

Assessing high school students' understanding of friction force: A cognitive diagnostic modeling approach

Benjie Wang¹, Wei Han^{1,*}, Qingdian Kong², Huanxia Wang³, and Yingjie Zhang^{1,†}

¹*School of Physics and Physical Engineering, Qufu Normal University, Qufu, 273165, China*

²*The School attached to Qufu Normal University, Qufu, 273165, China*

³*Research Center for Basic Education Curriculum, Qufu Normal University, Qufu, 273165, China*



(Received 19 March 2025; accepted 12 August 2025; published 8 September 2025)

Enhancing students' conceptual understanding is one of the primary goals of science education. Existing literature suggests that the degree of knowledge integration among students can reflect their level of conceptual understanding. This study focuses on the concept of friction force, constructing a cognitive structure model of friction force to depict the cognitive structure features of students with different levels of conceptual understanding. Additionally, a cognitive diagnostic test tool for friction force was developed. Based on this tool, 827 Chinese high school students were tested. The results demonstrate that the cognitive structure model of friction force is reasonable and can effectively characterize the knowledge structures of students with varying levels of conceptual understanding. Furthermore, analysis revealed multiple viable pathways for cognitive development. Based on empirical data, we identified the progression pathway optimally aligned with students' cognitive schemata. The findings of this study provide a novel approach for mapping the cognitive development process of students and offer valuable information for personalized instructional guidance.

DOI: [10.1103/kenz-wrtr](https://doi.org/10.1103/kenz-wrtr)

I. INTRODUCTION

Enhancing students' understanding of scientific concepts is a key goal of science education [1–4]. Nevertheless, empirical evidence reveals persistent challenges: students frequently retain misconceptions and demonstrate context-bound reasoning patterns despite traditional instructional interventions [5–8]. The findings highlight the limitations of traditional instruction methods in fostering deep conceptual understanding. Therefore, identifying the key characteristics and developmental trajectories of students' conceptual understanding is essential for improving teaching effectiveness [9].

To address these persistent difficulties, a critical step is to understand the cognitive mechanisms through which misconceptions transform into scientific concepts—a process known as conceptual change. Learners often hold misconceptions prior to instruction. The transition from misconceptions to scientific concepts is an essential process for achieving deep understanding. Since the 1980s, exploring the cognitive mechanisms underlying these changes has

presented a significant challenge in scientific education research. Drawing inspiration from Piaget's ideas on the development of scientific concepts [10], researchers have proposed diverse theoretical models to simulate this process, with the core divergence lying in the knowledge evolution path from novice to expert. These conceptual changes can be roughly categorized into two major theoretical schools: the coherence theory and the fragmentation theory [11,12]. The coherent theory posits that students hold naïve theories that exhibit structural coherence and show significant parallels to historical prescientific frameworks, such as Aristotelian mechanics [13,14]. This perspective suggests that conceptual change requires a fundamental replacement of core beliefs, akin to a Kuhnian scientific revolution, rather than mere information addition. Its central argument is that misconceptions, once transformed, will be completely abandoned and cannot coexist with scientific concepts. In contrast, fragmentation theory posits that conceptual change entails the systematic reorganization and coordination of learners' fragmented intuitive knowledge elements, through which coherent conceptual structures are constructed [15–18]. Minstrell believes that students' naïve views are not obstacles to conceptual change but resources for constructing new knowledge in instruction [15,16]. Moreover, diSessa introduces the view that conceptual knowledge initially comprises fragmented, context-bound elements (p-prim), and the goal of learning is to integrate them into a coherent structure. For instance, understanding Ohm's law begins with abstracting key properties from everyday experiences (such as the

*Contact author: hanwei0127@163.com

†Contact author: yingjiezhang@qufu.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

perceived difference in resistance when pushing a box on different surfaces) and gradually integrating them into complex concepts. This increase in the degree of integration signals the occurrence of conceptual change [18]. In addition to the research by diSessa and his colleagues, other prevalent theories of conceptual change also indicate this accumulative nature. Norman and Rumelhart list “accumulation” as one of the three main mechanisms of conceptual change, emphasizing that knowledge integration can trigger improvements in the level of understanding [19]. Vosniadou’s theory of framework theory reconstruction also acknowledges the progressive nature of change [20]. Chi’s ontological category shift theory, while focusing on the “incompatibility” between misconceptions and scientific concepts in terms of ontology (i.e., cross-category attribution conflicts) and the learning barriers it causes, also emphasizes that within the same ontological category, the acquisition of new concepts relies on progressive knowledge accumulation (such as attribute modification), whereas cross-category shifts require the reconstruction of cognitive frameworks [21,22]. In summary, despite the significant controversy in the research on conceptual change [23], the mainstream consensus of current scientific education research tends to support the fragmentation theory perspective. This perspective emphasizes the progressive and cumulative nature of conceptual change and regards the enhancement of the integration degree of knowledge elements and the structural organization as the core indicators of deepening understanding. This view is further supported by expert-novice comparative studies: experts exhibit hierarchically integrated cognitive structures, whereas novices display fragmented, *ad hoc* cognitive structures [24–28]. Thus, evaluating the integration of students’ cognitive structures is a key dimension for revealing their conceptual understanding.

However, traditional psychometric methods (e.g., CTT and IRT) are limited in diagnosing the integration of cognitive structures due to their unidimensionality assumption and macrolevel assessment orientation [29–31]. Mislevy and co-workers [32] conceptualizes this epistemological framework as the “ability-level paradigm,” highlighting its psychometric limitations through two critical lacunae: (i) inadequate grounding in cognitive psychological foundations, manifested through undefined cognitive architectures in test interpretation, and (ii) insufficient specification of statistical structures underlying measured constructs [29–31]. Responding to these constraints, educational measurement has undergone a paradigm shift from ability-centric to cognition-oriented assessment frameworks [29,30]. At the forefront of this transformation lies cognitive diagnostic assessment (CDA), a cornerstone of new-generation measurement theory. Emerging from the convergence of cognitive psychology and psychometrics, CDA emphasizes the examination of students’ cognitive structures, information-processing mechanisms, and skill proficiencies during problem-solving tasks, transcending traditional unidimensional

macrolevel evaluations. Unlike conventional assessments constrained by unidimensional measurement, CDA offers multidimensional insights into various aspects of learning, yielding more diagnostically valuable information [30,33]. CDA operates by analyzing students’ task performance patterns to diagnose domain-specific knowledge states and mastery profiles, thereby elucidating the structural organization of their knowledge systems [34]. Through systematic clustering of cognitive attributes and their interdependencies, CDA not only diagnoses domain-specific knowledge states but also traces developmental trajectories from novice to expert proficiency [35,36]. This analytical framework not only advances theoretical understanding of cognitive progression but also equips educators with empirically grounded tools for precision pedagogy aligned with learners’ individual needs.

This study focuses on the concept of friction force. As one of the fundamental concepts in mechanics, a solid understanding of friction is beneficial for students to form a comprehensive conception of motion. However, in the field of education, research on friction force is relatively limited, mainly concentrating on diagnosing students’ misconceptions and enhancing teaching design [37–47]. There is a lack of explanation regarding the reasons for these misconceptions at the micropsychological level and the process of conceptual change. Therefore, this study aims to assess students’ understanding of friction by examining the integration within their cognitive structures from a knowledge integration perspective. The following tasks were accomplished:

Part 1: Identify the cognitive attributes and attribute hierarchy of friction force, and construct a cognitive structure model of friction force.

Part 2: Develop a cognitive diagnostic test for friction based on the cognitive structure model to depict the cognitive development process of students learning friction by analyzing their cognitive structures at different levels.

II. PART 1: CONSTRUCTION OF A COGNITIVE STRUCTURE MODEL OF FRICTION FORCE

A. Students’ misconceptions of friction force

Misconceptions serve as observable indicators of underlying cognitive flaws—precisely the phenomena that cognitive models seek to diagnose and resolve. By systematically synthesizing misconceptions about friction force from prior studies, we identified recurrent patterns of knowledge fragmentation, which directly informed the abstraction of cognitive attributes in subsequent sections. Existing literature [37–47] documents substantial student misconceptions predominantly across four dimensions: generation conditions, magnitude, direction, and effects of friction. Accordingly, this study examines learning difficulties along these key dimensions.

1. Misconceptions regarding the conditions for friction force generation

Students frequently depend on their everyday experiences to ascertain the existence of friction force. For instance, they may mistakenly believe that a stationary object will not experience friction force, even when external forces are applied to them [38]. Similarly, students tend to assert that moving objects invariably experience friction force [42,48]. And some students erroneously believe that friction force exists only within individual objects, overlooking the fact that friction force simultaneously acts on the surfaces of two objects in mutual contact [49,50].

2. Misconceptions regarding the direction of friction force

Students often determine the direction of friction force based on the misconception that friction force hinders the motion of objects. This specific misunderstanding manifests in four main aspects: first, the belief that the direction of friction force opposes the actual motion of the object, such as when an individual is cycling in a forward motion, both wheels of the bicycle experience friction force directed rearward [42]; second, the misconception that when an external force acts on an object, the direction of friction force is always opposite to the direction of the external force [42,49]; third, the perception that when an object undergoes acceleration, friction force acts opposite to the direction of acceleration, as seen in the majority of students concluding that the friction force on the front wheel of a bicycle during a turn is directed outward [42,48]; fourth, the misconception that the direction of friction force is always horizontal [42,49,51]. This is attributed to the presentation of friction force examples in textbooks predominantly in the horizontal direction, making it challenging for students to comprehend that friction force can occur in vertical or inclined directions.

3. Misconceptions regarding the magnitude of friction force

The calculation of friction force magnitude poses a challenge for students, as it requires them to accurately distinguish between different types of friction force.

Regarding the magnitude of static friction, four misconceptions are prevalent. First, when an object remains at rest, though there is an external force acting on it, the friction force acting on the object is always equal to 0 [52]. Second, the misconception that static friction is always equal to μF_N [43,44,51], representing the maximum static friction value. These cases often hinge on the critical values of 0 and μF_N , neglecting the awareness that the static friction force magnitude ranges from 0 to μF_N . The aforementioned examples illustrate that these students overlook the passivity of friction force, consequently failing to acknowledge that friction force's magnitude varies in response to the object's state of motion or the external forces exerted upon it. Third, the misconception that a

stationary object experiences a friction force greater than the applied force [45,53]. For instance, in interviews conducted by Sari *et al.* regarding students attempting to push a cart on a horizontal surface without success, students attributed the stationary state to the friction force between the cart and the surface being greater than the applied force. Fourth, the misconception that the magnitude of friction force between two objects is directly proportional to the area of the contact surface between two objects [11,38,53].

