

Diagnosing middle school students' proficiency in constructing scientific explanations with the integration of chemical reactions and patterns: a cognitive diagnostic modeling approach

Peng He, Yu Zhang, Tingting Li, Yuan Zheng & Jie Yang

To cite this article: Peng He, Yu Zhang, Tingting Li, Yuan Zheng & Jie Yang (13 Oct 2024): Diagnosing middle school students' proficiency in constructing scientific explanations with the integration of chemical reactions and patterns: a cognitive diagnostic modeling approach, International Journal of Science Education, DOI: [10.1080/09500693.2024.2413926](https://doi.org/10.1080/09500693.2024.2413926)

To link to this article: <https://doi.org/10.1080/09500693.2024.2413926>



Published online: 13 Oct 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Diagnosing middle school students' proficiency in constructing scientific explanations with the integration of chemical reactions and patterns: a cognitive diagnostic modeling approach

Peng He ^a, Yu Zhang ^b, Tingting Li ^a, Yuan Zheng^c and Jie Yang ^d

^aCollege of Education, Washington State University, Pullman, WA, USA; ^bQiaoyi Guyunhe Middle School, Wuxi, People's Republic of China; ^cFaculty of Education, Northeast Normal University, Changchun, People's Republic of China; ^dResearch Institute of Science Education, Beijing Normal University, Beijing, People's Republic of China

ABSTRACT

Scientific explanation has been widely recognised as one of the primary proficiencies in global science education. From research and practical perspectives, investigating the cognitive process of constructing scientific explanations by integrating core ideas is essential to make sense of phenomena or solve problems. This study applied the cognitive diagnosis modeling (CDM) approach to diagnose students' cognitive patterns of constructing scientific explanations by integrating chemical reactions and patterns. We localised three well-designed assessment tasks and correspondingly developed five cognitive attributes. Responses from 244 Grade 9 students in two middle schools were collected, scored, and analyzed. Results show that the five developed cognitive attributes are reliable and effective in identifying students' cognitive challenges in constructing scientific explanations in a content-specific domain. In addition, we articulated students' potential cognitive process of constructing scientific explanations. We identified a primary cognitive pathway that students may follow: making correct claims, finding sufficient evidence based on data patterns, and providing scientific principles for the reasoning process. The cognitive diagnosis results could be used to guide teachers in selecting teaching materials and strategies and arranging lesson sequences to support the development of students' scientific explanations by integrating chemical reactions and patterns.

ARTICLE HISTORY

Received 20 March 2023
Accepted 4 October 2024

KEYWORDS

Scientific explanation;
cognitive diagnostic
modelling; cognitive
attributes; cognitive
pathways

Introduction

Scientific explanations require students to provide written or oral responses to explain how or why a phenomenon occurs with supporting evidence and scientific knowledge (Gotwals & Songer, 2013; McNeill et al., 2006; Yeo & Gilbert, 2014). Constructing scientific explanations goes beyond descriptions of natural patterns to support students in

providing plausible explanations for real-world phenomena by seeking evidence and responding to the ideas or knowledge claims in science (Driver et al., 2000; Sandoval & Millwood, 2005; Schwarz et al., 2017). From an educational reform perspective, ‘constructing scientific explanations’ is widely recognised as one of the primary scientific proficiencies that all students should obtain in national science standards across countries such as the U.S. Next Generation Science Standards (NGSS, the Lead States, 2013) and the Chinese Compulsory Education Science Curriculum Standards (He et al., 2021, 2022; Ministry of Education, P. R. China, 2022). In the NGSS, students deepen their understanding of disciplinary core ideas (DCIs) and expand the connections with crosscutting concepts (CCCs) through participating in science and engineering practices (SEPs), such as constructing scientific explanations (the NGSS Lead states, 2013). Similarly, the Chinese Compulsory Education Science Curriculum Standards (He et al., 2021, 2022; Ministry of Education, P. R. China, 2022) also emphasise that ‘students should learn to collect evidence and make evidence-based reasoning when conducting scientific inquiry.’ From a research perspective, supporting and developing students’ proficiency in constructing scientific explanations has been investigated globally in the field of science education in the past two decades (e.g. McNeill et al., 2006; Sevian & Gonsalves, 2008; Tang, 2016; Yeo & Gilbert, 2014). So far, a substantial amount of research has explored to understand the framework of constructing scientific explanations (e.g. Sandoval, 2003; Yao & Guo, 2018), investigate the instructional strategies to support students in constructing scientific explanations (e.g. McNeill et al., 2006), and examine the learning progressions in constructing scientific explanations (e.g. Songer & Gotwals, 2012). However, challenges remain in uncovering students’ cognitive processes when they construct scientific explanations to explain real-world phenomena or solve problems, particularly how to integrate specific DCIs and CCCs in the cognitive process. It is unknown about the cognitive sequence of how students apply multiple dimensions (e.g. DCIs, CCCs, and SEPs) in their sensemaking or problem-solving process. More importantly, it is essential to diagnose where do students encounter a challenge cognitively when they construct a scientific explanation to address the real-world situations. Therefore, the field of science education calls for a deeper understanding of the cognitive process involved in constructing scientific explanations (Braaten & Windschitl, 2011; Osborne & Patterson, 2011).

Unpacking the cognitive features of constructing scientific explanations would provide teachers and students with diagnostic, elaborative, and concrete information to further support student scientific proficiency. More empirical studies are needed to unfold the underlying students’ cognitive process of constructing scientific explanations when they make sense of phenomena or solve problems. So far, researchers have employed a new measurement approach, cognitive diagnostic modeling (CDM; de la Torre & Minchen, 2014), to diagnose complex constructs, such as problem-solving (Li et al., 2020) and scientific explanations (Hu et al., 2021). CDMs are restricted latent class models in which the latent variables or attributes are discrete, usually dichotomous (de la Torre & Minchen, 2014). The existing studies mainly used two measurement approaches (e.g. item response theory and classical testing theory) to validate assessments of students’ scientific explanations and report evaluating students’ performances in constructing scientific explanations (e.g. Gotwals & Songer, 2013). However, constructing scientific explanations is a complex activity that cannot be performed without incorporating disciplinary core ideas

(e.g. chemical reactions) and crosscutting concepts (e.g. patterns), especially in explaining real-world phenomena (Gotwals & Songer, 2013; NGSS, 2013; Schwarz et al., 2017). This argument is also evident by reviewing previous studies (e.g. Chang et al., 2014; Cheng & Gilbert, 2017; Cooper et al., 2016; Talanquer, 2013; Weinrich & Talanquer, 2016; Yan & Talanquer, 2015) of investigating students' reasoning and visualisation of chemical reactions in describing, explaining, and predicting chemical phenomena.

This study explores the emerging features of middle school students' cognitive process when they construct scientific explanations with the DCI of chemical reactions and the CCC of patterns in the context of Chinese science classrooms. We conducted this study in China because the new Chinese science standards share the same vision with the NGSS, emphasising the importance of constructing scientific explanations by integrating DCIs and CCCs. In this study, we adopted well-designed assessment tasks for assessing students' constructing scientific explanations with the integration of chemical reactions and patterns. We analyzed students' responses to the tasks, developed cognitive attributes, and applied the CDM approach (Ma & de la Torre, 2016) to explore the essential cognitive features and emerging cognitive processes of middle school students' constructing scientific explanations. We anticipate this study will provide a deeper and more comprehensive understanding of students' science proficiency in integrating scientific explanations with DCIs and CCCs, which would further support student science learning and teacher classroom instruction. Our study's approach would provide research insights for science education researchers to explore the latent cognitive process of students' complex science learning.

Research questions

We explore three research questions (RQs):

RQ1. What are the validity and reliability of the assessment tasks and the associated cognitive attributes for diagnosing middle school students' proficiency in constructing scientific explanations with chemical reactions and patterns?

RQ2. What are the cognitive characteristics of middle school students' proficiency in constructing scientific explanations with chemical reactions and patterns?

RQ3. What are the emergent cognitive pathways of middle school students' proficiency in constructing scientific explanations with chemical reactions and patterns?

Literature review

This section reviews the relevant literature on scientific explanations and CDMs to contextualise our study. We further demonstrate how the previous research informs our design of the CDM approach to investigate the underlying cognitive attributes and potential cognitive pathways of constructing scientific explanations in content-specific domains.

Scientific explanations

Scientific explanation is considered an essential learning goal in science education to promote conceptual understanding and reasoning about mechanisms (Strike &

Posner, 1985). It is also a crucial component of scientific inquiry reforms that help students develop and support their explanations through their investigations (Driver et al., 2000; Duschl, 2000).

Constructing scientific explanations is a complex process involving multiple cognitive elements (McNeill et al., 2006). To better understand and support this complex process, researchers have approached the construct of scientific explanations from various perspectives, such as philosophical views (e.g. Hempel & Oppenheim, 1948), theoretical and empirical evidence (e.g. Braaten & Windschitl, 2011; McNeill et al., 2006; Songer et al., 2009), or educational standards documents (e.g. Ministry of Education, P. R. China, 2022; NGSS Lead States, 2013; NRC, 2012). Efforts have been made to address the challenge of supporting students in constructing scientific explanations through exploring scientific explanation frameworks (Braaten & Windschitl, 2011; NRC, 2012; Osborne & Patterson, 2011; Sandoval, 2003). For instance, McNeill et al. (2006) introduced the claim-evidence-reasoning (CER) model to help students 'justify their claims using appropriate evidence and scientific principles.' This model aligns with the nature of constructing scientific explanations, emphasising the explicit application of theories to reveal the causal relationships and underlying mechanisms of a specific situation or phenomenon (Braaten & Windschitl, 2011; NRC, 2012; Osborne & Patterson, 2011).

