

Investigating transitional processes in learning progressions: A case study of force and motion

Yi Ding^{1,*} and Lei Bao^{2,†,‡}

¹*School of Teacher Education, Nanjing University of Information Science and Technology, Nanjing 210044, China*

²*School of Biological Science and Medical Engineering, Research Center for Learning Science, Key Laboratory of Child Development and Learning Science (MOE), Southeast University, Nanjing, 211189, China*



(Received 20 January 2026; accepted 14 July 2026; published 6 August 2026)

Learning progressions (LPs) are widely used to describe the long-term development of students' understanding of core scientific ideas. However, as increasing attention has been paid to the context-dependent nature of student learning, the limitations of traditional LP models in capturing the complexity and dynamics of cognitive development have become more apparent. This study integrates the partial credit model (PCM) and model analysis to systematically investigate students' cognitive developmental characteristics within the Force and Motion Learning Progression (FM-LP). Using Force Concept Inventory data from 1619 Chinese students, we repeated the study by Fulmer *et al.* on U.S. students to examine the applicability of the established FM-LP across Chinese student populations. The results indicate that the established FM-LP does not exhibit satisfactory model fit in the Chinese sample, providing further evidence that science learning is highly context- and population-dependent. To characterize students' intermediate cognitive states, this study introduces a model-state combination analysis framework. The findings show that most students do not stably occupy a single LP level; instead, they display thinking patterns spanning multiple levels, with clear differences observed between middle school and high school students. Further analyses based on PCM reveal that, as students' ability increases, their model states shift from strongly mixed states to less mixed states and eventually toward the relatively consistent use of the scientific conception. The model-state-based approach offers a new perspective on capturing cognitive complexity in learning progressions and provides important implications for progression-oriented assessment.

DOI: [10.1103/5v86-hkrk](https://doi.org/10.1103/5v86-hkrk)

I. INTRODUCTION

Learning progressions have been a focus of science education researchers over the last two decades [1–5]. Generally defined, learning progressions (LPs) are descriptions of successively more sophisticated ways of thinking about a topic. Duncan and Hmelo-Silver [6] offered a four-part, comprehensive definition: (i) LPs focus on a few basic ideas and practices developed over time. (ii) LPs have an upper and lower bound. The upper bound describes what students are expected to know and is

determined by academic standards and research in the content area of the learning progression. The lower bound is determined by students' prior knowledge and skills as they enter the progression. (iii) LPs comprise levels that describe the steps students take between the lower and upper bounds. (iv) In essence, learning progressions aim to represent an instruction-based, idealized developmental pathway, along which students' conceptual understanding is expected to advance step by step toward increasingly sophisticated levels.

There are multiple well-defined LPs in physics education research. Some LPs focus narrowly on specific topics, such as force and motion [1], celestial motion [7,8], and buoyancy [9,10]. Others focus broadly on big ideas, such as energy [11] and matter [12,13]. There are also LPs related to scientific practices, such as scientific argumentation [2] and scientific modeling performance [14]. Methodologically, early research on LPs typically constructed hypothetical learning sequences according to studies on students' misconceptions or the logic of disciplinary development and then empirically validated the LP model using data

*Contact author: 004005@nuist.edu.cn

†Contact author: lei.bao@outlook.com

‡Address for correspondence: No.2 Southeast University Road, Jiangning District, Nanjing, JiangSu, China.

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

from structured interviews, open-ended responses, and/or multiple-choice assessments [1,6]. These models commonly assume that students' comprehension of science concepts can be categorized into distinct levels and students situated at a certain level can use a consistent understanding across different contexts [15].

As an increasing number of researchers have noted misalignments between LP-predicted levels and students' response patterns [16–18], they have begun to realize the limitations of LPs [15,19]. First, the early LP studies are based on theoretical ideas about science learning, which assume that students often develop coherent explanatory frameworks even before they encounter formal instruction [15]. Science learning is viewed as a process of conceptual change, in which students must abandon their existing theories [20,21]. Under the guidance of these ideas, many LPs describe the development of student thinking as transitions from many different non-normative conceptions to scientific conceptions [1] or from less to more sophisticated thinking levels [9–11] over an instructional unit or multiple years of schooling. For example, Alonzo and Steedle [1] describe LPs as “the thinking that students at that level could be expected to exhibit, including both the correct ideas that can be carried to the next level and the misconceptions that will need to be revised before students can reach the next level.”

Second, it is problematic to assign a level to a given student at a given time since individual students' thinking is context sensitive and embedded in their prior experience [22]. Scherr, for example, documented how a college student could easily distinguish mass and density in one context but conflated them in another [23]. Bao *et al.* found that a significant number of students showed both correct reasoning patterns and incorrect reasoning patterns in a set of questions about Newton's third law [24]. Evidence from the studies by Alonzo *et al.* revealed the coexistence of multiple levels of knowledge within individual students, such that the same student was classified at different learning progression levels based on their performance on ordered multiple-choice items versus open-ended questions [1,18]. Steedle and Shavelson's study [18] also showed that usually only a small portion of students reasoned using a systematically coherent level of ideas across items, but many other students did not.

Overall, the main problem with the existing research on LPs lies in the difficulty of conceptualizing the complexities and dynamics of the learning process [19]. As knowledge-in-pieces theory states, students' ideas should be regarded as a variety of p-prims that are activated depending on unique contextual features and conditions [25,26]. Performance on a given item depends on how contextual features interact with students' prior experience [24,27–29]. Certain situational features tend to activate particular p-prims, which are then appropriately connected to the current situation through a process known as context

shift [30]. Since students' cognition is a complex system comprising numerous elements, science learning involves the process of reorganizing these elements into a more scientifically productive knowledge structure rather than eradicating them [26,31]. The dynamics of student learning processes should be revealed in LP studies by showing the mixed intermediate states before students can consistently use scientific-level ideas [25,30].

In this sense, model analysis may provide some insight into expressing many coexisting and competing levels exhibited in the multifaceted pathways of the intermediate processes of learning progressions. Model analysis is a probabilistic framework motivated by quantum mechanical thinking, where a student's state of understanding is represented by a multidimensional model collection encompassing the correct model and some other incorrect models within each concept dimension [29]. Related to LP studies, a student's state of thinking can be represented with a model collection encompassing understanding corresponding to different LP levels. By associating multiple questions with diverse contexts to a target concept, this approach facilitates an exploration of the possibility that students may exhibit coherent levels of understanding or coexisting and competing levels of understanding in a set of questions, which can further depict the sophisticated and mixed states during the LPs. The consistent use of a certain model indicates that the student's understanding of a concept is relatively stable up to a certain level during the learning progression. In the developmental journey to consistently use the correct or the most sophisticated model, students often go through a stage of mixed use of multiple conceptual models, which indicates the coexistence and possible competition between different ideas in the process of learning [24,28,29].