Concerning the sliding friction force, two categories of misconceptions exist. On the one hand, due to the influence of inertial thinking, students often mistake gravity for the normal force when applying the equation $f = \mu F_N$ to calculate the magnitude of the sliding friction force, leading to the erroneous equation $f = \mu mg$ [49,51]. On the other hand, students assess the magnitude of the friction force based on the object's motion state [50]. In the case of uniform motion, students erroneously believe that an object is not subject to friction force because its state of motion remains unchanged. Conversely, for accelerated motion, students perceive that the magnitude of the friction force varies with changes in the object's acceleration.

4. Misconceptions regarding the effect of friction force

Regarding the effects of friction force, students commonly hold two persistent misconceptions. First, many equate friction solely with resistance or drag [44,50]. Second, a significant proportion believe that any loss of mechanical energy is invariably attributable to friction [54,55]. For instance, when considering a stationary ball on an inclined plane, students may incorrectly assert that static friction causes the ball to lose energy [55]. In reality, under these conditions, static friction acts to prevent the ball from sliding down the incline, thereby maintaining its stationary position without causing any energy loss.

Collectively, these recurring misconceptions—persisting across educational contexts—reveal systematic gaps in knowledge integration. Students' overreliance on context-bound intuitions and neglect of force passivity demonstrate fragmented cognitive structures, wherein attributes like relative motion and Newtonian mechanics remain isolated rather than coherently linked. To bridge this gap, the following section constructs a cognitive structure model to explicate how these fragmented attributes integrate during conceptual change.

B. The cognitive structure model of friction force

Building on the identified patterns of knowledge fragmentation, a domain-specific cognitive structure model is required to map the integration pathway for friction concepts. While cognitive models theoretically capture problem-solving mechanisms [56], Gorin [57] emphasizes their limited applicability to academic testing contexts—particularly for multifaceted concepts like friction force where passive interaction with Newtonian mechanics

creates unique diagnostic demands. Consequently, we construct a customized cognitive structure model grounded in the misconception analysis of Sec. II A and developed through iterative expert validation. The construction process of the cognitive structure model primarily involved two critical steps: (i) the abstraction of cognitive attributes, which entailed identifying and defining the specific cognitive skills and knowledge components relevant to understanding friction force and (ii) the establishment of a hierarchical structure among these attributes, reflecting the logical and developmental progression of students' cognitive processes.

1. Abstract the attributes of friction force

Cognitive attributes encompass the knowledge, skills, and strategies required for an individual to perform specific tasks [58]. These attributes are organized based on logical relationships and developmental hierarchies in psychology. Therefore, this study focused on two key factors when abstracting these attributes: hierarchical properties and misconceptions. The hierarchical property aids in identifying subconcepts related to friction force, while misconceptions help in pinpointing key factors affecting students' understanding.

Understanding a force requires comprehending its conditions of generation, magnitude, and direction. For friction force:

- (1) Conditions for friction force: First, objects must be in contact and exert pressure on each other; second, the contact surfaces must be rough; and third, there must be relative motion or a tendency for relative motion between the objects.
- (2) Direction: The direction of the friction force is always opposite to the relative motion or the tendency for relative motion (with the caveat that

for complex objects, such as a car, the friction force on the front and rear wheels relative to the ground must be discussed separately).

- (3) Magnitude: The magnitude of the friction force is related to the properties of the contact surfaces (e.g., roughness) and the normal force between them. The formula for sliding friction is given by $f = \mu N$.

A thorough understanding of friction necessitates grasping the concept of relative motion and the ability to identify and quantify other forces acting on an object, such as gravity and normal forces. Leveraging this hierarchical structure, two essential attributes of friction emerge: (i) the presence of relative motion and (ii) the identification and measurement of interacting forces.

Furthermore, friction force is a passive force [51,59]. Its existence, magnitude, and direction depend entirely on other applied forces and the object's state of motion. Consequently, the specific effect of friction in a given scenario must be inferred based on the object's actual motion. Analysis of common misconceptions about friction reveals that many stem from an inadequate understanding of the relationship between force and motion. Examples include beliefs such as "a stationary object experiences no friction" [52], "a moving object always experiences friction" [42,48], or "the magnitude of friction is invariably μF_N " [43,44,51]. Previous research has substantiated this perspective on misconception origins [43,59,60]. Newton's laws of motion provide the crucial link between force and motion, enabling the evaluation of forces based on observed motion. Critically, comprehending the passive nature of friction is essential for this integration of concepts [60].

Therefore, the cognitive structure model for friction comprises the following core attributes: (i) the fundamental concept of friction itself, (ii) Newton's laws of motion, and (iii) the passive nature of friction. Table I lists these cognitive attributes along with their definitions.

TABLE I. Cognitive attributes and their definitions.

Attribute		Operational definition
A1	Relative movement	Master the knowledge of reference frames and understand the relativity of mechanical motion
A2	Description of force	Ability to identify specific forces acting on an object (limited to gravitational, elastic, and external forces), describe their magnitude and direction, and perform vector addition/decomposition exclusively on these specified forces (gravitational, elastic, external)
A3	The basic concept of friction force	Understand the conditions for the generation and effects of friction force
A4	Newton's laws of motion	Understand the relationship between force and motion, and be able to deduce the motion state of an object based on its force conditions; conversely, one should also be able to infer the force conditions of an object based on its motion state
A5	The passivity of friction force	Understand that friction is a passive force, whose effect is influenced by other external forces and the object's motion state. Able to solve friction-related problems from statics and dynamics perspectives.

2. Determine the attribute hierarchy

Building on extensive research in cognitive psychology, Leighton *et al.* [56] proposed that cognitive attributes are not isolated operations but elements within an interconnected network. They may follow certain psychological, logical, or hierarchical sequences. Accordingly, they introduced the attribute hierarchy, using attribute hierarchy diagrams to represent the cognitive structure models of relevant tasks.

Analysis of attribute descriptions reveals their hierarchical relationships. As established, mastery of the fundamental concept of friction force (A3) necessitates comprehension of “relative motion” (A1) and the ability to identify forces acting on an object (A2). Thus, A1 and A2 serve as prerequisites for A3.

Upon acquiring A3 (“fundamental concept of friction force”), learners are able to calculate sliding friction magnitude using $f = \mu N$ and determine its direction as opposing relative motion. However, challenges persist in analyzing static friction—its presence, magnitude, or direction—particularly when assessing stationary objects with nonintuitive relative motion tendencies.

Consequently, the application of Newton’s laws of motion (A4) becomes essential to analyze forces based on an object’s kinematic state. Integrating “Newton’s laws of motion” with the “fundamental concept of friction force” entails sophisticated comprehension of the passivity of friction force. Therefore, A3 and A4 are prerequisites for A5. Furthermore, proficiency in “force identification and measurement” (A2) underpins the application of Newton’s laws (A4), establishing A2 as a prerequisite for A4. The hierarchy among these five attributes is illustrated in Fig. 1.

C. Modeling student understanding using the friction force cognitive structure model

Existing research on students’ learning of friction force indicates that misconceptions are prevalent. To bridge these

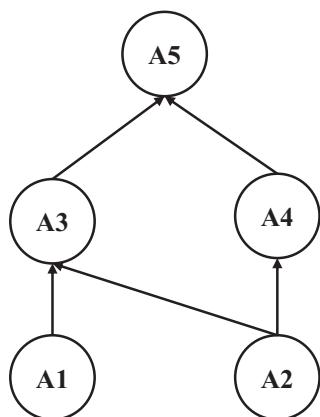


FIG. 1. Cognitive structure model of friction force. The arrows represent the sequence in which students master attributes during the cognitive process. For instance, before mastering A3, students are required to first acquire A1 and A2.

established findings, this study employs interviews to reexamine student misconceptions through the lens of knowledge integration. The interview results further serve to provide qualitative evidence for validating the cognitive structure model of friction force.

The interviews were conducted through three open-ended questions on sliding friction and static friction (see Fig. 2), which were derived from prior research on friction [43,60]. A total of 30 high school students from China were invited to participate in the interviews. During the interviews, students were asked to respond by thinking aloud so as to elaborate on their reasoning process. Each interview lasted approximately 20 min.

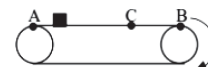
Building on existing research and interview findings, this study categorizes students’ conceptual understanding levels of friction force into novice, intermediate, and expert levels from a knowledge integration perspective. It further simulates their knowledge integration processes using the constructed cognitive structure model.

1. The novice level

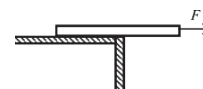
Students at this level possess an intuitive, experience-based understanding of friction derived primarily from everyday life. Interviews reveal a fundamental lack of conceptual comprehension, as illustrated in the following excerpts:

Student A: (Response to Q1) “*The object experiences friction to the left. Since the object moves*

Question 1: As shown in the figure, a conveyor device is placed horizontally and rotates at a constant speed. An object is gently placed at position A on the conveyor belt and reaches position C with the same velocity as the conveyor belt. What is the situation of the friction acting on the object during its conveyance from A to C? Explain your reasoning.



Question 2: As shown in the figure, a horizontal force F of 20 N is applied to the right to pull a wooden board, causing it to move to the right. The mass of the board is 10 kg, and the kinetic friction coefficient between the board and the table surface is 0.1. How does the magnitude of the friction force acting on the board change before it tips over?



Question 3: A magnet weighing 10 N is placed on the side of a refrigerator. A hand pushes upward with 6 N of force but the magnet does not move. Is the friction force exerted on the magnet upward, downward, or zero? Explain your reasoning.



FIG. 2. Interview questions used in part 1 of this research.

right, friction opposes its motion.” (Response to Q3) *“As I said earlier, friction opposes motion. The hand pushes the magnet upward, so friction acts downward to oppose that motion.”*

Student B: (Response to Q1) *“By definition, friction opposes relative motion. The object moves right relative to the conveyor belt, so friction acts leftward.”* (Response to Q2) *“As the plank tilts, pressure on the table decreases. Since $f = \mu F_N$, friction also decreases.”*

These responses indicate a fundamental lack of understanding regarding the concept of friction among novice learners. Two primary factors contribute to this learning difficulty: (i) an unfamiliarity with the definition of friction and (ii) lack of the concept of subordination. Persistent misconceptions arise when students lack a conceptual grasp of “relative motion.” Furthermore, while some students recall the kinetic friction equation ($f = \mu F_N$), they often fail to accurately analyze the normal force between objects. This leads to errors in calculating the magnitude of the friction force. This suggests that deficiencies in subordinate concepts impede the comprehension of superordinate principles.