The emphasis in scientific explanation is applying students' knowledge to solve real-world problems rather than just memorising information to answer questions (Li et al., 2021; NRC, 2012; Pellegrino & Hilton, 2012). A scientific explanation comprises three elements: claim, evidence, and reasoning (McNeill et al., 2006; Songer & Gotwals, 2012). A claim is a testable statement about a phenomenon; evidence is the data that supports the claim, and reasoning is the justification that shows why the evidence supports the claim using scientific principles (Gotwals et al., 2012). The nature of constructing scientific explanations is creating a claim or statement and making a reasoning process to show how evidence and scientific principles can be used to support a proposed claim. The structure of constructing scientific explanations requires higher demands for assessment. A handful of studies have applied large-scale assessment data to measure students' scientific explanation proficiencies in content-general domains (Hu et al., 2021; Kim et al., 2015; Zhan et al., 2019). This study addresses the gap in previous studies that aims to diagnose student scientific explanations in content-specific domains that incorporate the specific disciplinary core ideas and crosscutting concepts, such as chemical reactions and patterns.

Cognitive diagnostic assessment

Constructing scientific explanations is an essential science practices for students to apply what they have learned to solve real-world problems, not just memorise information to answer questions (McNeill et al., 2006). Despite its benefits, constructing scientific explanations is challenging for students and teachers (He et al., 2023a), requiring teacher-appropriate instructional support. However, teachers may need diagnostic information on student performance on assessment tasks to inform their instructional decisions (He et al., 2024). So far, a handful of studies have applied Item Response Theory (IRT) approaches to assess student proficiency in constructing scientific explanations

(e.g. He et al., 2023b; Songer & Gotwals, 2012; Yao & Guo, 2018), ranking a student's achievement against other students on a single proficiency continuum. The IRT-based assessments fail to provide diagnostic information for offering specific student feedback and informing teacher instructions. Compared to IRT, CDMs identify fine-grained cognitive attributes based on the measurement constructs (e.g. scientific explanations) and classify students' proficiencies based on their mastery of skills and attributes of interest (Ma & de la Torre, 2016). Notably, using the CDM approach in analyzing classroom assessments would offer deeper insights into the cognitive process of constructing scientific explanations and provide more detailed diagnostic information to inform teachers' instructional decisions and support students' development of scientific explanation proficiency.

Theoretical considerations

CDMs focus on identifying and diagnosing the underlying cognitive attributes required to complete specific tasks (e.g. de La Torre, 2009; DiBello & Stout, 2007; Tatsuoka, 2009). CDMs can bridge item response theory (IRT) and cognitive psychology by creating probabilistic models that reflect the cognitive processes involved in task performance (Mislevy, 1996). The CDM approach aligns with the vision of converting experts' cognitive models of a domain into probabilistic models for learners' observed outcomes.

CDMs were developed in response to Mislevy's (1994) call for more robust integration between cognitive science and measurement practices, significantly informed and supported by Knowledge Space Theory (KST, Doignon & Falmagne, 1999). KST conceptualises learners' knowledge as discrete states within a structured system, facilitating predicting observable responses based on a learner's current knowledge state. KST's representations of learning as discrete and structured nature of knowledge, enabling CDMs to diagnose and assess cognitive attributes and their interdependencies more effectively. This integration enhances the precision and diagnostic power of psychometric models, aligning with the goals of CDMs to reflect complex cognitive processes accurately. Thus, KST serves as a crucial cognitive science theory that underpins and enriches the theoretical consideration of CDMs. By leveraging KST's detailed representation of knowledge structures and cognitive pathways, CDMs can develop more accurate and practical diagnostic assessments, aligning with Mislevy's (1994) call for models that faithfully reflect the knowledge, skills, and abilities they measure and the practical demands of their use.

Applications of CDMs in science assessment

Applying CDMs in assessing students' scientific proficiency is a growing area in science education. Cognitive diagnostic assessment (CDA) offers a comprehensive approach to diagnosing students' learning by presenting attribute mastery patterns (AMPs) as their cognitive features. A series of attribute mastery patterns can be ordered sequentially to represent students' potential cognitive pathways toward a specific learning goal using corresponding assessments (e.g. Hu et al., 2021; Wu et al., 2021). Studies using large-scale datasets, such as PISA or TIMSS, increasingly examine students' scientific literacy through cognitive diagnostic assessments (e.g. Hu et al., 2021; Kim et al., 2015; Zhan et al., 2019). The cognitive attributes in these studies are either content-oriented (e.g. thermochemistry, Chen et al., 2017) or practice-oriented (e.g. scientific explanations,

Hu et al., 2021; scientific argumentation, Zhai et al., 2022). For instance, Hu et al. (2021) applied CDMs to diagnose elementary students' scientific explanations in six countries (China, the United States, Singapore, Australia, the United Kingdom, and Russia). Their study developed general cognitive attributes of constructing scientific explanations, such as observing and describing the phenomenon, obtaining and analyzing data, using facts, constructing reflections, systematic use of theory, and scientific reasoning. A typical cognitive pathway in their study was found to describe the phenomenon, construct reflections, scientific reasoning, obtain data, use facts, analyze data, and observe the phenomenon. Hu et al. (2021) also reflected that elementary students' ability to construct scientific explanations relates to their understanding of content knowledge (e.g. energy and force and motion). Their cognitive attributes of constructing scientific explanations are domain-general without consideration of specific DCIs and CCCs.

Therefore, this study highlights that students' proficiency in constructing scientific explanations must be distinct from the specific DCIs and CCCs when students make sense of phenomena or solve problems. To address this, this study introduces cognitive attributes focusing on three-dimensional constructs (NGSS Lead States, 2013), including the SEP of constructing scientific explanations, the DCI of chemical reactions, and the CCC of patterns. We also explore the potential cognitive pathways using the three-dimensional cognitive attributes.

Methods

To address the research questions, we employed an evidence-centered design (ECD; Mislevy & Haertel, 2006) approach to create assessment tasks for diagnosing students' proficiency in constructing scientific explanations that integrate chemical reactions and patterns (Harris et al., 2019). We then assessed the psychometric properties of these tasks. Subsequently, we identified cognitive attributes for constructing scientific explanations incorporating chemical reactions and patterns. Finally, we validated these cognitive attributes using the CDM approach and explored potential cognitive pathways for students' constructing scientific explanations involving chemical reactions and patterns.

Instrumentation

Assessment tasks

We adopted three assessment tasks from the Next Generation Science Assessment project to assess middle school students' constructing scientific explanations with the integration of chemical reactions and patterns (NGSA, 2024). All three tasks were developed using an ECD design process (Mislevy & Haertel, 2006), including unpacking three dimensions from NGSS, articulating learning performance goals, determining evidence statements and task design features, and developing tasks and rubrics (Harris et al., 2019). Such a systematic design process ensured the tasks' content validity. In addition, during the iterative design process, several rounds of internal and external expert reviews ensured the tasks' face validity. So far, these tasks have been widely used in science classrooms across countries. For this study, we translated and modified them by replacing the original names with Chinese names and preserving their key features, such as task

scenarios and data information. We further modified the prompts in the three tasks and provided entire scaffolding for constructing scientific explanations, such as claims, evidence, and reasoning (McNeill et al., 2006). We conducted this way because Chinese students have yet to experience practice-based instruction in everyday science classrooms. All three tasks were structured using the same three prompts (see Figure 1 and Appendix A). The three prompts ask students to make an appropriate claim (prompt 1), cite sufficient evidence that would support their claims (prompt 2), and connect the claim and evidence with the DCI of chemical reactions and the CCC of patterns in the reasoning process (prompt 3). Two graduate assistants and three collaborative teachers reviewed the tasks. They provided feedback on the appropriateness of language used

Xiao Fan experimented using ammonia and hydrogen chloride. She poured the hydrogen chloride and ammonia solutions into different beakers. She could tell by the strong smell from each beaker that ammonia gas was coming from one and hydrogen chloride gas from the other. When she placed the two beakers close together, she observed a white cloud of tiny particles form where the gases mixed. This is shown in the video: Mixing ammonia and hydrogen chloride gas. She asked a lab student to test and calculate the density of ammonia gas, hydrogen chloride gas, and the white substance. The data are in Table 1.

Table 1. Data of substances before and after mixing

Gas	Appearance	Odor	Density
Ammonia	Colorless gas	Strong smell	0.77 g/cm ³
Hydrogen Chloride	Colorless gas	Strong smell	1.49 g/cm ³
White substance (formed after two gases mixed)	White fume	Odorless	1.53 g/cm ³

Use the data in Table 1 to support your answers to the three questions below.

Question #1

Write a claim relating ammonia and hydrogen chloride before mixing and the white substance formed after mixing to state if a chemical reaction occurred.

Question #2

Show evidence to support your claim.

Question #3

Give reason(s) that the evidence you use supports your claim.

