ISSN: 2637-6067 | DOI: 10.6918/IJOSSER.202512_8(12).0041

An Exploration of Junior High School Biology Curriculum Evaluation Based on Q-Matrix Cognitive Diagnostic Theory

Li Zhu *

The Affiliated Middle School of Mianyang Teachers' College, Mianyang, China

* Corresponding author

Abstract

This study introduces and applies a Q-Matrix cognitive diagnostic theory-based framework for the comprehensive evaluation of competency development within the junior high school biology curriculum. Departing from traditional summative assessment, the model establishes a systematic mapping between nine defined attributes—spanning life concepts, scientific thinking, inquiry, and socio-ethical responsibility—and specific test items via a structured Q-matrix. By transforming student response data into a binary score matrix (R) and employing a sequence of matrix operations, the method calculates individualized probabilities of mastery for each attribute. A case study involving two students with nearly identical total scores (80/100 vs. 79/100) demonstrates the model's diagnostic power: it reveals starkly contrasting competency profiles, with Student A excelling in critical thinking and model construction but weaker in evolutionary understanding, while Student B shows superior performance in evolution-adaptation and social responsibility but lags in practical application skills. These nuanced insights, unobtainable from aggregate scores, validate the framework's capacity for precise, personalized diagnosis. The results underscore its utility in identifying instructional strengths, pinpointing areas for pedagogical intervention, and providing actionable feedback for curriculum refinement. This approach offers a robust, theory-driven tool for advancing formative assessment and promoting the integrated development of scientific knowledge and ethical competencies in science education.

Keywords

Biology curriculum; Cognitive diagnostic; Q-Matrix; Academic evaluation.

1. Introduction

Cognitive Diagnostic Theory (CDT) has been studied for many years, and the goal of cognitive diagnosis is to discover the cognitive attributes from the person’s response. This has been used in many fields such as clinical measurement [1-2] and education [3-4]. For the cognitive diagnosis model, many methods have been proposed, such as Deterministic Inputs, Noisy-And gate model (DINA) [5], Item Response Theory (IRT) [6-7] , inner product in matrix factorization [8], the NPC [9-10] model, GDINA model [11] and DINMix model [12-13]. In this work, the Q-Matrix cognitive diagnostic model [14] is adopted.

For the education of junior high school, the assessment of learning in science education is undergoing a paradigm shift, moving beyond the measurement of factual recall toward the evaluation of integrated competencies that encompass conceptual understanding, procedural skills, and socio-scientific values. This shift is particularly salient in curricula aiming to integrate ideological-political education, where the goal is to cultivate not only scientifically literate but also ethically responsible citizens. However, a significant methodological gap persists: traditional evaluation methods, often reliant on composite test scores, lack the granularity to diagnose the multidimensional and often latent structure of student

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

competencies. They provide limited insight into how specific cognitive and affective attributes develop and interact, thereby offering teachers scant guidance for personalized instruction. To address this limitation, the present study explores the application of Q-Matrix Cognitive Diagnostic Theory (CDT) to the context of junior high school biology. CDT provides a psychometric framework for attributing observed test performance to a profile of underlying, unobservable attributes. This research constructs a CDT model by explicitly linking curriculum-defined graduation requirements—Life concepts, scientific thinking, scientific inquiry, and attitude & responsibility—to a set of nine operational attributes through a carefully designed Q-matrix. The primary objective is to develop and demonstrate a computational methodology that transforms conventional test data into a detailed diagnostic profile for each student.

Through a comparative case analysis of two students, this paper illustrates how the model uncovers distinct learning patterns masked by similar total scores, thereby showing its potential to inform differentiated teaching strategies and contribute to a more nuanced, evidence-based approach to curriculum evaluation and the integration of value-based education.