The misconceptions observed during interviews align with findings documented in the existing literature [42,43,59,60]. Within the friction force cognitive structure model, these novice-level misconceptions can be ascribed to either the absence of cognitive attribute A3 or a failure to integrate attributes A1, A2, and A3.

2. The intermediate level

Students at this level demonstrate basic conceptual understanding. They recognize friction as a force opposing relative motion (or its tendency) and distinguish “relative motion” from absolute motion. They correctly identify forces like gravity and normal force, enabling preliminary quantitative analysis of kinetic friction via $f = \mu F_N$. However, they face challenges in assessing whether a tendency for relative motion exists between surfaces at rest, leading to difficulties in determining the effect of static friction, as evidenced below:

Student C: (Response to Q1) *“The object moves slower than the conveyor in section AC, so its relative motion is leftward. Friction opposes relative motion, thus acting rightward.”* (Response to Q2) *“The normal force on the plank remains constant (equal to gravity) and vertical during tilting. By $f = \mu F_N$, friction is constant.”* (Response to Q3) *“I cannot determine the relative motion tendency. Gravity gives the magnet downward tendency, but the hand’s support creates upward tendency.”*

Intermediate learners grasp friction’s core mechanism rather than simplistically equating it with resistance. Within

the friction force cognitive structure model, the cognitive attributes A1, A2, and A3 demonstrate effective integration in intermediate-level learners.

3. The expert level

Students with expert-level proficiency can analyze friction from the perspective of the relationship between force and motion. They recognize that “force is the cause of changes in an object’s state of motion.” By examining an object’s motion state, they can deduce the resultant force acting on it and further analyze the actual effects of friction through force composition and decomposition. The following interview excerpt illustrates this.

Student D: (Response to Q1) *“The object accelerates rightward in section AC, so the conveyor exerts rightward kinetic friction.”*

Student E: (Response to Q3) *“The magnet is static, so $F_{\text{net}} = 0$. Gravity acts downward (10 N), and the hand exerts an upward force (6 N). Thus, friction must be upward (4 N) to satisfy equilibrium.”*

Experts utilize Newtonian mechanics—deducing friction from motion dynamics—rather than relying solely on definitional friction properties. Their cognitive structure integrates attributes A1–A5 into a coherent framework according to the friction cognitive structure model.

The interviews reveal distinct reasoning patterns across proficiency levels, characterized by variations in the number and integration of applicable cognitive attributes. These differences reflect fundamental disparities in cognitive structure, which the friction cognitive structure model qualitatively captures, thereby offering preliminary validation of its rationality. In the subsequent phase of this research, we implement this model to develop a multiple-choice assessment. This instrument will elucidate students’ specific cognitive architectures and provide quantitative validation of the model’s efficacy.

III. PART 2: EXPLORING STUDENTS’ COGNITIVE PROGRESSION IN LEARNING FRICTION FORCE

A. Development of friction force cognitive diagnostic assessment tool

To assess students’ cognitive structure characteristics regarding the concept of friction and validate the friction cognitive structure model quantitatively, a multiple-choice tool, the friction cognitive diagnostic assessment tool, was developed. Following the cognitive diagnostic assessment paradigm, after determining the cognitive attributes and their hierarchical structure, it’s necessary to construct a Q matrix to represent the relationship between test items and attributes. The Q matrix is a crucial component that outlines the relationship between test items and cognitive attributes,

TABLE II. Q matrix. The matrix rows correspond to 13 test items (Items 1–13), and the columns to 5 cognitive attributes (A1–A5). “1” means the item measures the attribute, and “0” means it does not.

Item	Attribute				
	A1	A2	A3	A4	A5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	1	0	1	0
4	0	1	0	1	0
5	0	1	0	1	0
6	1	1	1	0	0
7	1	1	1	0	0
8	1	1	1	0	0
9	1	1	1	0	0
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	1	1	1	1	1

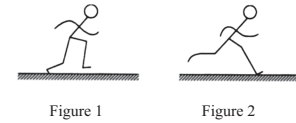
serving as a framework for constructing cognitive diagnostic assessments [61–63]. It is structured as a binary matrix with J rows (representing the number of test items) and K columns (representing the number of measured attributes). A value of $Q_{jk} = 1$ signifies that item j measures attribute k , while $Q_{jk} = 0$ indicates that item j does not measure attribute k . The Q matrix is considered complete if it satisfies two principles: first, the assessment should enable diagnosis of each cognitive attribute, and second, each cognitive attribute should be measured multiple times (at least three times). The Q matrix for the cognitive diagnostic test in this study is presented in Table II.

Based on this Q-matrix, the friction cognitive diagnostic assessment tool was developed in this study. After two—round iterative revisions, the final version of the assessment tool contains 13 items. Of these, item 3, item 4, item 5, item 6, item 7, item 9, item 11, and item 13 were developed and validated in prior research [60,64,65], while the remaining items come from China’s national physics-related college and high school entrance exams, which are standardized tests developed and validated by China’s National Education Examinations Authority. The full survey is in Supplemental Material [66]. The test uses a 0-1 scoring method.

The assessment tool incorporates five measurement modes. Mode (10000) evaluates students’ understanding of “relative motion” and their ability to differentiate between “motion” and “relative motion.” Mode (01000) assesses students’ knowledge of gravity and elasticity, as well as their capability in vector synthesis and decomposition. Mode (01010) examines students’ grasp of Newton’s laws of motion, including three scenarios that test their use of the prediction approach (predicting motion from force) and explanation approach (inferring force from motion) for Newton’s second law, and their understanding

7. As shown in the figure, there are two instants during Jake’s run. Let f_1 and f_2 represent the friction forces acting on Jake in Figures 1 and 2, respectively. Which of the following statements about the directions of f_1 and f_2 is correct? ()

- A. f_1 is forward, f_2 is backward
- B. f_1 is forward, f_2 is forward
- C. f_1 is backward, f_2 is backward
- D. f_1 is backward, f_2 is forward



13. As shown in the figure, Jack is writing on a horizontal desktop. To prevent slipping, an object is placed near the edge of the white paper on the left side. When writing to the right, both the paperweight and the paper remain stationary relative to the desktop. Which of the following statements is true?

- A. The object exerts a friction force on the paper, directed to the left.
- B. The object exerts a friction force on the paper, directed to the right.
- C. The desktop exerts a friction force on the paper, directed to the left.
- D. The pen exerts a friction force on the paper, directed to the left.

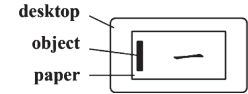


FIG. 3. Examples of items in the friction cognitive diagnostic assessment tool.

of Newton’s third law. Mode (11100) focuses on students’ comprehension of basic friction concepts, such as its conditions of generation, effects, and factors influencing its magnitude. Mode (11111) checks if students can analyze friction from the perspective of the relationship between force and motion. Figure 3 shows examples of two assessment modes. For item 7 (mode 11100), students need only apply the definition that the friction force opposes the direction of relative motion or the tendency of relative motion between surfaces to arrive at the solution. In contrast, item 13 (mode 11111) requires determining the friction force acting between the paper and the object. Since the presence of a tendency for relative motion cannot be definitively established in this case, it becomes necessary to analyze the actual effect of the frictional force based on the relationship between the applied forces and the observed motion. As the object is stationary relative to the desktop, its net force $F_{\text{net}} = 0$, so $f = 0$.

B. Cognitive diagnostic model

To effectively achieve the function of cognitive diagnosis, a cognitive diagnostic test is inherently dependent on a specific psychometric model, that is, the cognitive diagnosis model (CDM). Cognitive diagnostic assessment has advanced significantly, leading to the proposal of various CDMs by researchers. The selection of an appropriate CDM is critical for cognitive diagnostic analysis, as different models operate under distinct assumptions regarding attribute functioning mechanisms [30,67,68]. CDMs can be categorized into simplified models and saturated

models. Simplified models, such as the DINA [67] and DINO [30] models, offer advantages such as fewer parameters, conciseness, and stronger interpretability, but they also make strong assumptions about attribute functioning. In contrast, saturated models, like the G-DINA model [69], do not impose strict assumptions, allowing for the handling of more generalized attribute mechanisms. Moreover, the G-DINA model can be transformed into simplified models like DINA or DINO by imposing specific constraints. This study employs the more flexible G-DINA model, which is better suited for real-world testing scenarios, for the analysis of cognitive diagnostic assessments. The formula of the G-DINA model is presented as follows:

$$P(Y_{ij} = 1|\alpha_{ij}) = \delta_{j0} + \sum_{k=1}^K \delta_{jk} \alpha_{ik} + \sum_{k'=k+1}^K \sum_{k=1}^{K-1} \delta_{jkk'} \alpha_{ik} \alpha_{ik'} + \cdots + \delta_{j12K} \prod_{k=1}^K \alpha_{ik} \quad (1)$$

where $P(Y_{ij} = 1|\alpha_{ij})$ denotes the probability that a participant with an attribute mastery pattern (AMP) type i answers item j correctly, δ_{j0} represents the intercept for item j , refers to the probability that a participant will answer item j correctly without mastering any of the attributes measured by the item, δ_{jk} signifies the main effect of attribute k for that item, it represents the effect of increasing the probability of correctly answering the item due to the participant's mastery of attribute k , $\delta_{jkk'}$ captures the interaction effect between attributes k' and k for item j , α_{ik} indicates the mastery of attribute k for a participant of an attribute mastery pattern (AMP) i , with a value of 1 if mastered and 0 otherwise, $\alpha_{ik'}$ represents the mastery of attribute k' for the same participants, with a value of 1 if mastered and 0 otherwise.

To assess whether the G-DINA model fits the data, we employed the root mean square error of approximation (RMSEA) and the L_z statistic index to analyze the fit of items and participants. The RMSEA fit statistic for each item evaluates the degree of congruence between the item and the overall cognitive diagnostic model [70]. The L_z statistic is used to identify examinees whose response patterns are highly inconsistent with their estimated attribute mastery patterns [71]. The critical value for the RMSEA index was set at 0.1 [70]. An RMSEA value less than 0.1 indicates a good fit of the item, while a value greater than 0.1 suggests a poor fit. Table III presents the RMSEA values for the 13 items calculated using the G-DINA model. Except for item 5, all other items had RMSEA values less than 0.1, with an average RMSEA value of 0.0620, indicating a good fit of the items to the G-DINA model. The L_z statistic values greater than -2 indicate that the participant's responses are highly consistent with the model [71]. The results showed that

TABLE III. RMSEA of each item.