Figure 1. An example of constructing scientific explanations tasks (Task 1).

Note: the task was modified from *Tiffany's strong smelling beakers* (ID#: 080-02-c02) in the NGSA project (2024). Access the original task via the link: <https://ngss-assessment.portal.concord.org/resources/200/tiffany-s-strong-smelling-beakers-id-080-02-c02>

in the task scenarios and the clarity and explicitness of the three prompts, which ensured the face validity of the three tasks in the Chinese version. [Figure 1](#) shows the three assessment tasks (Task 1). Also, see the other two tasks in Appendix A.

Scoring rubrics

We developed analytic rubrics for the three tasks for scoring students' written responses to constructing scientific explanations (He et al., 2023b). For diagnostic purpose, we created five rubric criteria to score students' responses to the three prompts. We divided all aspects of evidence and reasoning into single criteria (see the five aspects in [Table 1](#)). The binary score approach (1 or 0) was used to score each specific item based on the criteria in the rubrics. If the student's response meets the criterion, we scored 1; otherwise, we scored 0. Instead of using a holistic score, we decided to use a set of binary sub-scores for two reasons. First, a set of binary sub-scores can tease the criteria in detail, offering specific and sufficient information about diagnosing students' proficiency (He et al., 2023b). Second, we adopted the binary scoring approaches from previous CDM studies (e.g. Hu et al., 2021; Wu et al., 2021; Zhai et al., 2022), which would be more applicable to CDM models. [Table 1](#) provides a scoring example. Based on the five criteria in the analytic rubric, we scored the student's response and provided the points as (1, 0, 1, 0, 1). The student missed the points on Items 1.2-1 and 1.3-1 because the response needed to include information about the odour data and whether the odour is a property.

Cognitive attributes

Cognitive attributes play a crucial role in cognitive diagnostic assessment. Cognitive attributes are invisible cognitive states hidden behind assessment tasks that are difficult to observe directly. Students' mastery can be judged by their responses to specific items in assessment tests. Therefore, the quality of attributes is critical and directly determines the effectiveness of cognitive diagnosis (Wu et al., 2020). We unpacked the five attributes (see [Table 2](#)) based on the NGSS (NGSS Lead States, 2013), the Chinese new science

Table 1. The rubric and scoring examples for Task 1.

Student Response:			
1. Yes, a chemical reaction occurs when mixing ammonia and hydrogen chloride together.			
2. The density of ammonia, hydrogen chloride, and white substances is different, which indicates that white substances are new substances.			
3. I used density because it is a property that can be used to identify whether substances are the same or not. A new substance produced indicates a chemical reaction occurs.			
Questions	Scoring items	Criteria	Scores (0/1)
1.1	1.1-1	The student states that a chemical reaction occurs.	1
1.2	1.2-1	The student describes the pattern (comparing before and after) in odour data as evidence to support the claim.	0
	1.2-2	The student describes the pattern (comparing before and after) in density data as evidence to support the claim.	1
1.3	1.3-1	The student indicates that odour is a characteristic property used to identify substances, and a change in odour indicates that a new substance has been produced, that is, a chemical reaction has occurred.	0
	1.3-2	The student indicates that density is a characteristic property used to identify substances, and a change in density indicates that a new substance has been produced, that is, a chemical reaction has occurred.	1

Table 2. The cognitive attributes and their descriptions

Code	Attribute	Description
A1 (SEP-Claim)	Build an appropriate claim	Students can build an appropriate claim to explain the phenomenon or real-world problem (e.g. whether a chemical reaction occurs).
A2 (SEP-Evidence)	Use evidence/scientific data to support the claim	Students can use sufficient data (changes in density, solubility, and/or odour) as evidence to support their claims.
A3 (SEP-Reasoning)	Construct reasoning to justify claim with evidence	Students can construct a reasoning process that connects the data of characteristic properties (e.g. density, odour, and solubility) to the claim (e.g. whether a chemical reaction occurs).
A4 (DCI-Chemical Reaction)	Apply the ideas of characteristic properties to identify whether a chemical reaction occurs or not	Students can apply the ideas of characteristic properties of substances (e.g. density, odour, solubility) and changes in properties to identify new substances and determine whether a chemical reaction occurs.
A5 (CCC-Patterns)	Identify the patterns (the similarities and differences) of data	Students can identify the similarities and differences in data on the properties of substances such as density and solubility.

standards (Ministry of Education, P. R. China, 2022), and the relevant literature of scientific explanations (McNeill et al., 2006; Songer & Gotwals, 2012). These five cognitive attributes include three SEP attributes related to constructing science explanations: SEP-Claim (A1), ‘Build an appropriate claim,’ SEP-Evidence (A2), ‘Use evidence/scientific data to support the claim,’ and SEP-Reasoning (A3) ‘Construct reasoning to justify claim with evidence’; one DCI attribute: DCI-Chemical Reactions (A4) ‘Apply the ideas of characteristic properties to identify whether a chemical reaction occurs’; and one CCC attribute: CCC-Patterns (A5) ‘Identify the patterns of data similarities and differences.’ These cognitive attributes were designed by unpacking the ‘Chemical Reactions’ (using characteristic properties of substances to identify new substances and determining whether a chemical reaction occurs), the subcomponents of the ‘Constructing scientific explanations’ practices based on the Claim-Evidence-Reasoning framework (McNeill et al., 2006; Songer & Gotwals, 2012), and unpack the crosscutting concept ‘Patterns’ (similarities and differences in data). Table 2 displays the cognitive attributes and their descriptions, which are expressed as dichotomous categories that reflect the examinee’s latent cognitive state.

Q-matrix

The Q matrix is a two-way specification table that links student responses with their cognitive states by indicating the specific attributes in each item (Wu et al., 2021). The Q-matrix table provides a straightforward representation of the relationship between items, categories, and attributes (Tu et al., 2019). Therefore, we created a Q matrix for the three assessment tasks, mapping the items to the attributes (See Table 3). The matrix assigns a ‘1’ to items that measure the attribute and a ‘0’ to those that do not. For instance, item 1.1-1 only contains A1 (SEP-Claim), given that it is designed to assess a student’s claim (which is also reflected as a cognitive attribute). For another example, item 1.3-1 contains four attributes (A1, A2, A3, and A4) because the item requires students to connect evidence and their claims using appropriate core ideas. We had external experts review the attribute assignments before we administrated our tasks.

Table 3. Q Matrix.

Task	Item	Attributes				
		A1 (SEP-Claim)	A2 (SEP-Evidence)	A3 (SEP-Reasoning)	A4 (DCI-Chemical Reaction)	A5 (CCC-Patterns)
Task 1	1.1-1	1	0	0	1	0
	1.2-1	1	1	0	0	1
	1.2-2	1	1	0	0	1
	1.3-1	1	1	1	1	0
	1.3-2	1	1	1	1	0
Task 2	2.1-1	1	0	0	1	0
	2.2-1	1	1	0	0	1
	2.2-2	1	1	0	0	1
	2.3-1	1	1	1	1	0
	2.3-2	1	1	1	1	0
Task 3	3.1-1	1	0	0	1	0
	3.2-1	1	1	0	0	1
	3.2-2	1	1	0	0	1
	3.3-1	1	1	1	1	0
	3.3-2	1	1	1	1	0

Participants and data collection

Two hundred and fifty Grade 9 students from two middle schools participated in this study. One school ($N = 132$) is in the northern region of mainland China, and another school ($N = 118$) is located in the southern part of mainland China. Even though the two schools are from different provinces, their students come from middle-class families in urban areas and have a wide range of academic achievement. After removing four incomplete data, we obtained the final analytic sample of 244 students' responses on our assessment test. 47.6% of the final sample were female students, whereas 52.4% were male. As mentioned above, like the NGSS, constructing scientific explanations is one of the essential science practices in the new Chinese science curriculum standards (He et al., 2021, 2022; Ministry of Education, P. R. China, 2017). Despite the practice-based science being recently introduced to China's science classrooms, the students who participated in our study had already experienced it because their teachers had implemented our CER-based curriculum in their classrooms before. The three tasks' assessment test (Chinese version) was administered using paper-pencil format. Because the primary purpose of this study was validating the cognitive attributes for diagnosing students' proficiency in constructing scientific explanations, we administered the test after students had completed their chemistry curriculum at the end of the middle school level. Forty-five minutes was allocated for students in a typical class period, sufficient for all students to complete the assessment test. We explained the purpose of this study to the students and their teachers before we collected their responses on our assessment test. Informed consent was obtained from all participants involved in the study. Their personal information was confidential and anonymous.

Scoring

We recruited three graduate research assistants who obtained a bachelor's degree in chemistry to score students' responses to the three tasks. They were enrolled in a full-time Master's chemistry teacher preparation programme at a national teacher education

university in China. Participating in our research project, they had sufficient experience designing assessments, creating rubrics, and scoring students' responses on open-ended tasks. The scoring process consists of a training stage and a formal scoring stage (He et al., 2023a). In the training stage, the first and second authors had a 3-hour meeting with them. They introduced assessments and rubrics to the three raters and also provided scored examples for deeper understanding. Then, we prepared a training dataset and randomly selected 10% of students ($N = 30$) from the sample for their scoring practices. We let three raters work on the training dataset separately within a week. After they completed, we organised a 3-hour meeting with them to discuss any discrepancies. The raters achieved a reasonable consensus in the training stage, leading to the formal scoring stage. In the formal stage, we overlapped another 10% of students to check the inter-rater reliability and evenly assigned them the rest of the data so they could work individually. We used Fleiss's kappa (Fleiss & Cohen, 1973) to measure the inter-rater reliability (IRR). Fleiss's kappa is a statistical measure used to assess agreement reliability between more than two raters when assigning scores to several items, which fit our scoring purpose. We calculated the final IRRs for each task (Task 1 is 0.816; Task 2 is 0.792; and Task 3 is 0.942), with an average of 0.84, indicating the three raters' scores are consistent. The final data is reliable for further CDM analysis.