2. Q-Matrix Method

Assuming the number of students is l, the number of test items is m, and the number of competencies assessed is n, with matrix R representing the students' test scores, then matrix R can be expressed as formula (1).

$$\mathbf{R}=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & & \ddots & \\ R_{l1} & R_{l2} & \cdots & R_{lm} \end{bmatrix} \qquad R_{ij}\in\{0,1\} \tag{1}$$

Among them, R_{ij} represents the score of the i-th student on the j-th test question. To facilitate subsequent calculations, it only takes values of 0 and 1, where 0 indicates an incorrect answer, and 1 indicates a correct answer or a scoring rate of 0.6 or higher. The Q-matrix is defined to represent the relationship between test questions and competencies, where Q_{ij} indicates whether the i-th test question involves the j-th competency. If it involves the attribute in cognitive diagnosis, then $Q_{ij}=1$; if not, then $Q_{ij}=0$.

$$\mathbf{Q}=\begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & \cdots & Q_{2n} \\ \vdots & & \ddots & \\ Q_{m1} & Q_{m2} & \cdots & Q_{mn} \end{bmatrix} \qquad Q_{ij}\in\{0,1\} \tag{2}$$

Using matrix multiplication, the matrix N representing the number of questions answered correctly by each student across different competencies can be calculated, where N_{ij} denotes the number of questions involving the j-th competency that the i-th student answered correctly.

$$\mathbf{N}=\mathbf{R}\cdot\mathbf{Q}=\begin{bmatrix} N_{11} & N_{12} & \cdots & N_{1n} \\ N_{21} & N_{22} & \cdots & N_{2n} \\ \vdots & & \ddots & \\ N_{l1} & N_{l2} & \cdots & N_{ln} \end{bmatrix} \tag{3}$$

Then the frequency f_{ik} at which student i possesses attribute k is given by:

$$f_{ik} = \frac{N_{ik}}{S_k} \tag{4}$$

Where S_k is the total number of test questions that involve attribute k, calculated as follows:

$$S_k = \sum_{j=1}^{m} Q_{jk} \tag{5}$$

Then the probability g_{ij} that student i answers test question j correctly is:

$$g_{ij} = \prod_{k=1}^{n}(max[\, f_{ik}, (1 - Q_{jk})]) \tag{6}$$

Among them, max [x, y] represents the maximum value between x and y, and

$$\prod_{k=1}^{n}(z_k) = z_1 \cdot z_2 \cdot z_3 \cdots z_n \tag{7}$$

Represents the continued product over zk. Finally, the cognitive diagnostic probability of student i for attribute k is given by:

$$P_{ik} = \frac{\sum_{j=1}^{m} g_{ij} \cdot min[\, Q_{jk}, R_{ij}]}{\sum_{j=1}^{m} g_{ij} \cdot Q_{jk}} \tag{8}$$

Among them, min [x, y] represents the minimum value between x and y, and $P_{ik} = 0$ when $\sum_{j=1}^{m} g_{ij} \cdot Q_{jk} = 0$.

This estimation method uses the estimated probability of the examinee answering each question correctly as a weighting factor for their scores, when assessing the probability of an examinee mastering each attribute. This approach aims to mitigate the interferences between different attributes. If an examiner has a low correct response frequency for a particular attribute, the probability of correctly answering questions involving that attribute becomes relatively small. By using this probability as the question weight when estimating the attribute mastery probability, the influence of such questions on the estimation of other attributes covered by the same question is reduced. In other words, this method diminishes the impact of attributes with low correct response frequencies on the estimation of other attributes.

Furthermore, when estimating the mastery probability of a specific attribute, this method not only considers the responses related to that attribute but also reduces the influence of other attributes. As a result, it enables a more accurate estimation of the examinee's attribute mastery probability.

The computational process is implemented in MATLAB, as shown in Fig. 1. The process consists of the following key stages:

(1) Data input and matrix definition:

The inputs include matrix R, which represents student item scores, and the values are binarized. In addition, the Q-matrix, which defines the relationship between test questions and attributes.

(2) Calculation of matrix N:

The matrix N, which records the number of questions each student answered correctly for each attribute, is computed, and this is achieved through a matrix multiplication operation between matrix R and the transpose of matrix Q.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

(3) Calculation of the frequency f_{ik}:
This is derived by dividing each element in matrix N by the corresponding attribute's total question count. This frequency serves as an initial estimate of the probability that student i possesses attribute k.
(4) Calculation of question response probability g_{ij}:
This step involves the mastery frequencies f_{ik} and the Q-matrix. A common approach is to model this probability as the product of the student's mastery probabilities for all attributes required by the question. This formula computes the product of all f_{ik} for which $Q_{jk} = 1$.
(5) Calculation of mastery probability P_{ik}:
The final mastery probability P_{ik} is calculated. This step uses the previously calculated g_{ij} probabilities and the original response data R to refine the initial mastery estimates f_{ik} into more accurate posterior probabilities P_{ik}.
The ultimate goal is to estimate the probability of each student having mastered each defined attribute, utilizing principles from cognitive diagnostic modeling.