Item	RMSEA	Item	RMSEA
1	0.0544	8	0.0701
2	0.0238	9	0.0687
3	0.0729	10	0.0517
4	0.0119	11	0.0700
5	0.1179	12	0.0845
6	0.0844	13	0.0371
7	0.0581	Mean	0.0620

98.1% of the students' responses were highly consistent with the G-DINA model, ensuring the validity and reliability of the cognitive diagnostic assessment. In addition, the SRMSR was 0.047, which is less than the benchmark of 0.05 [72]. Therefore, these statistical indices suggest that the G-DINA model exhibits a robust fit.

C. Participants

A total of 827 high school students from China participated in this study, comprising grade 10 ($N = 526$) and Grade 11 ($N = 301$). Participants were 41.1% female, 54.8% male, and 4.1% did not report gender. All participants had completed instruction on friction and Newton's laws of motion prior to the study. Before data collection, the study purpose was explained, and written informed consent was obtained. All personal information was kept strictly confidential and processed anonymously.

D. Data analysis

1. Psychometric validation

We employed cognitive diagnosis theory and item response theory to evaluate the reliability and validity of the assessment instrument. Our analysis focused on test-retest consistency across five attributes, a critical metric for establishing the psychometric robustness of cognitive diagnostic assessments. Complementary Rasch modeling further substantiated the instrument's validity and reliability.

2. Validation of the friction cognitive structure model

The validity of the friction force cognitive structure model was examined through a four-stage process.

First, multivariate regression analysis evaluated the statistical adequacy of the five extracted attributes.

Second, hierarchical consistency indices (HCI), attribute mastery probability and attribute characteristic curves derived from student responses were analyzed to verify hypothesized attribute hierarchies.

Third, latent class analysis was conducted to validate the conceptual understanding levels of friction—novice, intermediate, and expert—as defined by the cognitive structure model established in Part I. Mplus 8.0 was employed to perform the latent class analysis on students' conceptual understanding of friction. Models specifying 1 to 4 latent

classes were estimated and compared. Model fit was evaluated using multiple criteria: the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the sample-size adjusted BIC (aBIC), the entropy index, the Lo-Mendell-Rubin adjusted likelihood ratio test (LMR-LRT), and the bootstrapped likelihood ratio test (BLRT). The model demonstrating optimal fit across these indices was selected for validating the conceptual understanding levels [73]. Specifically, lower values of AIC, BIC, and aBIC indicate better model fit. The entropy index ranges from 0 to 1, with values closer to 1 signifying greater classification accuracy; an entropy value ≥ 0.80 corresponds to a classification accuracy exceeding 90%. Finally, statistically significant p values for both the LMR-LRT and the BLRT indicate that the model with k cluster provides a significantly better fit than the model with $k - 1$ cluster.

Ultimately, the G-DINA model was applied to analyze attribute mastery patterns within each cluster participant, in order to examine the consistency between the cognitive structures of students within each proficiency level and the theoretical hypotheses.

3. Analysis of cognitive processes

Initially, distributions of attribute mastery patterns (AMPs) and their corresponding ability parameters were analyzed, with mastery denoted as “1” and nonmastery as “0” [e.g., AMP (0, 1, 0, 0, 0) indicates mastery solely of A2, while (0, 1, 0, 1, 0) signifies mastery of A2 and A4]. Finally, potential cognitive pathways were mapped by establishing partial ordering relationships between AMPs, thereby constructing a comprehensive latent cognitive path diagram. This network traces progression trajectories from null mastery (0 attributes) to complete mastery (5 attributes), utilizing directed edges to denote permissible transitions [e.g., (0, 1, 0, 0, 0) $\rightarrow$ (0, 1, 0, 1, 0)], thereby elucidating feasible learning trajectories within the attribute hierarchy.

E. Result 1: Establish the validity and reliability of the assessment instrument

1. Reliability analysis of assessment tools based on cognitive diagnostic theory

In terms of reliability, based on the attribute retest consistency metric proposed by Templin and Bradshaw [74], the reliability coefficient is derived under the assumption that the probability of an examinee mastering an attribute remains constant. This is calculated by correlating the attribute mastery probabilities of the same examinees across two successive measurements. As shown in Table IV, the test-retest reliability coefficients for all five attributes exceed 0.7, indicating good reliability of the test.

Construct validity refers to the degree to which a test accurately measures the theoretical construct or characteristic it is intended to assess. In other words, it reflects the extent to which test scores can be interpreted as

TABLE IV. Templin reliability indicators.

Attribute	A1	A2	A3	A4	A5	Mean
Reliability	0.945	0.996	0.930	0.873	0.890	0.927

representing a specific psychological construct within a theoretical framework. This involves establishing alignment between empirical evidence and the underlying theory—specifically, whether the assessment genuinely measures the hypothesized construct.

The assessment tool developed in this study was guided by cognitive diagnosis theory. Consequently, its construct validity was evaluated based on the validity of the Q matrix [35,75]. Q matrix validity was quantified using the hierarchy consistency index (HCI).

The HCI for student i is expressed by the following formula:

$$HCI_i = 1 - \frac{2 \sum_{j \in S_{\text{correct}_i}} \sum_{g \in S_j} X_{ij}(1 - X_{ig})}{N_{c_i}}, \quad (2)$$

where S_{correct_i} is an index set that includes the items answered correctly by student i , X_{ij} represents student i 's score (1 or 0) on item j , where item j belongs to S_{correct_i} , S_j is an index set that includes all items measuring the same attribute as item j , X_{ig} denotes student i 's score (1 or 0) on item g , where item g is part of S_j , N_{c_i} refers to the total number of comparisons conducted for all items that student i has answered correctly.

Since the Q matrix follows the established design principles for its construction and the HCI index ($HCI = 0.6572 > 0.6000$) has confirmed the high quality of the Q matrix, it can be concluded that the assessment tool possesses strong construct validity.

Additionally, the test items exhibit a high degree of content validity, as confirmed by expert evaluations. In conclusion, the assessment tool demonstrates strong validity.

2. Reliability analysis of assessment tools based on item response theory

Initially, the first step in applying the Rasch model is to confirm the unidimensionality of the data. To evaluate this, a principal component analysis (PCA) of the Rasch residuals was performed. The PCA results indicated that the Rasch dimension explained 47.9% of the variance, with an eigenvalue of 12.0. The first contrast, representing the largest secondary dimension, had an eigenvalue of 1.4, accounting for 5.8% of the unexplained variance. The variance explained by the Rasch measures was more than eight times that of the largest secondary dimension. Collectively, these results suggest that the data met the unidimensionality assumption required by the Rasch model [76,77].

After establishing unidimensionality, item fit statistics were calculated to evaluate how well the test score data

TABLE V. Item statistics from Rasch analysis. The item difficulty estimates are derived from the Rasch analysis. The model SE refers to the standard error of the item difficulty estimates, as calculated based on the Rasch analysis.

Item	Item difficulty	Model S.E.	Infit MnSq	Outfit MnSq
1	-4.50	0.34	1.03	2.15
2	-4.04	0.28	1.19	2.33
3	-1.87	0.13	0.95	0.90
4	-0.54	0.09	0.90	0.79
5	0.06	0.09	1.12	1.18
6	-0.10	0.09	1.00	1.00
7	0.00	0.09	1.09	1.15
8	-0.03	0.08	1.09	1.05
9	1.19	0.09	0.93	0.88
10	1.76	0.09	0.84	0.70
11	3.31	0.11	0.99	0.92
12	1.89	0.09	0.92	0.79
13	2.88	0.10	1.03	0.87

align with the Rasch model. This analysis generates two types of fit statistics sensitive to different aspects of misfit: the inlier-sensitive (Infit) and the outlier-sensitive (Outfit) statistics. The outfit mean square (MnSq) residual is crucial for identifying outliers during the Rasch analysis process. This includes situations where students with lower abilities correctly guess difficult items or where higher-ability students mistakenly answer easy items due to carelessness. In contrast, the infit mean square (MnSq) residual is used to detect anomalies in responses to items that are commensurate with the student's ability level. Ideally, a perfect fit of the item to the Rasch model is indicated by an MnSq value of 1.00. Generally, both the infit and outfit MnSq values should fall within the range of 0.60 to 1.40 [76]. As shown in Table V, the Rasch analysis of the test data indicates that the infit MnSq values range from 0.84 to 1.19, while the outfit MnSq values range from 0.70 to 2.33. The fit indices

indicated that the outfit mean square (MnSq) values for item 1 and item 2 (2.15 and 2.33, respectively) were marginally higher than the desired range. Given the substantial sample size ($N = 827$) and the foundational nature of these items, the elevated outfit statistics are more likely attributable to the sensitivity of this statistic to sporadic unexpected responses (e.g., correct answers by low-ability examinees or incorrect answers by high-ability examinees) on relatively simple items, rather than systematic model misfit. More critically, the more robust infit MnSq indices remained within acceptable limits, indicating adequate model-data correspondence for core measurement precision [77]. Furthermore, removing item 1 would not only compromise the test's structural validity but, more significantly, would deprive attribute A1 in the cognitive model of its sole dedicated indicator, substantially increasing the risk of misclassifying examinees' mastery of A1. Therefore, based on the sample-sensitive nature of outfit statistics, acceptable infit values, and item 1's indispensable role in the theoretical framework, item 1 was retained. For the same reasons, item 2 was also retained. Overall, the test data demonstrated acceptable fit to the Rasch model, supporting subsequent measurement inferences.

Additionally, the Rasch model analysis was used to evaluate person separation reliability and item separation reliability. The person separation reliability was found to be 1.45, with a corresponding internal consistency coefficient of 0.68. Furthermore, the item separation reliability was measured at 14.61, resulting in an internal consistency coefficient of 1.00. These results indicate that the assessment tool demonstrates high reliability [76].