Cognitive diagnostic model

The CDM is a psychometric approach incorporating cognitive psychology to diagnose and evaluate students' internal psychological processing, providing comprehensive information about cognitive diagnosis. The core assumption is that a student's response meeting the criteria in a task depends on having mastery of a specific set of latent skills and knowledge (i.e. cognitive attributes). By asking students to attempt to respond carefully to sets of well-designed tasks and then analyzing their responses, researchers can make detailed inferences about the attributes they have and have not mastered (Leighton & Gierl, 2007). Instead of producing a total score (or multiple scores) or a grade level, CDMs assign students into discrete subgroups based on their mastery of the different attributes being measured in the test (Ma & de la Torre, 2016). In doing so, CDMs provide fine-grained information about individual students' proficiency profiles and the identification of latent subgroups in the targeting population. Therefore, to ensure the cognitive diagnosis is valid and reliable, selecting an appropriate CDM based on the theoretical and empirical model-data fit is critical.

To date, researchers have developed a variety of CDMs, which can be divided into two categories according to the parameter type. One is the reduced cognitive diagnostic model such as DINA (Deterministic Input, Noisy 'And' Gate; de La Torre, 2009), DINO (Deterministic Input, Noisy 'Or' Gate; Templin & Henson, 2006), rRUM (reduced Reparameterized Unified Model; Hartz, 2002), LLM (Linear Logistic Model; Maris, 1999), and ACDM (Additive Cognitive Diagnostic Model; de La Torre, 2011), which contain only single attribute parameters but not interaction parameters between multiple attributes. Another category is the saturated cognitive diagnostic model, such as the G-DINA (generalized DINA; de La Torre, 2011) model, which contains not only all single-attribute parameters but also multi-attribute interaction parameters. That is, the probability of a student's response meeting the criteria in our rubrics is

affected by the main and interaction effects. In addition, according to the complexity of the relationship between attributes, the models can be divided into compensated, partially compensated, and non-compensatory models. Simply put, the compensation model, such as DINO and LLM, refers to the fact that each cognitive attribute contributes differently to the correct answering item, and even if only some cognitive attributes are mastered, it is possible to answer the item correctly. The non-compensation model means that the subjects can only answer the item correctly if they have mastered all cognitive attributes, such as the DINA model.

Model selection

Before reporting the results to the three RQs, we processed the model selections using the model fitting test (Tatsuoka, 1984). The model fitting estimation is mainly carried out from three aspects: test (i.e. our assessment test with three tasks), item (i.e. our criteria), and subjects (i.e. middle school students). The test fitting is mainly used to examine the overall fit of the selected model and the data to ensure the applicability of the selected model and the accuracy of the diagnostic classification of the subjects. The item fitting test examines the fit between the item and the selected model, illustrating the item's quality. The subject fitting examines whether the selected model suits the test subjects.

We utilised the FlexCDMs platform (Tu, 2019) to determine the most appropriate CDM using relative fit indices such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviation. The Deviation measures the discrepancy between the model and reality, with a lower Deviation indicating a better fit. The AIC and BIC were used to assess the goodness of fit of CDM models, with a smaller value indicating a better fit of the data to the model (Vrieze, 2012). As indicated in Table 4, the LLM was the best-fitting model compared to other models such as DINA, DINO, and ACDM. We decided to choose the LLM also because the selected model meets our theoretical assumptions: (a) each cognitive attribute would contribute to the likelihood of constructing scientific explanations (Zhai et al., 2022); (b) the select model incorporates logistic regression principles to relate the probability of a student's response to the presence or absence of specific cognitive attributes. Based on our results (see Table 4), the LLM fits well to validate the cognitive attributes of students' proficiency in constructing scientific explanations.

Analytic strategies

Accordingly, this study reports the results of three RQs using the outputs from the LLM analysis. For RQ1, we first analyzed the items using the Root Mean Square

Table 4. Parameter comparisons in different models.

Models	Deviation	AIC	BIC
DINA	3103.43	3225.43	3440.23
DINO	3127.02	3249.02	3463.83
rRUM	2870.41	3058.41	3389.42
LLM	2782.44	2970.44	3301.46
ACDM	2958.12	3146.12	3477.13
GDM	3000.38	3158.38	3436.58
LCDM	2860.78	3204.78	3810.47
G-DINA	2672.66	3046.66	3705.17

Note. LLM is the final model used in this study.

Error of Approximation (RMSEA) as the absolute fit index (Oliveri & von Davier, 2011) and subject fitting using the L_z statistic index (Cui & Li, 2015). The RMSEA fit statistics of items examine how well each item fits within the overall CDM model. The L_z statistic can be used to flag students with response patterns that are highly inconsistent with their estimated attribute mastery profiles. We then estimated the item discrimination (Wang et al., 2018) and the reliability of cognitive attributes (Templin & Bradshaw, 2013). We assessed the classification accuracy and consistency of the attributes, as reported by Cui et al. (2012). This is a critical indicator to evaluate the reliability and validity of classification results in cognitive diagnostic assessment.

For RQ2, we reported the probability of attribute mastery and attribute mastery patterns (AMPs). An AMP explicitly represents whether the student has mastered each attribute, where '1' signifies mastery and '0' signifies non-mastery. Based on the LLM analysis, we calculated the number of emerging AMPs from our sample and presented the AMPs with a number over 5. Regarding individual students' cognitive diagnosis, the FlexCDMs platform (Tu, 2019) plotted a diagram to demonstrate a student's AMP using the probability of mastering the five cognitive attributes. Such analyses demonstrate an in-depth understanding of students' cognitive attributes and provide individual formative feedback for students (Tatsuoka, 2009). This study presented several diagrams of students' cognitive structures with the same total score.

Regarding RQ3, we articulated the cognitive pathways based on the relationship between different AMPs, connecting all possible relationships to form a diagram of potential cognitive pathways for the sampled students. Cognitive pathways refer to the hierarchical arrangement of AMPs, characterised by partial order relationships between them (Tatsuoka, 2009; Wu et al., 2021). In this study, the diagram of potential cognitive pathways depicts students' gradual acquisition of the number of attributes from 0 (none) to 5 (complete mastery). For instance, students with the AMP of 10010 have mastered two attributes (A1 and A4), whereas students with the AMP of 11010 have mastered three attributes (A1, A2, and A4). The AMP of 11010 contains the AMP of 10010, so a connection is articulated from the AMP of 11010 to the AMP of 10010. This study articulated all potential connections between AMPs to articulate a diagram of potential cognitive pathways.

Results

RQ1. Psychometric properties of the assessment tasks and the associate cognitive attributes

Item fit statistics

Based on the LLM analysis, we estimated the fit of the items using RMSEA. A critical value of 0.1 was set for the RMSEA index (Oliveri & von Davier, 2011). A value of RMSEA less than 0.1 indicates a good fit for the item, while a value greater than 0.1 implies a poor fit. Table 5 presents the RMSEA values of the 15 items calculated using the LLM model. Except for items 1.3-1 and 3.2-1, all other items showed RMSEA values less than 0.1, with an average RMSEA value of 0.0685, indicating a good fit of the items with the LLM model.

Table 5. RMSEA of each item.

Item	RMSEA	Item	RMSEA	Item	RMSEA
1.1-1	0.0481	2.1-1	0.0464	3.1-1	0.0300
1.2-1	0.0900	2.2-1	0.0684	3.2-1	0.1100
1.2-2	0.0648	2.2-2	0.0730	3.2-2	0.0546
1.3-1	0.1025	2.3-1	0.0420	3.3-1	0.0735
1.3-2	0.0704	2.3-2	0.0886	3.3-2	0.0654

We tested the subject fitting using the L_z statistic index (Cui & Li, 2015). A value greater than -2 indicates that the subject's response aligns well with the model. Results showed that 97.6% of the students' responses fit well with the LLM model, ensuring the cognitive attribute estimates are valid and reliable.

Item discrimination

Cognitive diagnostic assessment evaluates the accuracy of cognitive attributes and the quality of test items through item discrimination (Wang et al., 2018). Table 6 presents the discriminations of all 15 items, ranging from 0.279–0.999. A more considerable value indicates a higher discrimination. The results indicate that most items can distinguish students with different attribute profiles.

Reliability of cognitive attributes

To evaluate the reliability of the cognitive attributes, we calculated the test-retest consistency of the five specific attributes (A1–A5 in Table 2) based on Templin and Bradshaw's (2013) approach. The results showed that the five attributes were highly reliable, with scores of 0.9779 (A1), 0.9686 (A2), 0.9931 (A3), 0.9805 (A4), and 0.9079 (A5), yielding an average reliability of 0.9656.