To address these main concerns and problems in the existing LP research, this study focuses on three main questions:

RQ1: To what extent is the established FM-LP model population-dependent and context-dependent?

RQ2: What combinations of different FM-LP levels characterize students' intermediate model states along the learning progression?

RQ3: How do students' model states change as their ability increases?

II. ANALYSIS OF THE POPULATION-DEPENDENCE AND CONTEXT-DEPENDENCE OF THE FORCE AND MOTION LEARNING PROGRESSION

To explore RQ1, a prior study (Fulmer *et al.*, [4]) was repeated with a different sample in this section. In part A of this section, we describe the prior study, and then, in parts B–E, describe the replication undertaken in the current study.

TABLE I. Scheme for coding FCI responses according to the force and motion learning progression.

FCI item	Response				
	A	B	C	D	E
3	3	3	4	1	2
8	2	4	2	2	2
10	4	0	3	2	3
11	1	2	2	4	0
12	2	4	2	2	2
13	3	2	3	4	0
14	0	1	2	4	2
17	3	4	0	3	0
21	3	2	3	0	4
22	3	4	0	3	0
23	2	4	2	2	0
24	4	0	3	0	3
25	0	0	4	3	0
26	3	3	0	3	4
27	2	2	4	0	0
29	1	4	2	3	2
30	1	2	4	2	2

A. Prior study on force and motion learning progress

In the study by Fulmer *et al.*, the response options of 17 Force Concept Inventory (FCI) items were coded to match the proposed force and motion learning progression (FM-LP) model developed in the study by Alonzo and Steedle [1,4]. The coding scheme is shown in Table I. According to the FM-LP, correct responses were coded as 4 to indicate consistency with a Newtonian understanding of force and motion; that is, an understanding that the net force applied to an object is proportional to its resulting acceleration and that this force might not be in the direction of motion. Other responses were coded with values corresponding to the understanding patterns described by different LP levels, as shown in Table II. Students with level 3 understanding recognize that an object may remain at rest either because no forces act on it or because the net force is zero; however, they only have a partial understanding of the forces acting on moving objects, and their understanding remains incomplete. Students with level 2 understanding exhibit typical prescientific conceptions, believing that motion implies a force in the direction of motion and that the absence of motion implies the absence of force; conversely, they also believe that a force necessarily produces motion in its direction. This reflects an intuitive tendency to directly associate force with motion. Students with level 1 understanding regard force primarily as a push or pull but have not yet established a systematic relationship between force and motion, with their understanding remaining at an experiential and surface level. Students with level 0 understanding, by contrast, show responses that are far removed from scientific conceptions, reflecting

a lack of basic understanding of forces and often displaying irrelevant or incoherent ideas.

B. Method

To repeat the study by Fulmer *et al.*, data were collected from the Chinese population. The first sample consisted of 563 middle school students in grades 7–9. The grade 7 students were not yet enrolled in physics courses, while the grades 8 and 9 students were enrolled in middle school physics courses. The second sample consisted of 1056 high school students, all enrolled in high school physics courses. All data were collected near the middle of the fall semester in 2023. At this point, students in grades 8 and 9 of middle school have completed or are currently studying core topics related to force and motion. These topics typically include fundamental concepts such as Newton’s laws of motion, friction, velocity, and acceleration. High school students have entered a more advanced stage of mechanics learning, which usually involves the comprehensive application of Newtonian mechanics, vector analysis, and more complex tasks in physical modeling and problem solving. Thus, the sample consisted of students in a range of ages from middle school through high school with a variety of previous exposure to force and motion instruction.

A Chinese version of the 17 selected items (Q3, Q8, Q10, Q11, Q12, Q13, Q14, Q17, Q21, Q22, Q23, Q24, Q25, Q26, Q27, Q29, and Q30) from the revised FCI was used in this study [32]. Students were allotted 45 min to complete the test, and the majority of them were able to finish within the given time. For the raw scores on the FCI, which were scored dichotomously, out of the maximum of 17 points, the middle school sample had a mean of 6.18 (SD = 3.42) with scores ranging between 0 and 17, while the high school students had a mean of 10.03 (SD = 4.33) with scores ranging between 0 and 17.

Subsequently, we applied the coding scheme shown in Table I to process the data. The partial credit model (PCM) was used to repeat the main analysis in Fulmer’s study. The PCM, an extension of the Rasch model within the framework of item response theory, is designed for polytomous items with ordered categorical response structures. In the PCM analysis, the theta (θ) value represents each respondent’s latent ability on the measured dimension. Item difficulty refers to the location parameters of items or step thresholds on the latent continuum, indicating the level of ability required for a respondent to have a higher probability of achieving a given score category. Because items typically consist of multiple ordered score categories, item difficulty in the PCM is defined by a set of step thresholds rather than a single parameter. The typical ranges of θ and item difficulty values are between -3 and $+3$ logits. The infit and outfit mean square residual (MNSQ) statistics reflect how well the responses of individuals fit the model for those whose abilities are

TABLE II. Description of the levels of understanding about force and motion concepts.

Level	Description
4	Student understands that the net force applied to an object is proportional to its resulting acceleration (change in speed or direction) and that this force may not be in the direction of motion.
3	Student understands that an object is stationary either because there are no forces acting on it or because there is no net force acting on it. Student has a partial understanding of forces acting on moving objects.
2	Student believes that motion implies a force in the direction of motion and that nonmotion implies no force. Conversely, a student believes that force implies motion in the direction of the force.
1	Student understands force as a push or pull that may or may not involve motion.
0	Way off track.

close to (infit) or far from (outfit) the item's difficulty level. Acceptable infit and outfit values generally fall within the range of 0.7–1.3 and 0.5–1.5, respectively. Moreover, PCM models the probability of a respondent achieving higher response categories as a function of the difference between the respondent's ability and item difficulty. The transitions between adjacent response categories for an individual are governed by a series of thresholds. These thresholds reflect the level of ability required for a respondent to move from a lower-level response to a higher-level response [33].

C. Comparison between the results of this study and those of Fulmer *et al.*

Table III presents the difficulty estimates and the infit and outfit statistics of each item obtained from this study

and the study by Fulmer *et al.* For an easier comparison, the difficulty estimates of the 17 items from the two studies are also presented in Fig. 1.