Finally, the construct validity of the assessment tool was examined. In this study, the construct specifically refers to the cognitive structure model depicted in Fig. 1. The model posits that more complex cognitive constructs, which integrate multiple attributes, should be more challenging

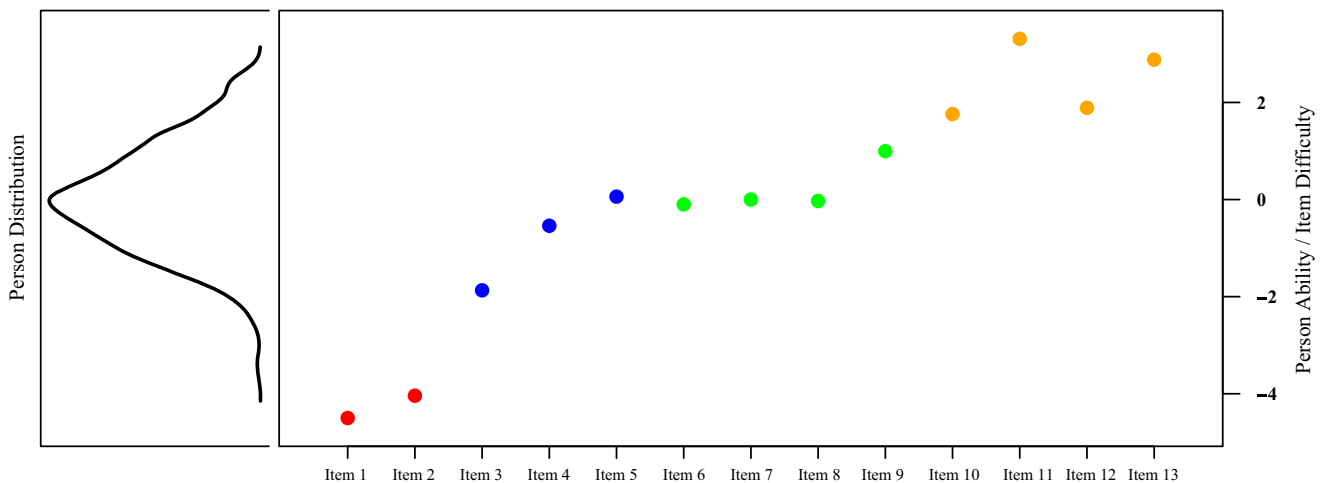


FIG. 4. Wright map. The scatterplot on the right represents the difficulty values of different items, with color coding used to identify four item sets. The histogram on the left displays the density distribution of all examinees across various ability levels.

TABLE VI. Explanatory analysis of the difficulty of the item.

Validation of the regression equation model			Coefficient of determination		
F	df	p	R^2	Adjusted R^2	S.E.
20.246	4	0.000	0.920	0.875	0.867

to master, thereby reflecting higher levels of cognitive ability. If the assessment tool effectively measures different levels of students' conceptual understanding, it provides evidence of strong construct validity. In other words, items that assess superordinate attributes should demonstrate a higher level of difficulty, while items that evaluate subordinate attributes should exhibit a lower level of difficulty. Figure 4 presents a wright map, which provides an overview of item difficulty and person ability distributions along the same logit scale. The left side of the map illustrates the distribution of persons sorted by their abilities, with the most proficient individuals positioned at the top and those with lower abilities at the bottom. The right side of the map displays the varying difficulties of different items, with the most challenging items placed at the top and the easiest items at the bottom. Items that assess the same attribute are represented by dots of the same color. A comparison of the difficulty levels of the items in Fig. 4 with the measurement properties outlined in the Q matrix reveals that the item difficulty levels are consistent with the theoretical constructs. As the integration of the measured concepts within the items increases, so does the difficulty value of those items. This alignment further suggests that the assessment tool possesses strong construct validity [78].

In summary, the assessment tool developed in this study, grounded in the cognitive structure model of friction force, complies with established psychological and educational measurement standards. It is effectively designed to evaluate the cognitive development of students regarding their understanding of the concept of friction force.

F. Result 2: Establish the validity and reliability of the cognitive structure model

1. Validation of attribute extraction rationality

The rationality of attribute extraction is assessed to ensure the comprehensive identification of fundamental attributes within the cognitive structure model. To this end, linear regression analysis was employed, using the explanatory power of extracted attributes on item difficulty as a quantitative indicator to evaluate their rationality. If the explanatory power reaches 60% or above, the extracted cognitive attributes can be deemed essentially reliable [34].

With the attributes contained in the items as independent variables and item difficulty as the dependent variable, multiple linear regression analysis was conducted. The results indicate that the regression model is statistically

significant ($F = 20.246$, $p < 0.001$), with the independent variables accounting for 87.5% of the variance in item difficulty (as shown in Table VI), demonstrating a high degree of explanatory power. Therefore, the five attributes extracted in this study can effectively predict item difficulty and are considered rational.

2. Validation of attribute hierarchy

The purpose of examining the reasonableness of the hierarchy relationship is to verify whether the attribute hierarchy aligns with logical knowledge relationships and cognitive psychological sequences.

The validity of the attribute hierarchy was assessed by comparing mastery probabilities. A higher mastery probability indicates lower attribute difficulty and a lower position within the hierarchy. As shown in Table VII, the mastery probabilities reveal that A5 is lower than those of A3 and A4, A3 is lower than those of A1 and A2, and A4 is lower than that of A2. This pattern aligns with the attribute hierarchy specified in the cognitive structure model of friction force.

To visualize the integration of cognitive attributes, attribute characteristic curves for five attributes were plotted (Fig. 5), depicting the relationship between attribute mastery probability and total test scores. Analysis of the figure reveals two key findings. First, the average mastery probability across all five attributes exhibits a consistent upward trend as total scores increase, thereby validating the hypothesis that greater attribute mastery corresponds to higher test performance. Second, variations in curve slopes reflect differential attribute difficulty. Attributes A1 and A2 show a marked increase in probability (from 0.4 to almost 1.0) between total scores of 2 and 3, indicating relatively low difficulty of mastery. Attributes A3 and A4 show significant probability growth beginning at 6 points, reaching full mastery (probability = 1.0) at 10 points. Attribute A5 initiates its ascent at approximately 6 points but only approaches full mastery (probability ≈ 1.0) at 13 points. This indicates that A5 is more difficult than A3 and A4, A3 is more difficult than A1 and A2, and A4 is more difficult

TABLE VII. Attribute mastery probability.

Attribute	A1	A2	A3	A4	A5
Mastery probability	0.988	0.979	0.485	0.588	0.307

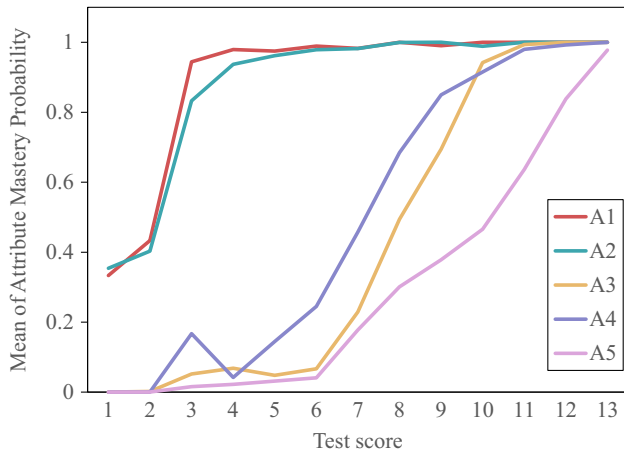


FIG. 5. Attribute characteristic curves. Curves depict mastery probabilities of attributes A1–A5 across total test scores (0–13 points). Steeper slopes indicate lower attribute difficulty (e.g., A1/A2), while gradual slopes (e.g., A5) reflect higher difficulty. The sequential mastery order (A1/A2 → A3/A4 → A5) validates the hierarchical structure of the friction cognitive model.

than A2, which is consistent with the hierarchy in the cognitive structure model.

In summary, the friction force cognitive structure model constructed in this study is rational and possesses satisfactory explanatory and predictive efficacy.

3. Validation of conceptual understanding levels

To validate the hypothesized conceptual understanding levels, we conducted latent class analyses specifying 1–4 clusters. Results are presented in Table VIII. As the number of clusters increased, the AIC, BIC, and aBIC values decreased, with the rate of decrease substantially diminishing beyond the two-cluster solution, indicating progressively better model fit. The three-cluster model yielded the highest entropy value (0.810), demonstrating its superiority over both the two-cluster and four-cluster solutions. The LMR-LRT and BLRT p values were statistically significant ($p < 0.001$) for both the two-cluster and three-cluster models. While the four-cluster model showed a significant LMR-LRT p value, its BLRT p -value was nonsignificant. Examination of class probability distributions further revealed one class in the four-cluster solution with an extremely low estimated probability, reinforcing the

TABLE IX. Average assignment probabilities for participants across latent classes.

Model	Attribution probability		
	c1	c2	c3
Cluster 1	0.913	0.065	0.022
Cluster 2	0.000	0.913	0.087
Cluster 3	0.000	0.084	0.916

superiority of the three-cluster model. Consequently, based on classification accuracy and model parsimony, the three-cluster solution was selected as optimal.

Using this model, the average probabilities of individual assignment to the three latent classes were computed (Table IX). For each participant, the probability of assignment to their most likely class exceeded 0.90, while the probability of assignment to any other class was below 0.10, corroborating the accuracy of the three-class solution.

Therefore, we employed a three-cluster model to test our hypotheses. Specifically, an item was deemed discriminatory for a given proficiency level if students at that level demonstrated a high probability of solving it, whereas students at lower levels were unlikely to succeed. To validate the items, we plotted the response probability distributions across the 13 items for each latent class (see Fig. 6). For instance, item 7, which assesses attributes A1, A2, and A3, was hypothesized to be solvable by students at the intermediate and expert levels but not by novices. As shown in Fig. 6, the response probability for cluster 1 (0.441) was lower than for cluster 2 (0.704) and cluster 3 (0.764), consistent with expectations. Analysis of all 13 items confirmed that each performed effectively in distinguishing student clusters, thereby supporting the hypothesized performance differences across proficiency levels.

4. Validation of cognitive structures across different proficiency levels

To assess whether the cognitive structures of students at different proficiency levels aligned with the hypothesis introduced in part 1, we computed the attribute mastery patterns (AMPs) for the three-cluster samples (Table X) and compared them to the expected patterns for novice, intermediate, and expert levels.

TABLE VIII. Latent class model fit indices for students' friction conceptual understanding level.