The classification accuracy at the attribute level for the five attributes ranged from 0.73–0.77, which indicates that they are reliable in diagnosing the integration of constructing explanations with scientific ideas.

RQ2. The cognitive characteristics of middle school students' proficiency

To answer RQ2, this section presents students' cognitive characteristics, including attribute mastery probability, AMPs, and cognitive structure analysis.

Attribute mastery probability

In terms of attribute mastery and non-mastery probabilities (See Figure 2), the students performed best in the attributes of A1, 'Build an appropriate claim,' and A5, 'Identify patterns (the similarity and differences) of data,' with mastery probabilities over 75%. Their

Table 6. Item discrimination.

Item	Discrimination	Item	Discrimination	Item	Discrimination
1.1-1	0.279	2.1-1	0.632	3.1-1	0.450
1.2-1	0.662	2.2-1	0.943	3.2-1	0.874
1.2-2	0.975	2.2-2	0.413	3.2-2	0.447
1.3-1	0.840	2.3-1	0.999	3.3-1	0.867
1.3-2	0.938	2.3-2	0.892	3.3-2	0.973

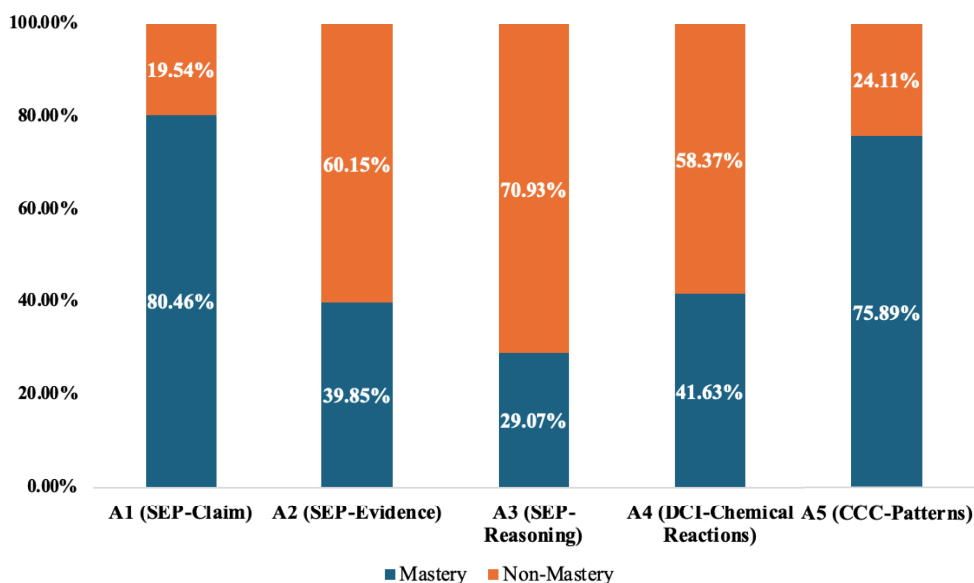


Figure 2. The mastery and non-mastery probability of the five attributes.

performance in the DCI attribute of A4, ‘apply the ideas of characteristic properties to identify whether a chemical reaction occurs or not,’ was moderate, with a mastery probability of about 41.63%. Performance was poor on A2, ‘Use evidence/scientific data to support the claim’ (39.85%), and A3, ‘Construct reasoning to justify claim with evidence.’ These results align with previous findings that students often struggled with fully comprehending how to use appropriate or sufficient evidence to support their claims (Sandoval, 2003; Sandoval & Millwood, 2005). Additionally, students struggled with reasoning and failed to adequately support their claims (Jiménez-Aleixandre et al., 2000).

Attribute Mastery Patterns

Theoretically, 32 Attribute Mastery Patterns (AMPs) should correspond to five cognitive attributes. However, the LLM analysis indicated that 244 students were grouped into 20 AMPs. We grouped the AMPs based on the number of mastery attributes (see Table 7). Most of the students concentrated on seven AMPs (see the Bold Font in Table 7), namely AMPs 2, 3, 6, 14, 15, 18, and 20, with the highest two probabilities being ‘10001’ (19.26%) and ‘11011’ (10.25%), respectively. These AMPs indicate that the students experienced challenges in providing adequate evidence to support their claims and using scientific principles in their reasoning.

A closer examination of the total number of AMPs in Table 7 revealed that 8.61% of the students (AMP2, ‘10000’) were able to claim that ‘a chemical reaction had taken place’ due to their mastery of attribute A1 but did not master any other attributes. 19.26% of the students (AMP6, ‘10001’) had the potential to locate evidence for claims by analyzing data patterns because of their expertise in attributes A1 and A5. Another 9.84% of the students (AMP3, ‘00011’) thoroughly understood the core concepts of chemical reactions because of their mastery of attributes A4 and A5. 8.61% of the students (AMP15, ‘11001’)

Table 7. The attribute mastery patterns.

Category	ID	AMP	Frequency (N)	Percentage (%)
0	1	00000	4	1.64
1	2	10000	21	8.61
2	3	00011	24	9.84
	4	01001	3	1.23
	5	01010	5	2.05
	6	10001	47	19.26
	7	10010	11	4.51
	8	10100	2	0.82
	9	11000	7	2.87
3	10	00111	11	4.51
	11	01011	2	0.82
	12	10011	3	1.23
	13	10101	2	0.82
	14	10110	16	6.56
	15	11001	21	8.61
	16	11010	5	2.05
4	17	01111	6	2.46
	18	11011	25	10.25
	19	11101	10	4.10
5	20	11111	19	7.79

Note: the five-digit in the AMP column (3rd column) indicate the presence and absence of the five cognitive attributes (A1, A2, A3, A4, and A5). For instance, AMP3 is 00011, indicating that students with the AMP3 had mastered A4 and A5 but not mastered A1, A2, and A3.

were able to provide sufficient evidence for claims by analyzing data patterns based on their mastery of attributes A1, A2, and A5. 6.56% of the students (AMP14, '10110') could reason about claims using partial evidence, having mastered attributes A1, A3, and A4, but could not provide adequate evidence. 10.25% of the students (AMP18, '11011') knew how to reason about claims using evidence, rooted in their mastery of attributes A1, A2, A4, and A5, but the reasoning process was not flawless. 7.79% of the students (AMP20, '11111') had mastered all attributes and possessed complete scientific explanation competencies.

Cognitive structure analysis

Figure 3 presents the pentagon diagrams of the four students' cognitive structures with the same total score (i.e. 10 points). Taking StudID8 as an example, the pentagon diagram shows the student's probabilities of mastering the five cognitive attributes. Points close to the centre of the diagram indicate less mastery for the attribute and vice versa. It is clear that the four students obtained the same total score but exhibited varying structures of cognitive attributes. For instance, StuID8 is highly likely to master all attributes except for A3 and A4. StuID85 demonstrates mastery of attributes 1, 4, and 5 but needs help with attributes 2 and 3. Meanwhile, StuID192 has a near-1 mastery probability for attributes A1, A2, and A4 but only a superficial understanding of attribute A5. StuID143 exhibits mastery of attributes A1 and A4, with a probability of approximately 0.5 for the mastery of attributes A2, A3, and A5, indicating a basic understanding but not complete mastery. The above exemplar analysis indicates that even if students obtained the identical total scores, their cognitive mastery of attributes may vary. The above exemplar analysis also supports using CDMs to diagnose students' constructing scientific explanations, which would offer meaningful information to teachers' classroom instruction.

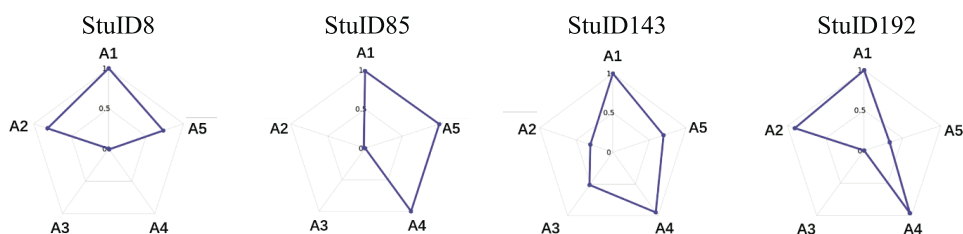


Figure 3. Comparison of different cognitive structures with the same total score.

Note. A1 refers to cognitive attribute 1, and so on and so forth. Cognitive structure refers to the probability of mastering each cognitive attribute for four students with the same total score.

RQ3. The cognitive pathways of middle school students' proficiency

We articulated a diagram (see Figure 4) of potential cognitive pathways that depict students' gradual acquisition of the number of attributes from none (AMP1, 00000) to complete mastery (AMP20, 11111). As mentioned above, the connections between AMPs are determined by the inclusion relationships between AMPs (Tatsuoka, 2009; Wu et al., 2021). However, some AMPs did not satisfy the inclusion relationship, so the corresponding AMPs were excluded from the diagram. This study found 20 AMPs from the LLM analysis of 244 students (see Table 7). However, we included 13 AMPs in the diagram (see Figure 4). For example, we excluded AMP3 (00011) because the AMP could not include the AMP2 (10000). Following this approach, we articulated the potential cognitive pathways that students may follow to develop. Of these cognitive pathways, a cognitive pathway, followed by a sequence of $A1 \rightarrow A5 \rightarrow A2 \rightarrow A4 \rightarrow A3$, contains most students ($N = 141$, 97.9%). The results suggest that when students develop their proficiency in constructing scientific explanations, they might follow the cognitive pathway that they can make an appropriate claim (A1), identify the differences and similarities of data (A5), use data patterns as evidence (A2), apply the relevant core ideas (A4), and construct a reasoning process to justify claim with evidence and core ideas (A3). The primary cognitive pathway may provide a potential developmental process for advancing students' constructing scientific explanations, which may inform teachers' classroom instructions. In addition, the other cognitive pathways (e.g. $A1 \rightarrow A4 \rightarrow A2 \rightarrow A5 \rightarrow A3$) also provide valuable information about students' potential cognitive processes.