As shown in Table III and Fig. 1, the estimates of the same item's difficulty among the two populations are considerably different. For some items (Q11, Q13, Q14, Q21, Q22, Q23, Q24, Q27, and Q30), the difficulty estimates using the Chinese sample data are higher than those of the U.S. students, while for other items (Q3, Q8, Q10, Q12, Q17, Q25, Q26, Q29), the trend is reversed. In addition, the infit statistics in this study have a mean square of 1.51, 1.36, 1.36, and 1.56 for items 14, 17, 24, and 25, which are all beyond the acceptable range. The large infit values on items 14, 17, 24, and 25 indicate that some students whose abilities are close to the difficulties of these items have not responded in a way consistent with the

TABLE III. Item difficulty estimates and fit statistics for FCI items coded according to the FM-LP in this study and the study by Fulmer *et al.*

Item	Item difficulty estimates		Infit MNSQ		Outfit MNSQ	
	This study	The study by Fulmer <i>et al.</i>	This study	The study by Fulmer <i>et al.</i>	This study	The study by Fulmer <i>et al.</i>
3	−0.87	−0.49	1.21	1.02	1.12	1.05
8	0.32	0.56	0.57	0.96	0.77	0.98
10	0.03	0.30	1.06	0.90	1.09	0.91
11	0.28	0.14	0.74	1.09	0.95	1.08
12	−0.21	0.17	0.78	0.99	0.98	0.99
13	−0.17	−0.36	0.79	0.96	0.89	0.95
14	0.47	0.33	1.51	0.97	1.53	0.94
17	−0.08	0.21	1.36	1.01	1.24	0.99
21	0.25	−0.02	0.99	1.05	1.03	1.07
22	0.09	−0.15	0.85	1.05	0.78	1.08
23	0.37	−0.09	0.85	1.05	0.95	1.03
24	0.05	−0.26	1.36	0.96	1.25	0.96
25	−0.02	0.44	1.56	1.03	1.28	1.05
26	0.17	0.31	0.84	0.99	0.87	0.92
27	0.01	−0.46	1.02	0.98	1.16	0.98
29	−0.71	0.27	1.17	0.99	1.40	1.00
30	0.01	−0.29	0.85	1.00	1.46	1.01

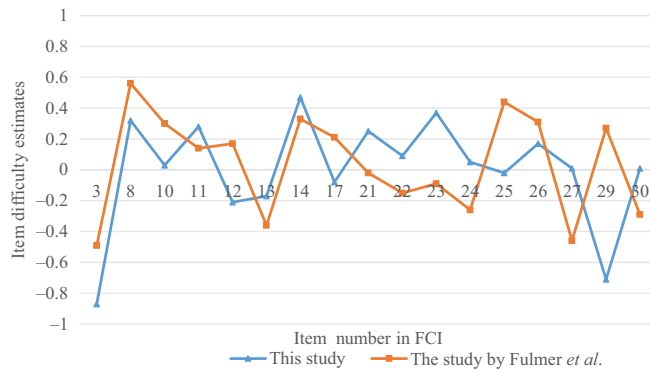


FIG. 1. Item difficulty estimates for FCI in this study and the study by Fulmer *et al.*

prediction of the FM-LP model. The large outfit value on item 14 indicates that students whose abilities are far from the difficulty level of item 14 have responded very differently from the FM-LP's predictions. In contrast, the mean-square fit statistics for the population in the study by Fulmer *et al.* were all within the acceptable ranges, which indicates that the U.S. students' responses to all items fit the FM-LP model well.

A Wright map is constructed in Fig. 2, plotting the students' ability estimates and the item thresholds by estimated difficulty on the same logit scale. In the PCM framework, logits are log-odds units used to place both person ability and item difficulty on a common interval scale. The logit scale typically ranges from -3 to $+3$, where values around 0 indicate average ability, negative values indicate below-average ability, and positive values indicate above-average ability. When a person and an item option are located at the same position on the logit scale, the person has a 50% probability of selecting the option corresponding to a specific level [33]. In Fig. 2, student ability estimates are plotted on the left side of the axis, item difficulty estimates for answer options are plotted on the right side of the axis, and the logit scale is displayed on the far left side of the figure. On the axis of the Wright map, M indicates the mean location (either the ability mean or the item difficulty mean), S represents the positions of ± 1 standard deviation, and T marks the boundaries for outliers (the regions of very high or very low ability or difficulty).

The Wright map can clearly show the thresholds of response categories, which represent the ability points at which students are likely to shift from selecting a lower-level response option to a higher-level one. For example, in Fig. 2, the ability point at which U.S. students shift from selecting level 1 response options to selecting level 2 response options is around -1 logit, and the threshold values of different levels of response for the U.S. students can be clearly distinguished from one another on the Wright map. However, the thresholds in the Chinese sample are more mixed; the thresholds for level 1 response

options overlap almost completely with those of level 2 response options, and the thresholds for level 3 response options also overlap with those of level 4 to a considerable extent.

D. Discussion

The above results reveal that the same item exhibits different difficulty across populations, indicating a form of sample dependence in item functioning. Such differences may stem from variations in item contexts and linguistic cues, which can activate different cognitive elements in students' thinking. At a more macrolevel, they may also arise from differences in curricula and instructional experiences shaped by students' distinct cultural and educational backgrounds [34,35]. For example, items 23 and 27 were more difficult for the Chinese student sample than for the U.S. sample, with a considerable proportion of Chinese students selecting level 2 options (54% and 31%, respectively). Both items examine the subsequent motion of an object after a force acting on it is suddenly removed. In this situation, the naïve conception that "no force implies no motion" (corresponding to the level 2 idea) appears to be more deeply entrenched among Chinese students than among their U.S. counterparts. At the curricular and instructional level, although Newton's first law is typically presented in textbooks as a core conclusion, it tends to receive insufficient emphasis in classroom instruction in China. In particular, instructional modeling of the transition from force removal to the continuity of subsequent motion, as well as exposure to diverse contextualized situations, is limited. As a result, students have fewer opportunities to systematically engage with integrative problems that require the simultaneous coordination of changes in force, the direction of velocity, and the magnitude of velocity. In contrast, items 25 and 29 were noticeably more difficult for U.S. students than for Chinese students. Both items assessed students' understanding of the forces acting on objects in equilibrium, specifically a box moving at a constant speed and a stationary office chair. In the Chinese context, equilibrium situations—such as objects at rest or moving at constant velocity—are heavily emphasized through standard free-body diagram analysis, which may strengthen students' procedural fluency in identifying balanced forces.