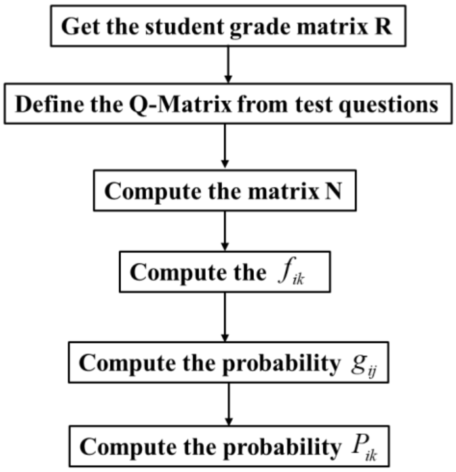

Figure 1. Computational process using MATLAB

3. Test Design

Table 1 presents a systematic mapping of the curriculum's graduation requirements to specific cognitive and behavioral attributes along with their corresponding core teaching content. Life concepts, scientific thinking, scientific inquiry, and attitude & responsibility are considered. Each requirement is decomposed into 2-3 specific attributes. For each attribute, representative main teaching content is provided. The table illustrates a comprehensive educational objective. It moves beyond foundational knowledge (life concepts) to encompass higher-order cognitive skills (scientific thinking, scientific inquiry) and culminates in value formation and social responsibility (attitude & responsibility). This reflects an integrated approach aiming to develop not just knowledgeable but also ethically conscious and scientifically literate individuals. The sequence from attribute 1 to 9 suggests a pedagogical progression from understanding core biological principles (1, 2) to applying scientific methods (3-6) and finally engaging with the societal and ethical dimensions of science (7-9).

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

Table 1. Graduation requirements and attributes considered

Graduation Requirements	Attributes ID	Attributes	Main Teaching Content
Life concepts	1	Structure-function perspective	Cell as the basic unit of biological structure and function; structural hierarchy of organisms; flower structure.
	2	Evolution-adaptation perspective	Biological adaptation to the environment through morphology (e.g., bird beak shape), physiology (e.g., cold resistance), and behavior (e.g., migration).
Scientific Thinking	3	Model Construction	Physical models (e.g., cell structure models), mathematical models (e.g., growth curves), conceptual models (e.g., ecosystem diagrams).
	4	Critical Questioning	Ethics of genetic technology; ecological protection debates.
Scientific Inquiry	5	Experimental Design	Experimental design (e.g., control variable method); inquiry activities (e.g., yeast respiration methods); discussion of social issues (e.g., ethics of genetic technology).
	6	Data Analysis	Dolly the sheep's genetic material came entirely from donor mother B, whose mammary cell nucleus contained a full set of genetic information.
Attitude Responsibility	7	Scientific Ethics	Genetically modified food controversies: Analyze scientific value and ecological risks through the "Golden Rice" case study.
	8	Social Responsibility	Understanding biodiversity.
	9	Lifelong Learning	Secrets of genes.

Table 2 specifies which attributes are assessed by each test item. Multiple choice primarily assesses two attributes per question, often pairing a core concept with another competency. Fill-in-the-blank assesses two or three attributes, introducing more complexity and potential for integrated application. Short answer assesses the highest number of attributes, requiring students to synthesize multiple competencies in extended responses. To ensure reliability, all nine attributes are assessed multiple times across the test. For instance, Structure-function perspective and evolution-adaptation perspective appear frequently, underscoring their foundational importance.