Discussions and conclusions

This study utilised a CDM approach to analyze students' proficiency in constructing scientific explanations integrating chemical reactions and patterns. We employed well-validated assessment tasks (NGSA, 2024) that required students to construct scientific explanations using claims, evidence, and reasoning (McNeill et al., 2006) to make sense of phenomena related to chemical reactions. The validity and reliability of the assessment tasks and cognitive attributes were evaluated using the CDM approach (de la Torre & Minchen, 2014). The identified cognitive attributes highlight the critical

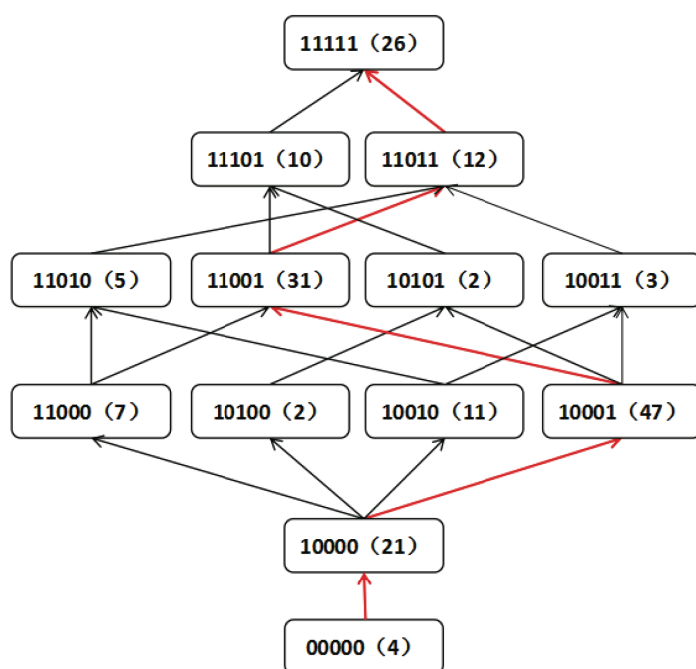


Figure 4. The cognitive pathways of students' constructing scientific explanations.

Note: The five-digit in the boxes is representing an AMP. The number in the parentheses is the frequency of the AMP. For instance, 10000 (21) is AMP2 (see Table 7) and contains 21 students.

components for students to construct scientific explanations in a content-specific domain from an integrated perspective. This extends previous studies focusing on developing practice-oriented attributes (Hu et al., 2021; Zhai et al., 2022). Our results indicate that the assessment tasks effectively diagnose middle school students' cognitive challenges in constructing content-specific scientific explanations. Based on the mastery of the cognitive attributes, the sample was divided into 20 groups, each representing a specific cognitive pathway of students' proficiency in constructing scientific explanations that integrate chemical reactions and patterns. We found that most students had challenges providing 'sufficient evidence' and making 'evidence-based reasoning.' The diagram of potential cognitive pathways reveals that students' proficiency in constructing scientific explanations that integrate chemical reactions and patterns might exhibit multiple pathways, which can support teachers in clearly understanding the students' proficiency and provide teachers with insights for providing targeted feedback and support in classroom.

Unfolding complex construct through CDM approach

The CDM approach was utilised to assess the effectiveness of adopted assessment tasks in capturing students' proficiency in constructing scientific explanations that integrate the core ideas of chemical reactions and patterns. The study also delves into the cognitive

process, provides insight into students' cognitive levels, and portrays potential multiple cognitive pathways. The adopted assessment tasks are designed to measure students' integrated proficiency, which involves different components working together to solve real-world problems or comprehend phenomena and requires advanced cognitive processes. Compared to previous studies of only focusing on science practices (e.g. Hu et al., 2021; Zhai et al., 2022), this study diagnosed students' proficiency in constructing scientific explanations integrated with chemical reactions and patterns. Therefore, the study addressed a challenge that was previously explored in prior research by uncovering complex latent cognitive attributes and processes.

Compared to traditional IRT and CTT models, the CDMs provide a more nuanced understanding of complex constructs using constructed assessment tasks (de La Torre, 2009), particularly regarding complex usable knowledge constructs (Harris et al., 2019; He et al., 2022). This study also enhances the applications of CDMs, providing a comprehensive set of fine-grained skills or attributes when evaluating students' mastery in a domain-specific area (Carpenter & Moser, 2020; Zhan et al., 2018). The CDM approach empirically uncovered students' cognitive proficiency in constructing scientific explanations and offered pedagogical insights by presenting the variety of students' cognitive attribute patterns. As such, it offers a deeper understanding of the cognitive attributes of constructing scientific explanations in a content-specific domain.

Furthermore, this study empirically validates using NGSA tasks (NGSA, 2024) in other educational contexts (i.e. China). Previous studies, such as those by Harris et al. (2019), have only discussed the design of these tasks and have yet to provide empirical evidence of the task quality. Our CDM analysis suggested that most items have good discrimination, demonstrating the reliability and validity of these assessment tasks in measuring students' integrated proficiency in constructing scientific explanations in a content-specific domain to make sense of phenomena or solve real-world problems.

Validating cognitive attributes of complex constructs

This study advances the field by developing and validating five cognitive attributes crucial in constructing scientific explanations, revealing the underlying cognitive processes involved in a content-specific domain. The reliability of these cognitive attributes has been confirmed through high test-retest agreement (average of 0.97), classification accuracy (0.77), and agreement (0.73). These validated attributes provide a means of diagnosing students' proficiency in constructing scientific explanations and identifying potential learning obstacles by examining individual attribute mastery. They also offer a valuable tool for teachers to select instructional strategies and assessments, targeting the development of some specific attributes.

This study goes beyond previous works such as Hu et al. (2021) and Zhai et al. (2022) by incorporating the CDM approach to evaluate practice-based and content-based attributes in constructing scientific explanations. To achieve this, we categorised the cognitive attributes into three groups: disciplinary core ideas (DCI, i.e. chemical reactions), science and engineering practice (SEP, i.e. constructing explanations), and crosscutting concepts (CCC, i.e. patterns). The SEP attributes align with McNeill et al.'s (2006) CER framework, which emphasises making claims, citing evidence, and reasoning with evidence. The DCI attribute is specific to chemical reactions, while the CCC attribute is crucial

as identifying patterns in data that can be used as evidence to support the created claim in a scientific explanation. Our results revealed students' proficiency in making claims, providing evidence, and reasoning with evidence but also provided insights into their cognitive mastery in chemical reactions and patterns. These findings highlight the critical aspects of diagnosing students' constructing scientific explanations in a content-specific domain.

Moreover, our results indicated that nearly all students mastered at least one cognitive attribute. The highest mastery rates could be found in 'making correct claims' and 'identifying patterns in data.' However, students were struggling with 'providing evidence' and 'reasoning based on evidence,' with only a 27.80% mastery rate in reasoning. Our findings align with existing studies on scientific explanation and scientific literacy at a coarse-grained level (McNeill et al., 2006; Yao & Guo, 2018; Zhan et al., 2019), indicating that students often struggle with the sufficiency and adequacy of evidence and using core ideas to support the reasoning for claims. The relationship between students' mastery of disciplinary core ideas and the construction of complete scientific explanations remains to be determined. Studies such as Zhan et al. (2019), which applied CDM analysis on PISA2015 science test data, found that content knowledge has the most significant influence on scientific literacy. Hu et al. (2021) also found that students' ability to construct scientific explanations is topic-specific, emphasising the connection between content knowledge and explanation construction.

Visualising latent cognitive pathways of complex constructs

Through the CDM approach, this study visualises latent cognitive pathways required for constructing explanations of chemical reactions and identifying patterns. These cognitive pathways revealed the various trajectories in which students develop their integrated scientific explanation proficiencies. Our study found 20 different patterns among the 244 students, with the most common pattern (i.e. AMP6, 10001) accounting for approximately 19% of the students. These students could make a claim and provide one piece of evidence. The evidence in the tasks was in the form of data or information, and students had to determine if factors like mass, odour, and density could be used as characteristic properties to identify substances. However, most students could only provide one piece of evidence. Approximately 10% of the students had mastered the disciplinary core ideas but were unable to apply them in constructing scientific explanations. They either failed to connect the evidence to the claims despite having all the necessary evidence or could not use their knowledge to solve the problem. This highlights the importance of developing usable knowledge, which our assessment task can effectively diagnose (Harris et al., 2019; Li, 2021).