In addition, the misalignments between FM-LP-predicted levels and Chinese students' response patterns in some items in this study are consistent with many previous studies, which have found that a good fit is difficult to achieve [16,17,36]. Fit to data is the criterion of good scientific models on which LP researchers have focused [37]. The infit and outfit outliers in the Chinese population for some items may imply that the difficulty of these items is influenced not only by thinking levels but also by factors such as item context and representational forms. That is, item functioning is not only the result of interaction

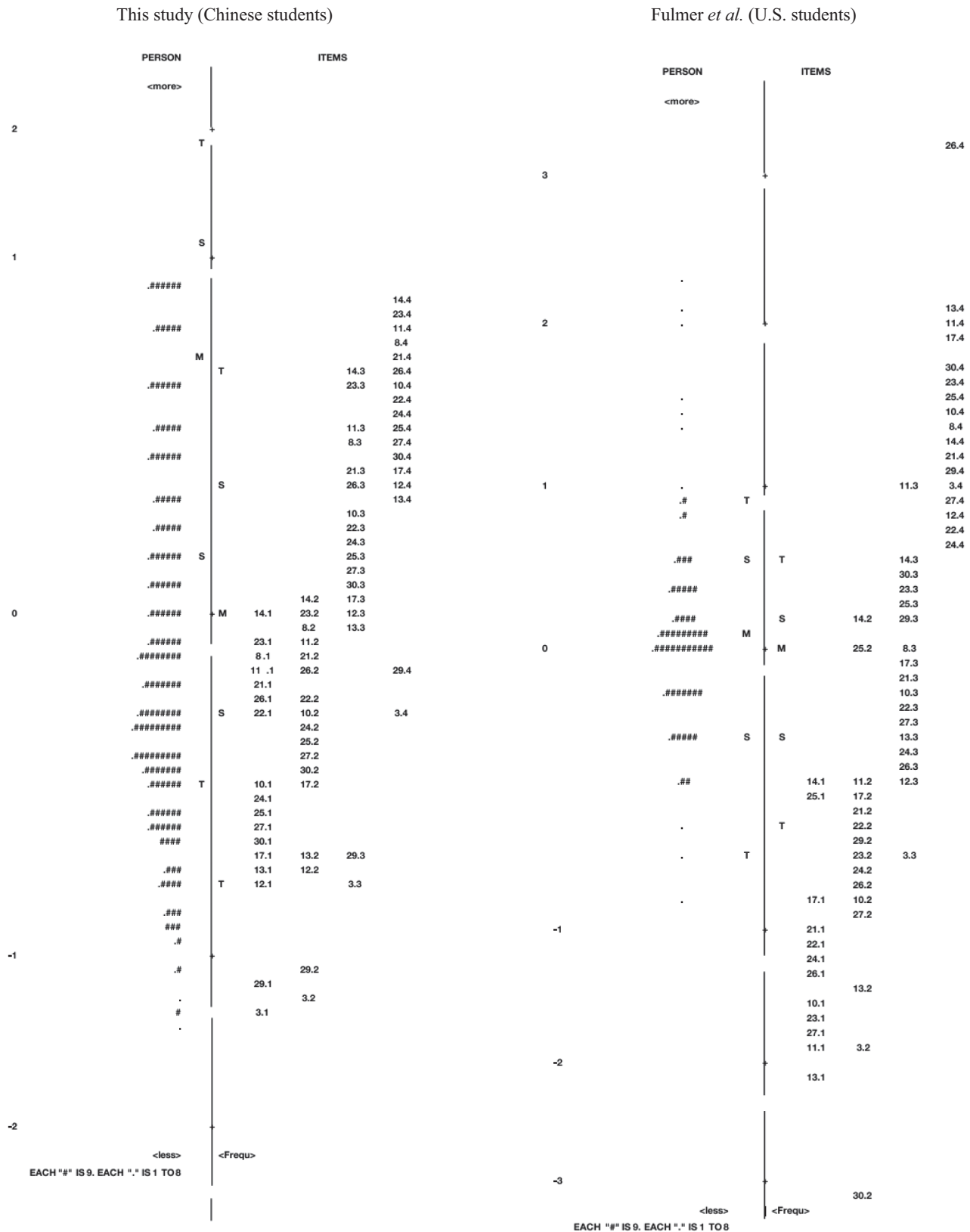


FIG. 2. Wright map for FCI item responses when coded according to the proposed LP, showing distribution of persons by estimated ability (left side of axis) and items by estimated difficulty (right side of axis). Each ‘.’ represents 1–8 persons; each ‘#’ represents 9 persons. The item responses are identified by their question number on the FCI and the LP level (e.g., Q13.4 is the LP level 4 response on item 13).

between students and the concept itself but also interaction with specific context, structure, and language features of items. This interaction between cognitive resources and task features is well documented in science education [28,38,39]. Moreover, in certain situations, students may rely on informed guessing, which highly influences the fit between student response patterns and the predicted FM-LP model [28].

Ideally, in the learning progression model, the threshold values on the Wright map should be progressively ordered, reflecting a gradual increase in students' thinking levels with the increase in estimated ability [40]. Clearly ordered thresholds indicate that response categories function as intended, which successfully differentiate students along a unidimensional latent trait, such as the conceptual understanding of force and motion. In the U.S. sample, the Wright map clearly shows a set of separated and monotonically increasing thresholds across the four thinking levels. This suggests that the development of students' thinking aligns well with the proposed FM-LP structure: as students' latent ability increases, they systematically shift toward higher levels of the LP with more normative scientific responses. The FM-LP, in this case, functions as a valid interpretive model that captures how students' thinking evolves in a structured and predictable manner.

In contrast, the Chinese sample exhibits substantial overlaps in adjacent thresholds on the Wright map, particularly between levels 1 and 2, and between levels 3 and 4, showing that the LP levels can no longer effectively distinguish students of different abilities. This result suggests that, on the one hand, learning progression models need to be adjusted and refined for different student populations. For example, in the Chinese student sample, the proportion of students selecting level 1 answer choices is extremely low (with the highest percentage being 3% of students selecting the level 1 option on item 14). Therefore, within the context of the sample used in this study, it may be appropriate to exclude the typical ideas represented by level 1 from the FM-LP model. On the other hand, the threshold mixing indicates that students do not shift from lower-level response options to higher-level ones in the orderly progression predicted by the FM-LP model as their abilities increase. While this may partly result from sample-specific characteristics, such as greater heterogeneity in thinking patterns, it also points to a deeper structural issue: that students' conceptual development is not universally stagelike, but instead reflects a mixed use of different thinking patterns at the same time.

E. Summary

These findings challenge the LP's intent and ability to assign a level to a given student at a given time and the PCM measurement principle that thresholds should ideally be ordered [40]. This, in turn, underscores the importance

of moving beyond a stagelike LP model toward an approach that embraces contextual variation [28], localized coherence [25], and the dynamic activation of knowledge resources [41]. In other words, rather than treating LP levels as fixed constructs, LP frameworks should be adapted to reflect how students in different cultural or curricular contexts may interpret and respond differently to scientific ideas.