Table 2. Test questions design and attributes

Question ID	Question type	Attributes
1	Multiple choice	Structure-function perspective; Evolution-adaptation perspective
2	Multiple choice	Model construction; Critical questioning
3	Multiple choice	Evolution-adaptation perspective; Experimental design
4	Multiple choice	Structure-function perspective; Data analysis
5	Multiple choice	Lifelong learning; Scientific ethics
6	Multiple choice	Experimental design; Lifelong learning
7	Multiple choice	Evolution-adaptation perspective; Social responsibility
8	Multiple choice	Social responsibility; Model construction
9	Fill-in-the-blank	Data analysis; Structure-function perspective
10	Fill-in-the-blank	Model construction; Data analysis
11	Fill-in-the-blank	Critical questioning; Scientific ethics; Social responsibility
12	Fill-in-the-blank	Critical questioning; Structure-function perspective; Model construction
13	Fill-in-the-blank	Data analysis; Scientific ethics; Social responsibility
14	Fill-in-the-blank	Experimental design; Lifelong learning; Scientific ethics
15	Short answer	Critical questioning; Structure-function perspective; Model construction; Evolution-adaptation perspective
16	Short answer	Experimental design; Scientific ethics; Lifelong learning
17	Short answer	Data analysis; Critical questioning; Lifelong learning; Social responsibility
18	Short answer	Structure-function perspective; Experimental design; Scientific ethics; Lifelong learning; Evolution-adaptation perspective

Figure 2 provides a powerful visual summary of the data in table 2, depicting the frequency and pattern of how the nine attributes are distributed across the 18 test questions. The vertical axis lists the attribute IDs, and the horizontal axis lists the question IDs. A mark at an intersection indicates that the corresponding attribute is assessed in that specific question. From the figure 2, the density of marks for each attribute row visually confirms its assessment frequency. Structure-function, evolution-adaptation, and lifelong learning show high frequency, and attribute 8 appears less frequently but is consistently paired with other attributes. The density of marks for each question column visually represents its complexity. Questions with many marks are the most complex, assessing multiple integrated attributes, which aligns perfectly with their designation as short answer questions in Table 2.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

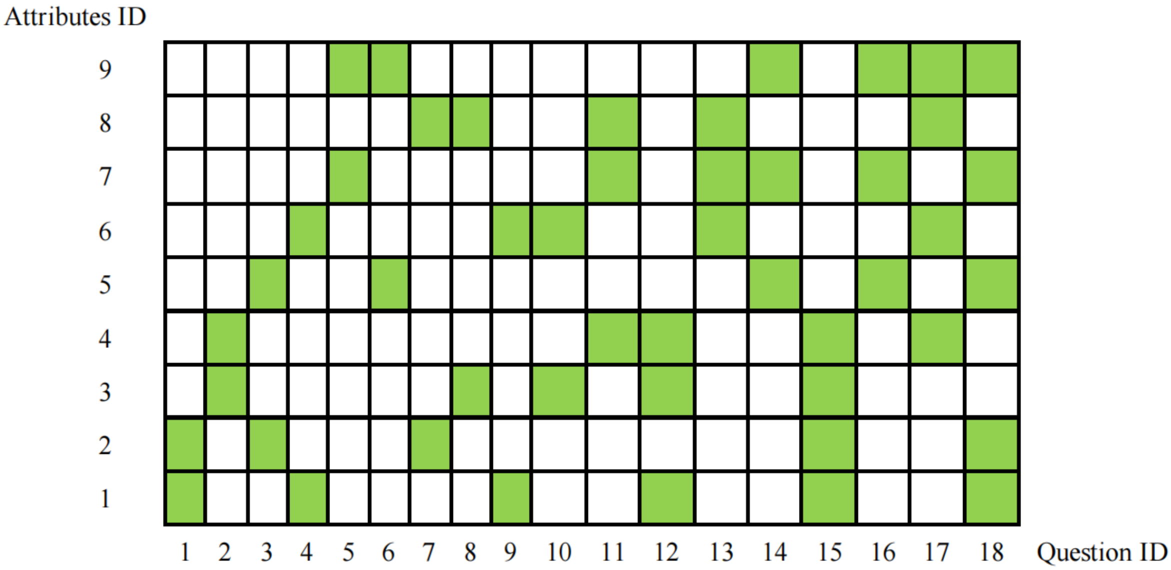

Figure 2. The distribution of attributes

Table 3 presents a detailed comparison of test performance between student A and student B across 18 questions, along with their corresponding binary representations in the R-matrix used for cognitive diagnosis. The table includes the question ID (1–18), maximum score for each question, and actual test score and binary value in the R-matrix for student A and student B. The student A scored 80/100, while Student B scored 79/100 totally, indicating comparable overall performance. A value of 1 is assigned in R matrix if a student's score meets or exceeds 60% of the question's maximum score. Both students performed well on explanation-type questions (Q1–Q5) and complex tasks (Q15–Q18).