Model	AIC	BIC	aBIC	Entropy	BLRT (p value)	LMR-LRT (p value)
One-cluster	11028.137	11098.776	11051.142	...	...	...
Two-cluster	10246.452	10392.440	10293.996	0.714	<0.001	<0.001
Three-cluster	10125.388	10346.725	10197.471	0.810	<0.001	<0.001
Four-cluster	10037.857	10334.544	10134.480	0.711	0.288	<0.001

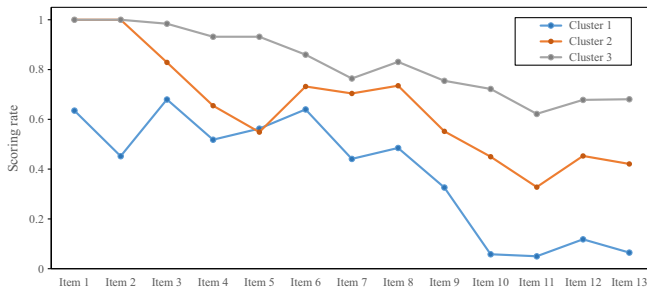


FIG. 6. Response probability distributions across 13 items for three latent classes of conceptual understanding.

Table X indicates that cluster 1 samples exhibited five distinct AMPs. However, some patterns contained very few students and were consequently excluded from further analysis due to insufficient sample size. The predominant AMP for cluster 1 involved mastery of attributes A1 and A2. Similarly, cluster 2 samples demonstrated three primary AMPs: mastery of (i) A1, A2, and A4; (ii) A1, A2, and A3; and (iii) A1, A2, A3, and A4. Cluster 3 exhibited a single AMP, characterized by mastery of all attributes A1 through A5.

Based on the performance profiles of the three clusters shown in Fig. 6, cluster 1 can be classified as novice level, cluster 2 as intermediate level, and cluster 3 as expert level. Comparison with the part 1 hypothesis revealed that the cognitive structures of the novice and expert clusters were consistent with expectations. In contrast, the intermediate cluster exhibited a cognitive structure that differed slightly from the hypothesis. Specifically, while the hypothesis posited that intermediate students would master A1, A2, and A3, the cluster analysis identified three distinct AMPs within this group. Notably, some students appeared to master A4 (Newton's laws of motion) before mastering A3 (fundamental concepts of friction), a sequence contrary to the presentation order in standard Chinese high school physics textbooks. We hypothesize that students exhibiting this pattern, despite lacking explicit knowledge of friction's definition, may leverage their understanding of Newton's

TABLE X. Numbers of students and attribute mastery patterns in each cluster.

A1	A2	A3	A4	A5	Cluster 1	Cluster 2	Cluster 3
0	0	0	0	0	1		
1	0	0	0	0	1		
0	1	0	0	0	2		
0	1	0	1	0	2		
1	1	0	0	0	256		
1	1	0	1	0	13	81	
1	1	1	0	0		43	
1	1	1	1	0		156	9
1	1	1	1	1			163

TABLE XI. Revised cognitive structural framework features of students at different proficiency levels.

Level	A1	A2	A3	A4	A5
Novice	✓	✓			
Intermediate	✓	✓	✓		
	✓	✓		✓	
	✓	✓	✓	✓	
Expert	✓	✓	✓	✓	✓

laws to infer force relationships from motion. This could enable them to solve certain types of problems, while potentially leading to difficulties with tasks requiring specific friction knowledge (e.g., factors influencing friction magnitude). Analysis of response patterns for students in this AMP supports this interpretation.

Consequently, we have revised the potential cognitive structures for the Intermediate level to reflect these findings, as detailed in Table XI.

G. Result 3: Construct cognitive development pathways

To illustrate the progression of students with differing initial proficiency levels in learning the concept of friction force, learners were categorized into distinct attribute mastery pattern (AMP) groups based on cognitive diagnostic assessment results. Subsequently, the cognitive pathways were identified by analyzing both the mean ability estimates associated with each AMP and the partial order relationships among the AMPs.

Cognitive diagnostic model is designed to evaluate each student's mastery of various attributes based on their response data, providing a binary judgment on the mastery status of each attribute (with mastery denoted as 1 and nonmastery as 0). Consequently, the mastery status of different attributes for each student forms a vector, which is referred to as the attribute mastery pattern (AMP). For instance, the AMP for the participant i can be represented by a vector $\mathbf{K}_i = (\alpha_1, \alpha_2, \dots, \alpha_m)$, where α_m represents participant i 's mastery of attribute A_m , and $\alpha_j \in \{0, 1\}$ for $j = 1, 2, \dots, m$. Here, a value of 1 indicates that participant i has mastered attribute j , while a value of 0 signifies nonmastery. In this study, the G-DINA package [79] was utilized for expected *a posteriori* estimation to determine the attribute mastery patterns of the participants. Furthermore, the Mirt package [80] was employed to estimate the ability parameters of the participants using the Rasch model. The results indicate that 87.51% of participants exhibited an ideal mastery pattern. This category encompassed nine distinct variants of the ideal pattern, with further details provided in Table XII.

To precisely delineate students' potential cognitive pathways in understanding friction force, this study outlines

TABLE XII. Attribute mastery patterns of participants. The attribute mastery patterns for participants were derived using the expected *a posteriori* estimation. Within the attribute vector K_i , the digit “1” signifies that the attribute has been mastered, while “0” indicates that it has not been mastered. For example, the vector (0, 1, 0, 0, 0) signifies mastery of attribute A2.

Attribute mastery pattern	K_i	Average ability
AMP1	(0, 0, 0, 0, 0)	-3.24
AMP2	(1, 0, 0, 0, 0)	-2.55
AMP3	(0, 1, 0, 0, 0)	-1.50
AMP4	(0, 1, 0, 1, 0)	-0.92
AMP5	(1, 1, 0, 0, 0)	-1.03
AMP6	(1, 1, 0, 1, 0)	-0.29
AMP7	(1, 1, 1, 0, 0)	-0.07
AMP8	(1, 1, 1, 1, 0)	0.55
AMP9	(1, 1, 1, 1, 1)	1.31

their learning trajectories. These pathways are constructed based on two fundamental assumptions: First, the mastery of knowledge or skills by students is a gradual and continuous process. Second, the progression from lower to higher knowledge states among different participants within the sample group reflects the process of knowledge acquisition [81]. Within the framework of cognitive diagnostic theory, a learning pathway refers to a hypothetical trajectory abstracted from the collective consciousness within a specific group. It reveals the hierarchical relationship of knowledge states exhibited by students at different levels during the process of attribute acquisition. For instance, participants with the mastery pattern (1, 1, 1, 1, 1) have mastered the attributes {A1, A2, A3, A4, A5}, while those with the mastery pattern (1, 1, 1, 1, 0) have mastered {A1, A2, A3, A4}. Since $\{A1, A2, A3, A4, A5\} \supset \{A1, A2, A3, A4\}$, it follows that there exists a learning pathway $(1, 1, 1, 1, 0) \rightarrow (1, 1, 1, 1, 1)$. However, due to the complexity of learning pathways, this study only considers learning pathways composed of adjacent levels. Although $\{A1, A2, A3, A4, A5\} \supset \{A1, A2, A3\}$, the pathway $(1, 1, 1, 1, 0, 0) \rightarrow (1, 1, 1, 1, 1)$ is excluded from the analysis because it spans multiple levels and does not meet the criteria of gradual progression considered in this study. This approach ensures a clear, sequential understanding of students' cognitive development.

Figure 7 illustrates the advancement paths of the participant group, from which it can be observed that four distinct progression paths can be identified:

1. $(0, 0, 0, 0, 0) \rightarrow (1, 0, 0, 0, 0) \rightarrow (1, 1, 0, 0, 0) \rightarrow (1, 1, 1, 0, 0) \rightarrow (1, 1, 1, 1, 0) \rightarrow (1, 1, 1, 1, 1)$;
2. $(0, 0, 0, 0, 0) \rightarrow (1, 0, 0, 0, 0) \rightarrow (1, 1, 0, 0, 0) \rightarrow (1, 1, 0, 1, 0) \rightarrow (1, 1, 1, 1, 0) \rightarrow (1, 1, 1, 1, 1)$;
3. $(0, 0, 0, 0, 0) \rightarrow (0, 1, 0, 0, 0) \rightarrow (1, 1, 0, 0, 0) \rightarrow (1, 1, 1, 0, 0) \rightarrow (1, 1, 1, 1, 0) \rightarrow (1, 1, 1, 1, 1)$; and
4. $(0, 0, 0, 0, 0) \rightarrow (0, 1, 0, 0, 0) \rightarrow (0, 1, 0, 1, 0) \rightarrow (1, 1, 0, 1, 0) \rightarrow (1, 1, 1, 1, 0) \rightarrow (1, 1, 1, 1, 1)$.

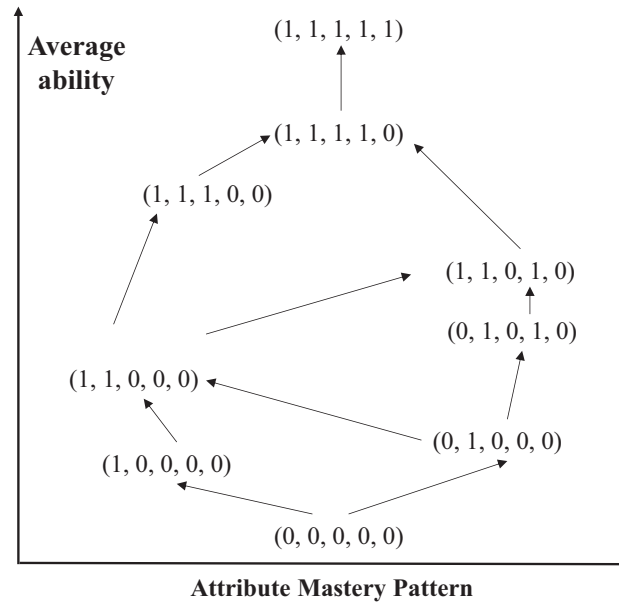


FIG. 7. Learning pathways. The learning pathways are constructed based on the hierarchical inclusion relationships among attribute mastery patterns. It is important to note that only sequential progression routes are included, such as $(1, 1, 1, 0, 0) \rightarrow (1, 1, 1, 1, 0)$, and skipped progression routes are excluded, such as $(1, 1, 0, 0, 0) \rightarrow (1, 1, 1, 1, 0)$. Additionally, the y axis represents the average ability values of participants with different attribute mastery patterns.