III. CHANGES IN STUDENTS' MODEL STATES ALONG A LEARNING PROGRESSION

This section focuses on the structural characteristics and dynamic evolution of students' model states. For RQ2, it identifies the types of intermediate model states represented by different combinations of FM-LP levels, thereby characterizing the diversity of students' cognitive structures along the learning progression. For RQ3, it examines the transition patterns and change trajectories of model states as students' ability increases, revealing the dynamic developmental features of conceptual understanding in science learning.

A. Method

According to prior studies, for the topic of force and motion, we can build a model space for each student with models corresponding to different LP levels [28]. Based on the coding method discussed in Table I, if student A selects level 0 options in 7 out of 17 items, level 1 options in 1 item, level 2 options in 4 items, level 3 options in 3 items, and level 4 options in 2 items, the student will obtain a model space as illustrated in Table IV.

We then operationally defined students' model states as follows. Based on these definitions, each student was assigned to a specific model state.

- (1) A student was classified as being "stable at a given level" when the probability of using a particular level exceeded 70%, while the probabilities of using each of the other levels were below 20%.
- (2) When the probabilities of using two levels were both above 20%, the student was classified as exhibiting a "mixed 2,4" state, a "mixed 3,4" state, and so on.
- (3) When the probabilities of using three levels were all above 20%, the student was classified as exhibiting a "three-level mixed" state.
- (4) When the probability of using a particular level was greater than 50% but less than 70%, the student was

TABLE IV. Give an example of the model space of student A.

Student A	Level 0	Level 1	Level 2	Level 3	Level 4
Number	7	1	4	3	2
Percentage	41.2%	5.9%	23.5%	17.6%	11.8%

TABLE V. Distribution of middle school and high school students across different model states.

	Other low frequency model states	Three-level mixed	Mixed 2, 3	Mixed 2, 4	Mixed 3, 4	Hybrid state dominated by level 4	Stable at level 4
All students ($N = 1619$)	4% (60)	26% (425)	7% (107)	18% (284)	9% (138)	9% (141)	29% (464)
Middle school students ($N = 563$)	5% (30)	39% (221)	14% (76)	18% (100)	9% (52)	6% (34)	9% (50)
High school students ($N = 1056$)	3% (30)	19% (204)	3% (31)	17% (184)	8% (86)	10% (107)	39% (414)

classified as being in a “hybrid state dominated by that level.”

To examine the robustness of the model-state classifications, alternative cutoff schemes (75%/55%/25%) were applied, and the resulting classifications were compared with those obtained using the original thresholds. Although minor shifts in classification occurred, the majority of participants were assigned to the same model states (approximately 87% agreement), and the main findings were unaffected. In particular, the identification of dominant states and the prevalence of mixed states remained stable across alternative specifications, suggesting that the classification scheme is robust for capturing the overall distribution patterns and developmental trends of students’ model states.

B. Results

The number and percentage of students in each model state are summarized in Table V. In addition, to compare differences in model-state distributions between middle school and high school students, their data are also reported separately in the table.

Overall, 29% of the students were classified as being in a consistent state at level 4, whereas only one student was found to be stable at level 2, and no students were found to be stably positioned at level 0, level 1, or level 3. A substantial proportion of students were characterized by three-level mixed states and mixed states consisting of levels 2 and 4, accounting for 26% and 18% of the sample, respectively. Other combinations of levels were observed much less frequently, with the percentages below 10%.

A marked difference was observed between middle school and high school students. Middle school students were more likely to be classified into mixed model states, particularly the three-level mixed state (39%), the mixed 2,4 state (18%), and the mixed 2,3 state (14%), indicating a higher prevalence of transitional conceptual structures. In contrast, high school students were predominantly classified into the consistent-at-level-4 state (39%), and the percentages of students classified into three-level mixed states and most two-level mixed states decreased compared with those of middle school students, suggesting a greater tendency toward more consolidated and consistent understanding. In addition, hybrid states dominated by

TABLE VI. The change of the students’ model states as their estimated ability (θ) increases.

Theta (from low to high)	Other low frequency model states	Three-level mixed	Mixed 2,3	Mixed 2,4	Mixed 3,4	Hybrid dominated by level 4	Stable at level 4
Low θ $\theta \in (-1.19, -0.57)$ ($N = 327$)	15% (48)	45% (146)	28% (92)	11% (36)	2% (5)	0 (0)	0 (0)
Medium-low θ $\theta \in (-0.51, 0.06)$ ($N = 702$)	4% (26)	38% (268)	2% (15)	31% (218)	14% (95)	11% (76)	1% (4)
Medium-high θ $\theta \in (0.13, 0.77)$ ($N = 369$)	0(0)	3% (11)	0(0)	8% (30)	10% (38)	14% (51)	65% (239)
High θ $\theta \in (0.90, 1.46)$ ($N = 221$)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	100% (221)

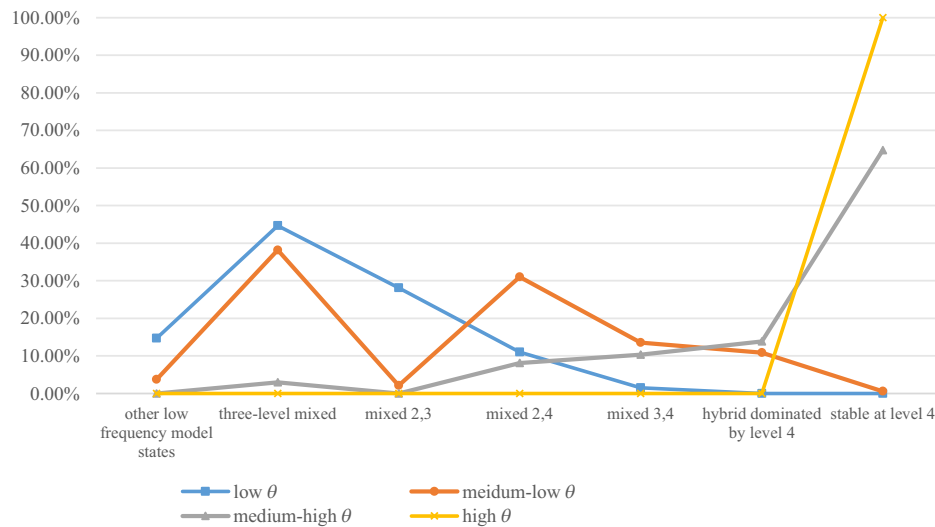


FIG. 3. Percentage of students with different model states across the four estimated ability intervals.

level 4 were more common among high school students (10%) than among middle school students (6%).