Table 3. Test score of student A and student A

Question ID	Score	Student A		Student B	
		Test score	Value in R matrix	Test score	Value in R matrix
1	4	4	1	4	1
2	4	4	1	2	1
3	4	2	0	4	1
4	4	4	1	4	1
5	4	4	1	4	1
6	2	2	1	0	0
7	2	0	0	2	1
8	2	2	1	0	1
9	2	0	0	2	1
10	2	2	1	0	0
11	5	5	1	5	0
12	5	3	1	3	1
13	5	2	0	5	1
14	5	5	1	3	1
15	10	8	1	8	1
16	10	9	1	8	1
17	10	8	1	8	1
18	20	16	1	17	1
Total score	100	80		79	

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

Figure 3 visualizes the binary representations (0 or 1) of student A and student B in the R-matrix across all 18 questions. The horizontal axis lists the question IDs (1–18), and the vertical axis distinguishes between student A and student B. Both students achieved "1" for most questions, reflecting strong overall competence. From Question 6 to Question 10, multiple variations occur, indicates differences in mastery of specific attributes. The binary patterns underscore the utility of the R-matrix to assess strengths and weaknesses of student.

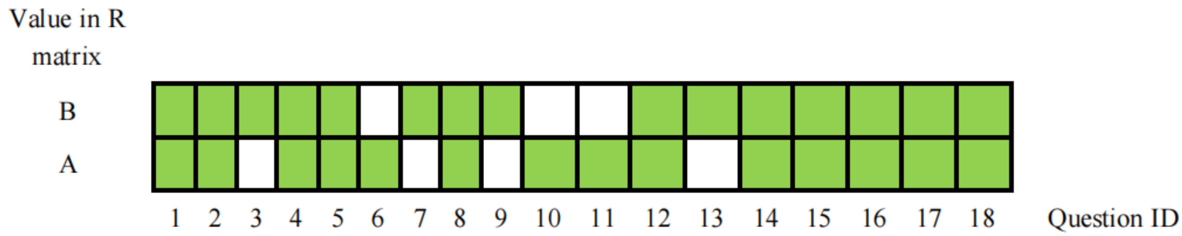

Figure 3. The distribution of attributes

4. Results and Discussion

Based on Q-Matrix cognitive diagnostic theory, the results are listed in table 4, and the radar charts are shown in figure 4 and figure 5. Although the two students achieved similar total scores on the final exam, both instructors and students have access to limited information. Without further analysis, these scores offer limited interpretability. By applying the Q-matrix theory for analysis, distinct differences in the competency profiles of the two students can be identified.

Student A demonstrates a significant advantage in the dimension of critical thinking, offering unique insights into the social impacts of gene editing technology. In the model construction task, they independently designed a demonstration device for genetic inheritance laws. Their lesson plan design shows systemic integration of ecological protection concepts, and their data analysis skills are reflected in the precise interpretation of population change statistical charts. This multi-dimensional strength indicates that the student has reached a high level of mastery in core knowledge areas such as the principles of genetic technology and ecosystem stability. They also display a strong interest in scientific inquiry activities, and their cognitive structure exhibits early characteristics of research-oriented thinking. However, a notable weakness exists in their understanding of the evolution-adaptation perspective, as their explanations of natural selection mechanisms remain superficial, lacking a deep grasp of the co-evolutionary relationship between organisms and their environment. In data analysis, while they can complete basic statistical tasks, they show a deficiency in interpreting advanced metrics such as the coefficient of variation. This discrepancy may stem from the student's learning strategy: an excessive focus on the vertical deepening of specialized knowledge in cutting-edge fields like genetic engineering, coupled with insufficient exposure to the humanistic aspects within the history of biology, leading to a lack of philosophical reflection on the nature of science. It is recommended that subsequent instruction strengthen the integration of the history of science and scientific ethics. Using classic cases such as "Mendel's Pea Experiments" instructors can guide students to understand the cultural context behind scientific discoveries, promoting the synergistic development of scientific spirit and humanistic literacy.