This suggests that students' cognitive development does not follow a singular, linear path but instead manifests through multiple possible trajectories. Such variation aligns with the nature of learning complex cognitive tasks, where the development of understanding is both multidimensional and hierarchical. For instance, in popular high school physics textbooks in China, students are typically introduced to the definition of friction force before Newton's laws of motion, following a learning path of $(1, 1, 1, 0, 0) \rightarrow (1, 1, 1, 1, 0)$. However, by analyzing the actual learning paths, we observe that students also follow alternative trajectories, such as $(1, 1, 0, 1, 0) \rightarrow (1, 1, 1, 1, 0)$. Building on the prior data, attribute A4 proved to be less difficult than attribute A3. Furthermore, the progression along the pathway $(1, 1, 0, 0, 0) \rightarrow (1, 1, 0, 1, 0) \rightarrow (1, 1, 1, 1, 0) \rightarrow (1, 1, 1, 1, 1)$ exhibited a more gradual increase in ability estimates, imposing a lower cognitive load on students. Consequently, this pathway aligns more closely with students' cognitive progression regarding the concept of friction than the learning sequence commonly presented in popular senior high school physics textbooks in China. Further analysis of the cognitive development processes of the 10th and 11th grade samples revealed that their cognitive development structures and potential optimal progression pathways regarding the concept of frictional force are similar. This finding further validates the rationality of the friction force cognitive structure model.

IV. DISCUSSION

Friction force is ubiquitous in daily life; however, its inherent complexity and abstract nature has led to widespread misconceptions among students. Understanding the origins of these misconceptions is critical for educators to design targeted instructional strategies and foster deeper conceptual understanding. This study addresses these issues through the lens of knowledge integration. This research first identified cognitive attributes related to friction and simulated their accumulation during conceptual change, thereby constructing a friction-specific cognitive structure model. Subsequently, cognitive diagnostic assessment tools were developed based on cognitive diagnostic theory. Through mixed-methods analysis, the validity of both the model and instruments was confirmed, yielding insights into the cognitive structures of students at varying conceptual levels and the developmental trajectory from lower to higher proficiency. These findings provide actionable guidance for personalized instruction.

Contrary to traditional views attributing misconceptions solely to everyday experiences, this study reveals that even students proficient in friction-related concepts (e.g., conditions for friction generation and directional definitions) struggle with problem solving. This persistence may arise from cognitive inertia—students often overlook the reactive nature of friction, which dynamically adapts to external forces and motion states. Consequently, they erroneously apply definition-based analyses akin to gravity, leading to enduring misconceptions such as “moving objects must experience friction” [42] and “stationary objects must not be subject to friction forces” [50]. Thus, this study argues that students must reconstruct their understanding of friction within the framework of motion and interaction. Crucially, integrating Newtonian mechanics into friction instruction is essential for deep conceptualization. Notably, some students compartmentalize “basic friction concepts” and “Newton’s laws” within their knowledge structures, hindering integration. Emphasizing friction’s reactive nature is key to bridging this gap [51,59,60].

Furthermore, this study delineates fine-grained cognitive development trajectories, revealing multiple nonlinear pathways. For instance, students mastering foundational attributes (A1, A2; pattern [1, 1, 0, 0, 0]) diverge into two paths: (i) learning friction basics (A3) before Newton’s laws (A4), aligned with the Chinese curriculum; or (ii) prioritizing Newton’s laws (A4) before friction (A3), as in U.S. curricula. Attribute mastery probabilities indicate A3 is more challenging than A4 ($p_{A3} < p_{A4}$), suggesting the latter pathway aligns better with cognitive logic and gradual difficulty progression.

Finally, this study validates the advantages of cognitive diagnostic assessments in modeling students’ cognitive development. In the field of scientific concept understanding, research on learning progressions is dedicated to

modeling students’ cognitive development [82]. Current learning progressions are primarily based on item response theory [83]. Some scholars argue that it is more reasonable to have an ordinal structure for latent variables when constructing progression models than a continuous structure [84,85]. However, in IRT analysis, students’ abilities are treated as latent variables, and their ability estimates are continuous variables, often requiring the setting of thresholds to classify students’ ability levels, a process that is inherently subjective. Moreover, cognitive development models based on IRT analysis can only describe linear growth in ability levels and cannot account for the multi-path phenomenon in cognitive development. In contrast, assessments based on cognitive diagnostic theory treat cognitive attributes as latent variables, which are typically binary (mastered or not mastered) and used to describe students’ states across multiple cognitive dimensions. By characterizing the cognitive structures of individuals at different levels, a more nuanced analysis and classification of students’ cognitive states can be achieved, providing a deeper understanding of the learning needs of different students and thus offering guidance for their subsequent learning.

While this study offers valuable contributions, several limitations warrant acknowledgment. First, the exclusive reliance on a Chinese sample limits the generalizability of the findings. The observed cognitive structures and developmental trajectories are likely closely tied to the specific ways knowledge is presented and sequenced within the Chinese physics curriculum. Consequently, the broader applicability of our conclusions is constrained. Future research should incorporate greater geographical and demographic diversity to enhance the generalizability of inferences. Second, although the study emphasizes knowledge integration, it does not sufficiently explore the potential influence of other factors on conceptual understanding levels. Factors such as problem-solving context characteristics or individual differences were not thoroughly examined. Subsequent research should further investigate how instructional interventions might modulate these factors to foster cognitive development. Third, attributes A1 and A2 were each measured by only a single item in the assessment instrument. While pilot testing and interviews indicated very high mastery rates for both attributes within the sample, and while the G-DINA model estimates attribute mastery patterns based on students’ overall response patterns across all items—rather than relying solely on single-item performance—which objectively mitigates the risk to some extent, it must be acknowledged that single-item measurement is inherently more susceptible to random error than multi-item measurement. This susceptibility increases the potential risk of misclassifying individual attribute mastery states. To enhance the robustness of future studies employing similar methodologies,

increasing the number of items measuring foundational attributes lower in the hierarchy is recommended.

ACKNOWLEDGMENTS

The authors would like to acknowledge the help of the editor and anonymous reviewers.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

-
- [1] E. Gaigher, J. M. Rogan, and M. W. H. Braun, Exploring the development of conceptual understanding through structured problem-solving in physics, *Int. J. Sci. Educ.* **29**, 1089 (2007).
 - [2] P. G. Hewitt, Millikan lecture 1982: The missing essential—A conceptual understanding of physics, *Am. J. Phys.* **51**, 305 (1983).
 - [3] P. Nieminen, A. Savinainen, and J. Viiri, Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010123 (2012).
 - [4] L. Bao and K. Koenig, Physics education research for 21st century learning, *Discip. Interdiscip. Sci. Educ. Res.* **1**, 2 (2019).
 - [5] A. B. Champagne, R. F. Gunstone, and L. E. Klopfer, A perspective on the differences between expert and novice performance in solving physics problems, *Res. Sci. Educ.* **12**, 71 (1982).
 - [6] M. B. Nakhleh, Are our students conceptual thinkers or algorithmic problem solvers? Identifying conceptual students in general chemistry, *J. Chem. Educ.* **70**, 52 (1993).
 - [7] A. E. Rivet and J. S. Krajcik, Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science, *J. Res. Sci. Teach.* **45**, 79 (2008).
 - [8] K. J. Roth, Developing meaningful conceptual understanding in science, in *Dimensions of Thinking and Cognitive Instruction* (Routledge, New York, 1990).
 - [9] W. Xu, Q. Liu, K. Koenig, J. Fritchman, J. Han, S. Pan, and L. Bao, Assessment of knowledge integration in student learning of momentum, *Phys. Rev. Phys. Educ. Res.* **16**, 010130 (2020).
 - [10] J. Piaget, *The Child's Conception of Physical Causality* (Littlefield, Adams & Co., Totowa, NJ, 1966).
 - [11] P. Potvin, L. Nenciovici, G. Malenfant-Robichaud, F. Thibault, O. Sy, M. A. Mahhou, A. Bernard, G. Allaire-Duquette, J. Blanchette Sarrasin, L.-M. Brault Foisy, N. Brouillette, A.-A. St-Aubin, P. Charland, S. Masson, M. Riopel, C.-C. Tsai, M. Bélanger, and P. Chastenay, Models of conceptual change in science learning: Establishing an exhaustive inventory based on support given by articles published in major journals, *Stud. Sci. Educ.* **56**, 157 (2020).
 - [12] R. K. Sawyer, *The Cambridge Handbook of the Learning Sciences* (Cambridge University Press, Cambridge, England, 2005).
 - [13] G. J. Posner, K. A. Strike, P. W. Hewson, W. A. Gertzog *et al.*, Accommodation of a scientific conception: Toward a theory of conceptual change, *Sci. Educ.* **66**, 211 (1982).
 - [14] *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge Taylor & Francis Group, New York, 2013), pp. 1–768.
 - [15] J. Minstrell, Explaining the “at rest” condition of an object, *Phys. Teach.* **20**, 10 (1982).
 - [16] J. J. Mintzes, J. H. Wandersee, and J. D. Novak, *Teaching Science for Understanding: A Human Constructivist View* (Academic Press, New York, 2005).
 - [17] A. A. DiSessa, N. M. Gillespie, and J. B. Esterly, Coherence versus fragmentation in the development of the concept of force, *Cogn. Sci.* **28**, 843 (2004).
 - [18] A. A. DiSessa *et al.*, A bird's-eye view of the “pieces” vs “coherence” controversy (from the “pieces” side of the fence), in *International Handbook of Research on Conceptual Change* (Routledge, New York, 2013), pp. 31–48.
 - [19] D. A. Norman and D. E. Rumelhart, The LNR approach to human information processing, *Cognition* **10**, 235 (1981).
 - [20] S. Vosniadou, Capturing and modeling the process of conceptual change, *Learn. Instr.* **4**, 45 (1994).
 - [21] M. T. Chi, J. D. Slotta, and N. De Leeuw, From things to processes: A theory of conceptual change for learning science concepts, *Learn. Instr.* **4**, 27 (1994).
 - [22] M. T. H. Chi, Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, New York, 2013), pp. 49–70.
 - [23] K. S. Taber, Understanding the nature and processes of conceptual change: An essay review. Review of Vosniadou, Stella (Ed.) (2008) *International Handbook of Research on Conceptual Change*. Vol. 14 No. 1, *Educ. Rev.* **14** (2011).
 - [24] D. R. Krathwohl, A revision of Bloom's taxonomy: An overview, *Theory Pract.* **41**, 212 (2002).
 - [25] J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980).
 - [26] O. L. Liu, H.-S. Lee, and M. C. Linn, Measuring knowledge integration: Validation of four-year assessments, *J. Res. Sci. Teach.* **48**, 1079 (2011).
 - [27] M. S. Sabella and E. F. Redish, Knowledge organization and activation in physics problem solving, *Am. J. Phys.* **75**, 1017 (2007).