To further investigate the dynamic changes in students' model states as their ability developed, all students were grouped into four θ intervals based on the ability estimates (θ) obtained from the PCM analysis, ordered from lower to higher values around the mean. For each theta θ interval, the number and percentage of students associated with different model states were calculated, as summarized in Table VI. To provide a more intuitive visualization of how model states change with estimated ability, the data in Table VI are presented graphically in Fig. 3. In Fig. 3, the horizontal axis represents different model states, while the vertical axis shows the percentage of students. Interval 1 corresponds to the group with the lowest mean θ value, whereas interval 4 represents the group with the highest mean θ value.

Table VI and Fig. 3 reveal a clear and systematic shift in students' model-state distributions as estimated ability (θ) increases. At lower θ intervals (interval 1 and interval 2), students are predominantly characterized by mixed model states, whereas higher θ levels (interval 3 and interval 4) are increasingly associated with stable and higher-level model states. This pattern suggests a gradual transition from fragmented and cross-level thinking toward more coherent and consistent thinking as students' proficiency develops.

In interval 1 ($N = 327$), students were overwhelmingly characterized by mixed model states. Three-level mixed states accounted for the largest proportion (45%), followed by mixed 2,3 states (28%) and mixed 2,4 states (11%). Consistent thinking states were almost absent, with no students stably positioned at level 4 and only one student at level 2. These results indicate that students at lower ability levels rely on fragmented thinking patterns spanning multiple levels. In interval 2 ($N = 702$), mixed states

remained dominant; however, the distribution became more concentrated. The proportion of three-level mixed states decreased to 38%, while mixed 2,4 (31%) and mixed 3,4 states (14%) became more prominent. At the same time, hybrid states dominated by level 4 emerged (11%), and consistent thinking states at level 4 (1%) began to appear, although they remained rare. This pattern suggests an ongoing reorganization of thinking structures toward higher-level models.

A pronounced shift occurred in interval 3 ($N = 369$). Stable-at-level-4 states became the dominant category, accounting for 65% of students. Mixed states declined sharply, with three-level mixed states dropping to 3% and mixed 2,4 and 3,4 states to 8% and 10%, respectively. Lower-level and broader mixed states were no longer observed. These results indicate a substantial consolidation of thinking at the highest level. In interval 4 ($N = 221$), all students (100%) were classified as being stably positioned at level 4, with no mixed or hybrid states observed. This complete convergence suggests that students at the highest ability level consistently and coherently employ the highest-level thinking model.

C. Discussion

The numbers and percentages of students in each state suggest that most students, particularly middle school students, were unable to consistently employ a single level of understanding. Instead, the majority exhibited thinking patterns spanning multiple levels. For example, among the three-level mixed state, the most common response pattern involved a combination of levels 2, 3, and 4. Student B provides a representative case. This student's responses were classified as level 4 for items 3, 14, 25, and 29; level 3 for items 13, 17, 21, 22, 24, and 26; and level 2 for items 8, 10, 11, 12, 27, and 30. A closer analysis reveals that items 25, 17, and 11 essentially assess

the same underlying concept: an object undergoing uniform linear motion (in equilibrium) experiences a net force of zero. However, the student's understanding differed substantially across these contexts. In item 25, the student's response correctly recognized that the pushing force acting on a box moving at a constant speed is balanced by the resistive force, leading to a net force of zero. In contrast, in item 17, the selected response was classified as level 3 and corresponded to the idea that an elevator moving upward at a constant speed experiences a greater upward force than downward force, indicating only a partial understanding of forces acting on moving objects. In the ice hockey scenario presented in item 11, the selected response was classified as level 2 and corresponded to the idea that a hockey puck moving across a smooth surface is acted upon by a force in the direction of motion.

The inconsistent response patterns exhibited by students across different contexts may be attributed to several major factors. First, differences in contextual representations and surface features may play an important role in shaping students' thinking across tasks. In item 25, the box moving at a constant velocity represents a prototypical context that is repeatedly encountered in physics instruction in China. The problem explicitly highlights the opposition between the applied force and the resistive force, which may draw students' attention to force balance and thereby activate level 4 understanding. By contrast, the elevator context in item 17 involves vertical motion and gravitational effects, making the situation more complex. The intuitive representation of the elevator "moving upward" may lead students to infer that the upward force must be greater than the downward force, thus favoring level 3 thinking, which reflects a partial understanding of the forces acting on a moving object. Second, the activation of everyday experience and intuitive knowledge differs substantially across contexts. The elevator is a highly familiar real-life situation in which students' embodied experiences—such as the sensation of being "lifted" during upward motion—can reinforce the intuitive belief that a larger force must act in the direction of motion. Similarly, in the ice hockey context of item 11, students' everyday experiences with motion dominated by friction may lead them to believe that an object must continue to be acted upon by a force in the direction of motion, resulting in level 2 understanding. Such intuitive experiences may, in specific contexts, suppress students' access to formalized physics principles.

The difference in prominent model states between middle school and high school students indicates a developmental trend from mixed and transitional model combinations toward more consistent and higher-level model states along the learning progression over an extended period of time. After years of structured science learning, high school students typically possess a more coherent framework of mechanics, with strong conceptual

connections and integration. As a result, their progress in understanding specific concepts tends to be less influenced by context. In contrast, middle school students often have underdeveloped cognitive structures and weaker links between concepts. Even when they apply correct scientific models in certain tasks, these may represent only localized activations that have not yet been integrated into a broader knowledge system. Consequently, their conceptual understanding is characterized by a series of mixed-level states reflecting competing modes of thinking.

As described by Corcoran *et al.* [37], levels in a learning progression are often assumed to denote periods of consistency in students' knowledge. From this perspective, a student who gives a level 2 response on a Newton's third law item, for example, would be expected to give similar responses across all Newton's third law items. Contrary to this expectation, the results of the present study indicate that only the scientific conception (level 4) is stably used by students at higher estimated ability levels, whereas no other alternative levels were found to be consistently adopted by any specific group of students. In other words, before students are able to stably apply the scientific conception, they do not use any single non-normative idea consistently. This can be explained through the lens of the knowledge-in-pieces theory. Students' non-normative ideas are not well structured or internally coherent alternative theories; rather, they consist of multiple context-sensitive p-prims that are temporarily activated and assembled in response to specific task features [31,42]. Consequently, compared with the scientific conception, non-normative ideas often lack internal coherence and generative power, making them difficult to be invoked in a stable and systematic manner across different items and contexts.