Student B's learning performance exhibits a marked differentiation in disciplinary competencies. This student shows outstanding advantages in the dimensions of the evolution-adaptation perspective and social responsibility. Their cognitive structure is deeply integrated with Darwin's theory of natural selection, enabling them to dialectically analyze the dynamic equilibrium of organism-environment interactions using fundamental principles such as survival of the fittest and genetic variation. The development of this higher-order thinking ability confirms the effective penetration of course instruction in the evolutionary biology

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

knowledge module, reflecting the student's profound understanding and solid grasp of the laws governing the origin and development of life. At the level of social responsibility awareness, Student B exhibits moral judgment surpassing that of their peers, capable of translating ecological ethics concepts such as biodiversity protection into concrete actions. However, the assessment data also reveals a structural deficiency in the student's practical application skills. In experimental design, issues such as ambiguous goal-setting and inappropriate method selection are present. Model construction skills are hampered by weaknesses in spatial imagination and abstract thinking. These weaknesses expose an imbalance in cultivating scientific inquiry competencies in the teaching process, suggesting a need to strengthen students' comprehensive practical skills through strategies like project-based learning.

In the three critical dimensions of critical questioning, structure-function perspective, and scientific ethics, both students performed well, with success rates exceeding 80%. This indicates that the teacher has achieved some success into these aspects. However, the assessment results also expose shortcomings in instruction. In the areas of model construction and the evolution-adaptation perspective, both students performed relatively poorly, with success rates significantly lower than in other dimensions. This weakness suggests that in the subsequent teaching evaluation, it would be beneficial to further integrate evaluation data from the entire class to establish a more comprehensive feedback mechanism. By analyzing the common needs of students in research methods and professional development strategies, instructors can adjust teaching strategies more effectively.

Table 4. Cognitive diagnostic results of student A and student B

	Student A	Student B
Critical questioning	0.90	0.85
Data analysis	0.63	0.84
Structure-function perspective	0.85	0.86
Experimental Design	0.81	0.61
Scientific ethics	0.95	0.89
Lifelong learning	0.9	0.66
Evolution adaptation perspective	0.63	0.90
Social responsibility	0.67	0.90
Model construction	0.90	0.61

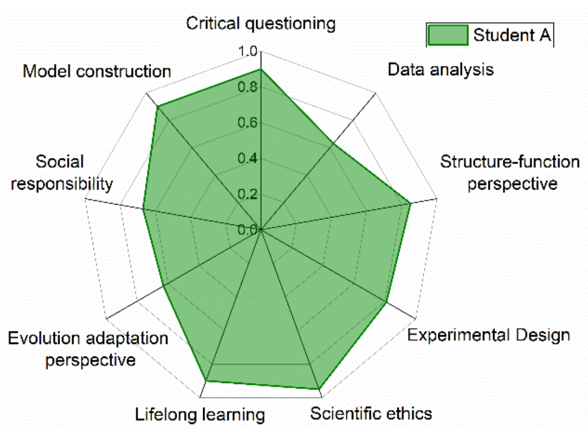

Figure 4. Cognitive Diagnosis of student A

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

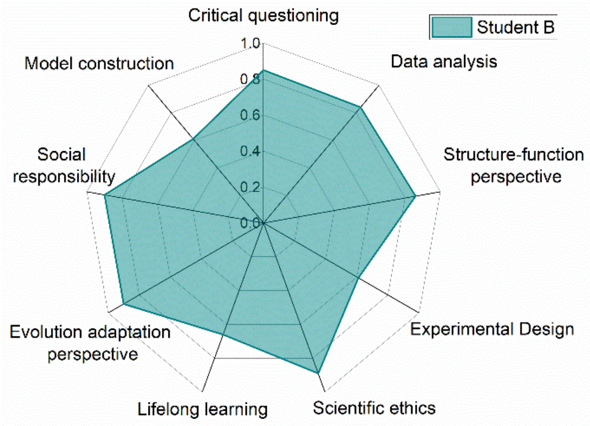

Figure 5. Cognitive Diagnosis of student B

Analyzing the results of the two students' performance using Q-matrix theory reveals their respective strengths and existing problems. This method allows for a detailed diagnosis of an individual student's development across various competencies. Therefore, the student competency evaluation scheme obtained through this computational method demonstrates strong characteristics and is well-suited for conducting individualized student assessment and diagnosis.