Applying CDMs in our adopted assessment tasks identified various AMPs, offering insight into students' cognitive characteristics for constructing scientific explanations. Analysis of those AMPs also informs a potential primary cognitive pathway, outlining the cognitive development process of students' scientific explanation competencies in a content-specific situation. Our findings indicate that several cognitive pathways cover most students in our sample. This primary learning pathway is characterised by students first making an appropriate claim, then seeking evidence based on information, and finally using evidence to be reasoning while mastering chemical ideas, providing a

valuable understanding of the cognitive process of students constructing scientific explanations.

In science education, using CDMs to delineate cognitive pathways of complex constructs still needs to be explored. The diagram of cognitive pathways provides an approach to articulate students' potential cognitive process, which may reflect learners' mastery sequence of a specific group of students in a particular context but not reflect the content sequence that is described in a curriculum standard. This differentiation between learner and content logic sets the boundary for a learner-centered approach (Confrey et al., 2014). The cognitive pathways and individual students' diagnosis information may guide science teachers in selecting and adjusting their instructional materials, classroom strategies, and assessment activities.

It should be noted that there is a crucial relationship between our cognitive pathways and the well-studied 'learning progression' in science education (e.g. Songer et al., 2009). Presenting cognitive pathways is a cognitive perspective of learning progressions. Our study aligns with the existing work on the learning progressions of scientific explanations (e.g. Gotwals et al., 2012; Yao & Guo, 2018), which captures students' learning progress in constructing scientific explanations. While there is no empirical evidence about the relationship between our cognitive pathways and existing research on learning progression, future studies must incorporate students' cognitive pathways with learning progressions and investigate teachers' instructional feedback in supporting students with diverse cognitive pathways. In addition, teachers may use the developed cognitive pathways in different ways to guide their classroom instruction, depending on their professional status or student backgrounds.

Limitations

Building upon a CDM approach, this study investigated the cognitive process by which students construct scientific explanations by integrating chemical reactions and patterns. The inclusion relationships between AMPs were used to outline potential cognitive pathways, reflecting students' cognitive development. Our findings contribute to ongoing research on validating assessments and provide a cognitive perspective on the diagnosis of students' constructing scientific explanations. Moreover, the findings of this study provide a new perspective to understand students' cognitions towards three-dimensional learning of integrating DCIs, SEPs, and CCCs to make sense of phenomena or solve real-world problems. This study also proposed a guided approach to develop cognitive attributes and apply CDMs to investigate the cognitive features of three-dimensional learning, which may inform researchers in other content areas or other educational contexts.

While our study sheds light on essential insights, it is critical to acknowledge the limitations and areas for further improvement. One limitation is using convenience sampling, which may limit the generalizability of our results to a larger population. While they were deemed adequate, this study only utilised three assessment tasks. Some items with mediocre fit statistics (i.e. items 1.3-1 and 3.2-1) may challenge the reliability and validity of the tasks for diagnostic purposes. While the CDM models we used are more applicable to binary criteria, we also acknowledge that using binary criteria in our rubrics may be constrained to capture the cognitive attributes for holistically diagnosing students' proficiency. We encourage future research investigating whether different types of rubrics

(He et al., 2023b) would impact the CDM estimations of cognitive attributes on such complex constructs. For the CDMs, this study did not account for the interaction between cognitive attributes when using the LLM, which may have affected the estimations of the cognitive attributes and the potential cognitive pathways. We acknowledged that the percentages of the cognitive attributes and the potential cognitive pathways may vary by other subject contents within the same science practices (e.g. constructing scientific explanations) or the same subject content with different practices (e.g. designing solutions). Future research should consider these limitations and areas for improvement to provide a more comprehensive understanding of cognitive pathways in our study's context. Despite these limitations, our study makes a valuable contribution to science education. This study's cognitive patterns and pathways would provide meaningful information to help teachers adjust their instructional strategies better to support students' cognitive development in science classrooms.

Ethics Statement

All procedures performed in studies involving human participants were in accordance with

the ethical standards of the institutional research committee of Northeast Normal University. Informed consent was obtained from all individual participants involved in the study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study is supported by the National Science Foundation [grant number DRL-2446701]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Conflict of Interest Statement

The authors declare that they have no competing interests.

ORCID

Peng He  <http://orcid.org/0000-0002-2877-0117>

Yu Zhang  <http://orcid.org/0009-0008-0180-6555>

Tingting Li  <http://orcid.org/0000-0002-5692-2042>

Jie Yang  <http://orcid.org/0000-0003-3185-0588>

References

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, 95(4), 639–669. <https://doi.org/10.1002/sce.20449>

- Carpenter, T. P., & Moser, J. M. (2020). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser, & T. A. Romberg. (Eds.), *Addition, and subtraction* (pp. 9–24). Routledge.
- Chang, H. Y., Quintana, C., & Krajcik, J. (2014). Using drawing technology to assess students' visualizations of chemical reaction processes. *Journal of Science Education and Technology*, 23(3), 355–369. <https://doi.org/10.1007/s10956-013-9468-2>
- Chen, F., Zhang, S., Guo, Y., & Xin, T. (2017). Applying the rule space model to develop a learning progression for thermochemistry. *Research in Science Education*, 47(6), 1357–1378. <https://doi.org/10.1007/s11165-016-9553-7>
- Cheng, M. M. W., & Gilbert, J. K. (2017). Modelling students' visualisation of chemical reaction. *International Journal of Science Education*, 39(9), 1173–1193. <https://doi.org/10.1080/09500693.2017.1319989>
- Confrey, J., Maloney, A. P., Nguyen, K. H., & Rupp, A. A. (2014). Equipartitioning, a foundation for rational number reasoning: Elucidation of a learning trajectory. In A. P. Maloney, J. Confrey, & K. H. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 61–96). IAP Information Age Publishing.
- Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating students' reasoning about acid–base reactions. *Journal of Chemical Education*, 93(10), 1703–1712. <https://doi.org/10.1021/acs.jchemed.6b00417>
- Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49(1), 19–38. <https://doi.org/10.1111/j.1745-3984.2011.00158.x>
- Cui, Y., & Li, J. (2015). Evaluating person fit for cognitive diagnostic assessment. *Applied Psychological Measurement*, 39(3), 223–238. <https://doi.org/10.1177/0146621614557272>
- de La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115–130. <https://doi.org/10.3102/1076998607309474>
- de La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179–199. <https://doi.org/10.1007/s11336-011-9207-7>
- de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, 20(2), 89–97. <https://doi.org/10.1016/j.pse.2014.11.001>
- DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, 44(4), 285–291. <https://doi.org/10.1111/j.1745-3984.2007.00039.x>
- Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. Springer-Verlag.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312. [https://doi.org/10.1002/\(SICI\)1098-237X\(200005\)84:3<287::AID-SCE1>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A)
- Duschl, R. A. (2000). Using and abusing: Relating history of science to learning and teaching science.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), 613–619. <https://doi.org/10.1177/001316447303300309>
- Gotwals, A. W., & Songer, N. B. (2013). Validity evidence for learning progression-based assessment items that fuse core disciplinary ideas and science practices. *Journal of Research in Science Teaching*, 50(5), 597–626. <https://doi.org/10.1002/tea.21083>
- Gotwals, A. W., Songer, N. B., & Bullard, L. (2012). Assessing students' progressing abilities to construct scientific explanations. In A.C. Alonzo & A.W. Gotwals (Eds.), *Learning progressions in science* (pp. 183–210). Brill Sense.
- Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-In-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice*, 38(2), 53–67. <https://doi.org/10.1111/emip.12253>