D. Summary

In this section, we provide evidence that students' conceptual development does not proceed in the manner proposed by some LPs. In contrast, students' learning is consistently characterized by a mixed use of thinking across different levels. As students' ability increases, this mixture becomes less uncertain, with a higher probability of employing higher-level, scientifically normative thinking. In this sense, conceptual development is better understood not as the complete eradication of non-normative ideas, but as an increasing capacity to inhibit the activation of non-normative conceptions in favor of more productive ones [43].

Under these circumstances, we argue that more advanced stages of the learning progression should be reflected in a gradual concentration of thinking toward one or a small number of levels—that is, a progression from a broad mixture of thinking across multiple levels or a consistent use of lower-level thinking to a mixed state dominated by one level, and eventually to the consistent

use of the scientific conception. Compared with traditional level-based representations, this model-state-based approach offers a distinct advantage in that it makes it possible to “explore possibilities” for understanding how the cognition system may behave under different scenarios [44]. That is, although the proportions of students occupying different combinations of model states vary across ability groups, the possible combinations themselves remain relatively stable.

IV. CONCLUSION

This study employed both PCM and model analysis to investigate the appropriateness and developmental characteristics of the force and motion learning progression. This new model-state-based analytical approach provides fresh evidence for the knowledge-in-pieces theory of students’ science learning, and offers insights into the complex, intermediate states of learning progressions that have been largely underexplored in the existing literature.

First, the results indicate that the established FM-LP is not fully stable across different populations, but instead exhibits a certain degree of contextual and population dependence. By repeating Fulmer *et al.*’s [4] study using FCI data from Chinese middle school and high school students, we found substantial differences in item difficulty parameters between the Chinese and U.S. samples. Several items that showed good model fit in the U.S. sample failed to fit the model adequately in the Chinese sample. In addition, the Wright map results revealed considerable overlap among the existing FM-LP levels, suggesting that these levels do not effectively distinguish Chinese students’ ability. Taken together, learning progression models should be empirically examined when applied across cultures or student populations, rather than being treated as context-independent pathways of cognitive development.

Second, by introducing a model-state-based analysis framework, this study reveals intermediate cognitive states that are difficult to capture using traditional single-level-based learning progression models. The results show that most students do not stably occupy a single FM-LP level; instead, they exhibit combinations of thinking patterns spanning multiple levels. Moreover, these combination patterns differ between middle school and high school students. Such “mixed” model states may reflect the transitional and unstable nature of students’ conceptual development, highlighting its nonlinear characteristics and suggesting the complexity of the ways students draw upon different thinking resources across contexts.

Finally, based on ability estimates (θ) derived from the PCM, this study further examines the dynamic changes in students’ model states as ability increases. The results indicate that, with increasing ability, students’ thinking patterns in this study shift from highly dispersed mixtures across multiple levels toward mixtures dominated by

higher-level thinking—particularly the scientific conception—and eventually toward the relatively consistent use of the scientific conception. This developmental trajectory does not involve prolonged stability at any single incorrect level.

However, we do not rule out an alternative possibility that the observed instability of non-normative levels in this study may be due to the multifaceted dependence exhibited within this specific sample on non-normative ideas in the domain of force and motion. The study by Fulmer *et al.* suggests that U.S. students exhibit a degree of consistency aligned with FM-LP levels when responding to the FCI. This may be attributed to the more systematic use of non-normative ideas within their sample, which enables students to consistently invoke the same explanatory framework across different contexts. If a different population or topic were examined, coherent and sustained use of a particular non-normative level of thinking may emerge [18,29].

Overall, the findings suggest that the progression of student learning may be better understood as a process of continual reorganization and convergence of LP-level combinations, rather than as linear transitions between discrete levels. The model-state-based representational approach offers a new analytical lens for capturing cognitive complexity in learning progressions and provides important empirical evidence for refining and applying learning progression models across diverse student populations. For science education researchers and teachers, this study calls for caution against directly transferring or applying a particular progression pathway across different groups when conducting progression-based instructional assessment or curriculum design [45]. Instead, a contextualized interpretation of models should be adopted, fully respecting the diversity of learners’ sociocultural and cognitive backgrounds [46] to provide more fine-grained and appropriate guidance on curriculum and instructional designs.

ACKNOWLEDGMENTS

This research is supported by the Startup Fund of Southeast University (4007032509), the Fundamental Research Funds for the Central Universities (2242026K30053), and the Jiangsu Province Education Science Planning Project (B/2025/01/26).