- [28] J. L. Snyder, An investigation of the knowledge structures of experts, intermediates and novices in physics, *Int. J. Sci. Educ.* **22**, 979 (2000).
- [29] T. R. Sikorski and D. Hammer, A critique of how learning progressions research conceptualizes sophistication and progress, in *Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010)* (International Society of the Learning Sciences, Chicago, 2010), pp. 1032–1039, <https://repository.isls.org/handle/1/2641>.
- [30] A. A. Rupp, J. Templin and R. A. Henson, *Diagnostic Measurement: Theory, Methods, and Applications* (Guilford Press, New York, 2010).
- [31] *Learning Progressions in Science: Current Challenges and Future Directions*, edited by A. C. Alonzo and A. W. Gotwals (Sense Publishers, Rotterdam, 2012).
- [32] N. Frederiksen, R. J. Mislevy, and I. I. Bejar, *Test Theory for a New Generation of Tests* (Erlbaum, Hillsdale, NJ, 1993).
- [33] K. K. Tatsuoka, *Cognitive Assessment: An Introduction to the Rule Space Method* (Routledge, New York, 2009).
- [34] D. Tu, Y. Cai, and S. Ding, *Cognitive Diagnosis: Theory, Methods and Applications* (Beijing Normal University, Beijing, 2012).
- [35] S. Bai, Developing a learning progression for probability based on the GDINA model in China, *Front. Psychol.* **11**, 569852 (2020).
- [36] X. Wu, Y. Zhang, R. Wu, X. Tang, and T. Xu, Cognitive model construction and assessment of data analysis ability based on CDA, *Front. Psychol.* **13**, 1009142 (2022).
- [37] E. D. Corpuz and N. S. Rebello, Investigating students' mental models and knowledge construction of microscopic friction. I. Implications for curriculum design and development, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020102 (2011).
- [38] J. K. Gilbert and D. M. Watts, Concepts, misconceptions and alternative conceptions: Changing perspectives in science education, *Stud. Sci. Educ.* **10**, 61 (1983).
- [39] I. Aviani, N. Erceg, and V. Mešić, Drawing and using free body diagrams: Why it may be better not to decompose forces, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020137 (2015).
- [40] R. J. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- [41] I. P. Canlas, Using visual representations in identifying students' preconceptions in friction, *Res. Sci. Technol. Educ.* **39**, 156 (2021).
- [42] S. Prasitpong and R. Chitree, What Thai students think about directions and types of frictional forces, *AIP Conf. Proc.* **1263**, 66 (2010).
- [43] A. Boudreaux and A. Elby, How curriculum developers' cognitive theories influence curriculum development, *Phys. Rev. Phys. Educ. Res.* **16**, 020144 (2020).
- [44] J. Clement, Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics, *J. Res. Sci. Teach.* **30**, 1241 (1993).
- [45] P. Mulhall and R. Gunstone, Views about learning physics held by physics teachers with differing approaches to teaching physics, *J. Sci. Teach. Educ.* **23**, 429 (2012).
- [46] H. Ş. Kızılcık, M. Aygün, E. Ş Şahin, N. Önder-Çelikkanlı, O. Türk, T. Taşkın, and B. Güneş, Possible misconceptions about solid friction, *Phys. Rev. Phys. Educ. Res.* **17**, 023107 (2021).
- [47] C. C. Tsai and M. Lydia Wen, Research and trends in science education from 1998 to 2002: A content analysis of publication in selected journals, *Int. J. Sci. Educ.* **27**, 3 (2005).
- [48] C. O. Ameh and R. F. Gunstone, The understanding held by Nigerian science teachers of some science concepts, in *Proceedings of the Annual Meeting of the American Educational Research Association (AERA)*, New Orleans, LA (1988), ERIC No. ED292633.
- [49] U. Besson, L. Borghi, A. D. Ambrosio, and P. Mascheretti, A three-dimensional approach and open source structure for the design and experimentation of teaching-learning sequences: The case of friction, *Int. J. Sci. Educ.* **32**, 1289 (2010).
- [50] H. Caldas and É. Saltiel, Le frottement cinétique: Analyse des raisonnements des étudiants, *Didaskalia* **6**, 55 (1995).
- [51] A. B. Arons, *Teaching Introductory Physics* (John Wiley & Sons, New York, 1997).
- [52] T. C. Chia, Common misconceptions in frictional force among university physics students, *Teach. Learn.* **16**, 107 (1996).
- [53] D. R. Sari, D. Ramdhani, and H. K. Surtikanti, Analysis of elementary school students' misconception on force and movement concept, *J. Phys. Conf. Ser.* **1157**, 022053 (2019).
- [54] M. E. Loverude, C. H. Kautz, and P. R. L. Heron, Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas, *Am. J. Phys.* **70**, 137 (2002).
- [55] A. Vidak, N. Erceg, E. Hasović, S. Odvzak, and V. Mevsić, Teaching about rolling motion: Exploring the effectiveness of an extreme case reasoning approach, *J. Balt. Sci. Educ.* **17**, 511 (2018).
- [56] J. P. Leighton, M. J. Gierl, and S. M. Hunka, The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach, *J. Educ. Meas.* **41**, 205 (2004).
- [57] J. S. Gorin, Test construction and diagnostic testing, in *Cognitive Diagnostic Assessment for Education: Theory and Applications*, edited by J. P. Leighton and M. J. Gierl (Cambridge University Press, New York, 2007), pp. 173–202, [10.1017/CBO9780511611186](https://doi.org/10.1017/CBO9780511611186).
- [58] J. De La Torre, Dina model and parameter estimation: A didactic, *J. Educ. Behav. Stat.* **34**, 115 (2009).
- [59] S. Ha, G. Lee, and C. S. Kalman, Workshop on friction: Understanding and addressing students' difficulties in learning science through a hermeneutical perspective, *Sci. Educ.* **22**, 1423 (2013).
- [60] B. Wang, W. Han, Y. Zhang, Q. Wang, D. Li, Z. Tang, and Q. Kong, Assessment of student knowledge integration in learning friction force, *J. Balt. Sci. Educ.* **23**, 767 (2024).
- [61] J. de la Torre and C.-Y. Chiu, A general method of empirical Q-matrix validation, *Psychometrika* **81**, 253 (2016).
- [62] S. E. Whitely, Multicomponent latent trait models for ability tests, *Psychometrika* **45**, 479 (1980).

- [63] K. K. TATSUOKA, Rule space: An approach for dealing with misconceptions based on item response theory, *J. Educ. Meas.* **20**, 345 (1983).
- [64] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [65] Y. Nie, Y. Xiao, J. C. Fritchman, Q. Liu, J. Han, J. Xiong, and L. Bao, Teaching towards knowledge integration in learning force and motion, *Int. J. Sci. Educ.* **41**, 2271 (2019).
- [66] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/kcnz-wrtr> for friction force cognitive diagnostic test.
- [67] B. W. Junker and K. Sijtsma, Cognitive assessment models with few assumptions, and connections with nonparametric item response theory, *Appl. Psychol. Meas.* **25**, 258 (2001).
- [68] K. K. Tatsuoka, Caution indices based on item response theory, *Psychometrika* **49**, 95 (1984).
- [69] J. de la Torre, The generalized DINA model framework, *Psychometrika* **76**, 179 (2011).
- [70] M. E. Oliveri and M. von Davier, Investigation of model fit and score scale comparability in international assessments, *Psychol. Test Assess. Model.* **53**, 315 (2011).
- [71] Y. Cui and J. Li, Evaluating person fit for cognitive diagnostic assessment, *Appl. Psychol. Meas.* **39**, 223 (2015).
- [72] A. Maydeu-Olivares, Goodness-of-fit assessment of item response theory models, *Meas. Interdiscip. Res. Perspect.* **11**, 71 (2013).
- [73] J. Peugh and X. Fan, Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation, *Struct. Equ. Model.* **20**, 616 (2013).
- [74] J. Templin and L. Bradshaw, Measuring the reliability of diagnostic classification model examinee estimates, *J. Classif.* **30**, 251 (2013).
- [75] Y. Cui and J. P. Leighton, The hierarchy consistency index: Evaluating person fit for cognitive diagnostic assessment, *J. Educ. Measure.* **46**, 429 (2009).
- [76] J. M. Linacre, *A User's Guide to WINSTEPS-MINISTEP: Rasch-Model Computer Programs. Program Manual 3.68.0* (Winsteps.com, Chicago, IL, 2009).
- [77] X. Liu, *Using and Developing Measurement Instruments in Science Education: A Rasch Modeling Approach* (IAP, Charlotte, NC, 2010).
- [78] M. Planinic, W. J. Boone, A. Susac, and L. Ivanjek, Rasch analysis in physics education research: Why measurement matters, *Phys. Rev. Phys. Educ. Res.* **15**, 020111 (2019).
- [79] W. Ma and J. de la Torre, GDINA: An R package for cognitive diagnosis modeling, *J. Stat. Softw.* **93**, 1 (2020).
- [80] R. P. Chalmers, Mirt: A multidimensional item response theory package for the R environment, *J. Stat. Softw.* **48**, 1 (2012).
- [81] X. Wu, R. Wu, Y. Zhang, D. Arthur, and H.-H. Chang, Research on construction method of learning paths and learning progressions based on cognitive diagnosis assessment, *Assess. Educ.* **28**, 657 (2021).
- [82] *Taking Science to School: Learning and Teaching Science in Grades K-8*, edited by R. A. Duschl, H. A. Schweingruber, and A. W. Shouse (National Academies Press, Washington, DC, 2007).
- [83] K. Neumann, T. Vierung, W. J. Boone, and H. E. Fischer, Towards a learning progression of energy, *J. Res. Sci. Teach.* **50**, 162 (2013).
- [84] D. C. Briggs and A. C. Alonzo, The psychometric modeling of ordered multiple-choice item responses for diagnostic assessment with a learning progression, in *Learning Progressions in Science: Current Challenges and Future Directions*, edited by A. C. Alonzo and A. W. Gotwals (Sense Publishers, Rotterdam, 2012), pp. 293–316.
- [85] F. Chen, S. Zhang, Y. Guo, and T. Xin, Applying the rule space model to develop a learning progression for thermochemistry, *Res. Sci. Educ.* **47**, 1357 (2017).