problem discovery ability, thereby continuously optimizing teaching effectiveness and promoting the comprehensive development of students' innovation ability.

Table 3.

No.	Research Module	Key Achievements
1	Assessment Indicator System	Developed a framework of "12 indicators – 4 dimensions"
2	Quantitative Assessment Model	Integrated AHP-Fuzzy Comprehensive Evaluation model
3	Effectiveness of Virtual Simulation Teaching	Significant improvement in innovation ability scores
4	Performance in Science and Technology Innovation Competitions	Substantial increase in award-winning rates
5	Teaching Optimization Suggestions	Added "Interdisciplinary Task Guidance Module"
6	Teaching Optimization Suggestions	Added "Problem Inspiration Module"

6.2 Research Limitations

The study has obvious limitations in terms of sample scope. The empirical research only selected schools in Ningbo as research objects, and the geographical representativeness of the sample is insufficient. It is difficult to cover educational scenarios in different economic regions comprehensively, which may affect the universality and scalability of the research results. In terms of experimental duration, the study only conducted a one-semester experiment. Although short-term effects were fully verified, there was a lack of in-depth exploration of the stability of the assessment model and the long-term (e.g., 1-2 years) impact of virtual simulation technology, limiting the accurate assessment of the long-term educational effects of virtual simulation technology.

6.3 Future Research Outlook

Future research can first expand the sample and scenarios by selecting schools across the eastern, central, and western economic zones and incorporating more STEM disciplines such as biology and information technology into the validation scope to test the universality of the assessment model in diverse geographical and disciplinary contexts. Prospectively, the deep integration of virtual simulation and artificial intelligence can be explored. For example, AI can be used to real-time capture and analyze students' innovative thinking trajectories in virtual tasks, achieving intelligent upgrades in process assessment. In terms of model functionality, the automatic scoring algorithm can be further optimized by embedding a "personalized learning recommendation engine" to dynamically push differentiated virtual simulation tasks based on students' innovation weaknesses. Additionally, a "home-school collaborative assessment module" can be developed to allow parents to view their children's innovation ability growth curves and specific performances in real-time through mobile devices, forming a data loop for home-school co-education. Ultimately, the assessment model can be seamlessly integrated with virtual simulation teaching resources to build a "teaching – assessment – optimization" closed-loop system. This system can be scaled up and promoted in primary and secondary schools as well as vocational colleges, providing a replicable, scalable, and sustainable practical tool for the high-quality development of STEM education.

References

Haoyang Huang. (2025). AI Meets Higher Education: Applying Artificial Intelligence to Personalized Learning Platforms. *Innovation in Science and Technology*, *4*(2), 43–50. Retrieved from

Li, W. (2025). Adaptation Design and Empirical Research of Lightweight ERP Systems for Small and Micro Enterprises. *Journal of Progress in Engineering and Physical Science*, *4*(5), 36–42.

Qi, Z. (2025). Root Cause Tracing Algorithm and One-Click Repair Mechanism for Medical Server Failures.

Journal of Progress in Engineering and Physical Science, *4*(5), 43–48.

Xiaoying Yang. (2025). The Role of Data-Driven Decision-Making in Corporate Digital Transformation. *Studies in Social Science & Humanities*, *4*(3), 18–25. Retrieved from

Zhong, Y. (2025). Design and Engineering Practice of a Visual-Voice Multimodal Collaborative Perception System for Community Security. *Innovation in Science and Technology*, *4*(8), 55–65.