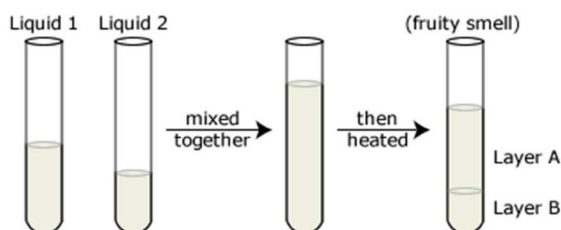
- Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign.
- He, P., Chen, I.-C., Touitou, I., Bartz, K., Schneider, B., & Krajcik, J. (2023a). Predicting student science achievement using post-unit assessment performances in a coherent high school chemistry project-based learning system. *Journal of Research in Science Teaching*, 60(4), 724–760. <https://doi.org/10.1002/tea.21815>
- He, P., Shin, N., Kaldaras, L., & Krajcik, J. (2024). Integrating artificial intelligence into learning progression-based learning systems to support student knowledge-in-Use: Opportunities and challenges. *Handbook of Research on Science Learning Progressions*, 461–487.
- He, P., Zhai, X., Shin, N., & Krajcik, J. (2023b). Applying Rasch measurement to assess knowledge-in-use in science education. In X.Liu & W. J. Boone (Eds.), *Advances in applications of Rasch measurement in science education* (pp. 315–347). Springer International Publishing.
- He, P., Zheng, C., & Li, T. (2021). Development and validation of an instrument for measuring Chinese chemistry teachers' perceptions of pedagogical content knowledge for teaching chemistry core competencies. *Chemistry Education Research and Practice*, 22(2), 513–531. <https://doi.org/10.1039/C9RP00286C>
- He, P., Zheng, C., & Li, T. (2022). Development and validation of an instrument for measuring Chinese chemistry teachers perceived self-efficacy towards chemistry core competencies. *International Journal of Science and Mathematics Education*, 20(7), 1337–1359. <https://doi.org/10.1007/s10763-021-10216-8>
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135–175. <https://doi.org/10.1086/286983>
- Hu, T., Yang, J., Wu, R., & Wu, X. (2021). An international comparative study of students' scientific explanation based on cognitive diagnostic assessment. *Frontiers in Psychology*, 12, 795497–795497. <https://doi.org/10.3389/fpsyg.2021.795497>
- Jiménez-Aleixandre, M. P., Bugallo Rodríguez, A., & Duschl, R. A. (2000). “Doing the lesson” or “doing science”: argument in high school genetics. *Science Education*, 84(6), 757–792.
- Kim, J., Kim, S., & Dong, H. (2015). International comparison of cognitive attributes using analysis on science results at TIMSS 2011 based on the cognitive diagnostic theory. *Journal of the Korean Association for Science Education*, 35(2), 267–275. <https://doi.org/10.14697/jkase.2015.35.2.0267>
- Leighton, J. P., & Gierl, M. J. (2007). Why cognitive diagnostic assessment? In J. P. Leighton, & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 3–18). Cambridge University Press.
- Li, T. (2021). Developing deep learning through systems thinking. In J. Krajcik, & B. Schneider (Eds.), *Science education through multiple literacies: Project-based learning in elementary school* (pp. 79–94). Harvard Education Press.
- Li, T., Miller, E., Chen, I. C., Bartz, K., Codere, S., & Krajcik, J. (2021). The relationship between teacher's support of literacy development and elementary students' modelling proficiency in project-based learning science classrooms. *Education 3-13*, 49(3), 302–316.
- Li, L., Zhou, X., Huang, J., Tu, D., Gao, X., Yang, Z., & Li, M. (2020). Assessing kindergarteners' mathematics problem solving: The development of a cognitive diagnostic test. *Studies in Educational Evaluation*, 66, 100879. <https://doi.org/10.1016/j.stueduc.2020.100879>
- Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, 69(3), 253–275. <https://doi.org/10.1111/bmsp.12070>
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64(2), 187–212. <https://doi.org/10.1007/BF02294535>
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1
- Ministry of Education, P. R. China. (2022). *Compulsory education chemistry curriculum standards*. Beijing Normal University Press.

- Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59(4), 439–483. <https://doi.org/10.1007/BF02294388>
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379–416. <https://doi.org/10.1111/j.1745-3984.1996.tb00498.x>
- Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6–20. <https://doi.org/10.1111/j.1745-3992.2006.00075.x>
- National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
- NGSA. (2024). Next generation science assessment. <https://ngss-assessment.portal.concord.org>.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.
- Oliveri, M. E., & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, 53(3), 315.
- Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, 95(4), 627–638. <https://doi.org/10.1002/sce.20438>
- Pellegrino, J. W., & Hilton, M. L. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Research Council.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, 12(1), 5–51. https://doi.org/10.1207/S15327809JLS1201_2
- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2
- Schwarz, C. V., Passmore, C., & Reiser, B. J. (2017). *Helping students make sense of the world using next generation science and engineering practices*. NSTA Press.
- Sevian, H., & Gonsalves, L. (2008). Analysing how scientists explain their research: A rubric for measuring the effectiveness of scientific explanations. *International Journal of Science Education*, 30(11), 1441–1467. <https://doi.org/10.1080/09500690802267579>
- Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal of Research in Science Teaching*, 49(2), 141–165. <https://doi.org/10.1002/tea.20454>
- Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, 46(6), 610–631. <https://doi.org/10.1002/tea.20313>
- Strike, K., & Posner, G. (1985). *A conceptual change view of learning and understanding, cognitive structure and conceptual change* (pp. 189–210). LHTWest and AL Pines, Press, Academic.
- Talanquer, V. (2013). How do students reason about chemical substances and reactions? In G. Tsapalis & H. Sevian (Eds.), *Concepts of matter in science education* (pp. 331–346). Springer Netherlands.
- Tang, K. S. (2016). Constructing scientific explanations through premise–reasoning–outcome (PRO): an exploratory study to scaffold students in structuring written explanations. *International Journal of Science Education*, 38(9), 1415–1440. <https://doi.org/10.1080/09500693.2016.1192309>
- Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, 49(1), 95–110. <https://doi.org/10.1007/BF02294208>
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Routledge.
- Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30(2), 251–275. <https://doi.org/10.1007/s00357-013-9129-4>
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11(3), 287–305. <https://doi.org/10.1037/1082-989X.11.3.287>
- Tu, D. (2019). FlexCDMs. Retrieved July 5, 2020, from <http://www.psychometrics-studio.cn>.

- Tu, D., Wang, S., Cai, Y., Douglas, J., & Chang, H. H. (2019). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted Q-matrix design. *Applied Psychological Measurement*, 43(4), 255–271. <https://doi.org/10.1177/0146621618765721>
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17(2), 228–243. <https://doi.org/10.1037/a0027127>
- Wang, W., Song, L., & Ding, S. (2018). The index and application of cognitive diagnostic tests from the perspective of classification. *Journal of Psychological Science*, 41, 475–483.
- Weinrich, M. L., & Talanquer, V. (2016). Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product. *Chemistry Education Research and Practice*, 17(2), 394–406. <https://doi.org/10.1039/C5RP00208G>
- Wu, X., Wu, R., Chang, H. H., Kong, Q., & Zhang, Y. (2020). International comparative study on PISA mathematics achievement test based on cognitive diagnostic models. *Frontiers in Psychology*, 11, 2230. <https://doi.org/10.3389/fpsyg.2020.02230>
- Wu, X., Wu, R., Zhang, Y., Arthur, D., & Chang, H. H. (2021). Research on construction methods of learning pathways and learning progressions based on cognitive diagnosis assessment. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 657–675. <https://doi.org/10.1080/0969594X.2021.1978387>
- Yan, F., & Talanquer, V. (2015). Students' ideas about how and why chemical reactions happen: Mapping the conceptual landscape. *International Journal of Science Education*, 37(18), 3066–3092. <https://doi.org/10.1080/09500693.2015.1121414>
- Yao, J. X., & Guo, Y. Y. (2018). Validity evidence for a learning progression of scientific explanation. *Journal of Research in Science Teaching*, 55(2), 299–317. <https://doi.org/10.1002/tea.21420>
- Yeo, J., & Gilbert, J. K. (2014). Constructing a scientific explanation—A narrative account. *International Journal of Science Education*, 36(11), 1902–1935. <https://doi.org/10.1080/09500693.2014.880527>
- Zhai, X., Haudek, K. C., & Ma, W. (2022). Assessing argumentation using machine learning and cognitive diagnostic modeling. *Research in Science Education*, 1–20.
- Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, 71(2), 262–286. <https://doi.org/10.1111/bmsp.12114>
- Zhan, P., Jiao, H., Liao, D., & Li, F. (2019). A longitudinal higher-order diagnostic classification model. *Journal of Educational and Behavioral Statistics*, 44(3), 251–281. <https://doi.org/10.3102/1076998619827593>

Appendix A

Xiao Fan placed two liquids in a test tube. The two liquids were soluble in each other and mixed together. She then heated the test tube to see if the liquids would react. After heating the liquids, two separate layers formed - Layer A and Layer B.



She tested and measured some properties of the liquids and layers and calculated the density of these substances, recording the data in Table 1.

Table 1. Data of sample before and after heating.

	Sample	Volume	Solubility in Water	Odor
Before Heating	Liquid 1	0.45 cm ³	Yes	alcohol
	Liquid 2	0.34 cm ³	Yes	vinegar
After Heating	Layer A	0.40 cm ³	No	none
	Layer B	0.30 cm ³	Yes	fruity

Question #1

State whether mixing and then heating Liquid 1 and Liquid 2 caused a chemical reaction to occur relating the liquids before heating to the layers after heating.

Question #2

Describe what information from the data table you would use as evidence to support your claim and explain why you used it.

Question #3

Support your claim with evidence and reasoning using the information in the data table.

Figure A1. Task 2-Layers in a test tube.

Note: The task was modified from Layers in a test tube (ID#: 019-03-c02) in the NGSA project (2024); Access to the original task via the link: <https://ngss-assessment.portal.concord.org/resources/202/layers-in-a-test-tube-id-019-03-c02>.

Xiao Fan dissolved two solids (Solid A and Solid B) into separate beakers, forming two clear liquids. Then, she mixed those two clear liquids together. She found that a yellow solid was formed (Solid P). recorded the color, solubility, and density of the solids before and after mixing in Table 1.

Table 1. Data of sample before and after mixing.

Solid	Color	Solubility in Water	Density
A (Before Mixing)	White	Yes	3.12 g/cm ³
B (Before Mixing)	White	Yes	4.53 g/cm ³
P (After Heating)	Yellow	No	6.16 g/cm ³

Question #1

Write a claim about whether a chemical reaction occurs when after mixing A and B.

Question #2

Describe what information from the data table you would use as evidence to support your claim and explain why you used it.

Question #3

Support your claim with evidence and reasoning using the information in the data table.

Figure A2. Task 3-Mixing clear liquids.

Note: The task was modified from Mixing clear liquids (ID#: 018.03-c02) in the NGSA project (2024); Access to the original task via the link: <https://ngss-assessment.portal.concord.org/resources/201/mixing-clear-liquids-id-018-03-c02>.