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

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] A. C. Alonzo and J. T. Steedle, Developing and assessing a force and motion learning progression, *Sci. Educ.* **93**, 389 (2009).
- [2] L. K. Berland and K. L. McNeill, A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts, *Sci. Educ.* **94**, 765 (2010).
- [3] 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).
- [4] G. W. Fulmer, L. L. Liang, and X. Liu, Applying a force and motion learning progression over an extended time span using the force concept inventory, *Int. J. Sci. Educ.* **36**, 2918 (2014).
- [5] Y. Gao, X. Zhai, Y. Cui, T. Xin, and O. Bulut, Re-validating a learning progression of buoyancy for middle school students: A longitudinal study, *Res. Sci. Educ.* **52**, 1761 (2022).
- [6] R. G. Duncan and C. E. Hmelo-Silver, Learning progressions: Aligning curriculum, instruction, and assessment, *J. Res. Sci. Teach.* **46**, 606 (2009).
- [7] J. D. Plummer and L. Maynard, Building a learning progression for celestial motion: An exploration of students' reasoning about the seasons, *J. Res. Sci. Teach.* **51**, 902 (2014).
- [8] J. D. Plummer and J. Krajcik, Building a learning progression for celestial motion: Elementary levels from an earth-based perspective, *J. Res. Sci. Teach.* **47**, 768 (2010).
- [9] S. Paik, G. Song, S. Kim, and M. Ha, Developing a four-level learning progression and assessment for the concept of buoyancy, *Eurasia J. Math. Sci. Technol. Educ.* **13**, 4965 (2017).
- [10] Y. Gao, X. Zhai, B. Andersson, P. Zeng, and T. Xin, Developing a learning progression of buoyancy to model conceptual change: A latent class and rule space model analysis, *Res. Sci. Educ.* **50**, 1369 (2020).
- [11] H. Lee and O. L. Liu, Assessing learning progression of energy concepts across middle school grades: The knowledge integration perspective, *Sci. Educ.* **94**, 665 (2010).
- [12] S. Y. Stevens, C. Delgado, and J. S. Krajcik, Developing a hypothetical multi-dimensional learning progression for the nature of matter, *J. Res. Sci. Teach.* **47**, 687 (2010).
- [13] X. Liu and K. M. Lesniak, Students' progression of understanding the matter concept from elementary to high school: Matter concept, *Sci. Educ.* **89**, 433 (2005).
- [14] Y. M. Bamberger and E. A. Davis, Middle-school science students' scientific modelling performances across content areas and within a learning progression, *Int. J. Sci. Educ.* **35**, 213 (2013).
- [15] A. C. Alonzo and A. Elby, Beyond empirical adequacy: Learning progressions as models and their value for teachers, *Cognit. Instr.* **37**, 1 (2019).
- [16] E. M. Furtak, D. Morrison, and H. Kroog, Investigating the link between learning progressions and classroom assessment: Learning progressions and classroom assessment, *Sci. Educ.* **98**, 640 (2014).
- [17] K. Neumann, T. Viering, W. J. Boone, and H. E. Fischer, Towards a learning progression of energy, *J. Res. Sci. Teach.* **50**, 162 (2013).
- [18] J. T. Steedle and R. J. Shavelson, Supporting valid interpretations of learning progression level diagnoses, *J. Res. Sci. Teach.* **46**, 699 (2009).
- [19] T.-R. Sikorski and D. Hammer, A critique of how learning progressions research conceptualizes sophistication and progress, in *Proceedings of the International Conference of the Learning Sciences (ICLS)* (International Society of the Learning Sciences, 2010), Vol. 1.
- [20] S. Carey and E. Spelke, Domain-specific knowledge and conceptual change, in *Mapping the Mind*, edited by L. A. Hirschfeld and S. A. Gelman (Cambridge University Press, Cambridge, England, 1994), p. 200.
- [21] S. Vosniadou, On the nature of naïve physics, in *Reconsidering Conceptual Change: Issues in Theory and Practice*, edited by M. Limón and L. Mason (Kluwer Academic Publishers, Dordrecht, 2002), p. 76.
- [22] C. Cutting and T. Lowrie, Bounded learning progressions: A framework to capture young children's development of mathematical activity in play-based contexts, *Math. Educ. Res. J.* **35**, 317 (2023).
- [23] R. E. Scherr, Gesture analysis for physics education researchers, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010101 (2008).
- [24] L. Bao, K. Hogg, and D. Zollman, Model analysis of fine structures of student models: An example with Newton's third law, *Am. J. Phys.* **70**, 766 (2002).
- [25] A. A. diSessa, The construction of causal schemes: Learning mechanisms at the knowledge level, *Cognit. Sci.* **38**, 795 (2014).
- [26] A. A. diSessa and B. L. Sherin, What changes in conceptual change?, *Int. J. Sci. Educ.* **20**, 1155 (1998).
- [27] L. Bao and J. C. Fritchman, Knowledge integration in student learning of Newton's third law: Addressing the action-reaction language and the implied causality, *Phys. Rev. Phys. Educ. Res.* **17**, 020116 (2021).
- [28] L. Bao and E. F. Redish, Model analysis: Representing and assessing the dynamics of student learning, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010103 (2006).
- [29] Y. Ding, G. Zhu, Q. Bian, and L. Bao, Analysis of students' conceptual change in learning Newton's third law with an integrated framework of model analysis and knowledge integration, *Phys. Rev. Phys. Educ. Res.* **20**, 020141 (2024).
- [30] T. G. Amin, Articulating knowledge-in-pieces with other theories of conceptual change, *Human Dev.* **60**, 44 (2017).
- [31] A. A. diSessa, A friendly introduction to "knowledge in pieces": Modeling types of knowledge and their roles in learning, in *Invited Lectures from the 13th International Congress on Mathematical Education*, edited by G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, and B. Xu (Springer International Publishing, Cham, 2018), p. 84.
- [32] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [33] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd ed. (Routledge, New York, NY, 2012).
- [34] D. M. Bolt and X. Liao, On the positive correlation between DIF and difficulty: A new theory on the correlation as methodological artifact, *J. Educ. Meas.* **58**, 465 (2021).
- [35] S. Gökçe, G. Berberoğlu, C. S. Wells, and S. G. Sireci, Linguistic distance and translation differential item func-

- tioning on trends in international mathematics and science study mathematics assessment items, *J. Psychoeduc. Assess.* **39**, 728 (2021).
- [36] M. Wiser, C. L. Smith, J. Asbell-Clarke *et al.*, Developing and refining a learning progression for matter: The inquiry project (Grades 3–5), in *Proceedings of the American Educational Research Association (AERA) Annual Meeting* (San Diego, CA, 2009).
- [37] T. B. Corcoran, F. A. Mosher, and A. Rogat, *Learning Progressions in Science: An Evidence-Based Approach to Reform*. Consortium for Policy Research in Education (University of Pennsylvania, 2009).
- [38] E. M. Stump, M. Dew, G. Passante, and N. G. Holmes, Context affects student thinking about sources of uncertainty in classical and quantum mechanics, *Phys. Rev. Phys. Educ. Res.* **19**, 020157 (2023).
- [39] S. Zhou, J. Han, K. Koenig, A. Raplinger, Y. Pi, D. Li, H. Xiao, Z. Fu, and L. Bao, Assessment of scientific reasoning: The effects of task context, data, and design on student reasoning in control of variables, *Think. Skills Creativ.* **19**, 175 (2016).
- [40] M. Wilson, Measuring progressions: Assessment structures underlying a learning progression, *J. Res. Sci. Teach.* **46**, 716 (2009).
- [41] L. B. Hammer, J. C. Cullen, M. B. Neal, R. R. Sinclair, and M. V. Shafiro, The longitudinal effects of work-family conflict and positive spillover on depressive symptoms among dual-earner couples, *J. Occup. Health Psychol.* **10**, 138 (2005).
- [42] A. A. diSessa, Toward an epistemology of physics, *Cognit. Instr.* **10**, 105 (1993).
- [43] L.-M. Brault Foisy, P. Potvin, M. Riopel, and S. Masson, Is inhibition involved in overcoming a common physics misconception in mechanics?, *Trends Neurosci. Educ.* **4**, 26 (2015).
- [44] J. Odenbaugh, Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology, *Biol. Philos.* **20**, 231 (2005).
- [45] C. L. Smith, M. Wiser, C. W. Anderson, and J. Krajcik, FOCUS ARTICLE: Implications of research on children’s learning for standards and assessment: A proposed learning progression for matter and the atomic-molecular theory, *Meas. Interdiscip. Res. Perspect.* **4**, 1 (2006).
- [46] K. I. Erickson, C. H. Hillman, and A. F. Kramer, Physical activity, brain, and cognition, *Curr. Opin. Behav. Sci.* **4**, 27 (2015).