5. Conclusion

This study successfully demonstrates the effect and practical value of a Q-Matrix cognitive diagnostic theory framework for conducting in-depth, attribute-specific evaluation within the junior high school biology curriculum. The proposed model moves decisively beyond the limitations of total-score analysis by decomposing student performance into probabilistic estimates of mastery across nine core competencies. The case comparison of Students A and B serves as a compelling validation: despite nearly identical final grades, their cognitive diagnostic profiles revealed fundamentally different strengths and weaknesses. Student A's profile highlighted advanced critical thinking and technical design skills alongside a need for deeper conceptual understanding in evolution, while Student B exhibited strong theoretical and ethical reasoning but required support in practical application and model construction. These diagnostic outcomes directly translate into actionable insights for pedagogy, suggesting tailored interventions such as incorporating history-of-science narratives for Student A and implementing scaffolded, project-based inquiries for Student B.

In conclusion, this approach offers a robust, scalable, and theory-grounded tool for the formative assessment of complex, integrated learning outcomes. It represents a significant step toward realizing truly diagnostic assessment in science education, one that aligns evaluation with the holistic goals of developing knowledgeable, skilled, and ethically engaged learners.

Acknowledgements

This work was financially supported by no fund.

References

[1] M. L. Thomas, "The value of item response theory in clinical assessment: A review," Assessment, vol. 18, no. 3, pp. 291–307, 2011.

[2] J. C. Cappelleri, J. J. Lundy, and R. D. Hays, "Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures,"Clin. Therapeutics, vol. 36, no. 5, pp. 648–662, 2014.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202512_8(12).0041

[3] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Engaging with massive online courses," in Proc. 23rd Int. Conf. World Wide Web, 2014, pp. 687–698.

[4] H. Burns, C. A. Luckhardt, J. W. Parlett, and C. L. Redfield, Intelligent Tutoring Systems: Evolutions in Design, Hove, U.K.: Psychology Press, 2014.

[5] J. De La Torre, "DINA model and parameter estimation: A didactic," J. Educ. Behav. Statist., vol. 34, no. 1, pp. 115–130, 2009.

[6] S. E. Embretson and S. P. Reise, Item Response Theory, Hove, U.K.: Psychology Press, 2013.

[7] M.D. Reckase, "Multidimensional item response theorymodels," in Multidimensional Item Response Theory, Berlin, Germany: Springer, 2009, pp. 79–112.

[8] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, no. 8, pp. 30–37, 2009.

[9] Chiu, C.Y.; Douglas, J. A Nonparametric Approach to Cognitive Diagnosis by Proximity to Ideal Response Patterns. J. Classif. 2013, 30, 225–250.

[10] Chiu, C.Y.; Sun, Y.; Bian, Y. Cognitive Diagnosis for Small Educational Programs: The General Nonparametric Classification Method. Psychometrika 2017, 83, 355–375.

[11] de la Torre, J. The Generalized DINA Model Framework. Psychometrika 2011, 76, 179–199.

[12] Ma,W.; de la Torre, J. GDINA: An R Package for Cognitive Diagnosis Modeling. J. Stat. Softw. 2020, 93, 1–26.

[13] Zhan, P. Deterministic Input, Noisy Mixed Modeling for Identifying Coexisting Condensation Rules in Cognitive Diagnostic Assessments. J. Intell. 2023, 11, 55.

[14] Zhu Jinxin, Zhang Shumei, Xin Tao. A modelof cognitively diagnostic base on q matrix: classifying model of the probability of attributes' mastery [J]. Journal of Beijing Normal University (Natural Science), 2009, 45(2): 117-122.