

Dynamics of collaborative modeling in an ill-structured real-world problemTom Reshef-Israeli^{*} and Shulamit Kapon[†]*Faculty of Education in Science and Technology, Technion Israel Institute of Technology, Haifa, Israel*

(Received 29 January 2026; accepted 13 April 2026; published 13 May 2026)

Collaborative mathematical modeling of real-world problems requires learners not only to engage in multiple epistemic practices but also to remain aligned around shared physical models as assumptions and solution paths evolve. However, few empirical studies have traced how such coordination unfolds moment by moment during authentic work. This study examined the *in vivo* discourse of a group of preservice physics teachers solving an ill-structured mechanics problem by tracking their collaborative reasoning turn by turn. We analyzed the group's work by coding their epistemic modeling phases and the operational physical models in use, alongside a time-resolved measure of discourse model coherence (DMC) that indexes alignment around shared model assumptions. By combining turn-level qualitative coding with time-resolved quantitative visualizations, we traced how the group's shared problem space evolved over time and identified moments of convergence and divergence. The analysis revealed that the group progressed through six epistemic phases of modeling while iterating across multiple operational physics models of increasing complexity. Periods of heightened complexity were marked by slower progression and frequent fluctuations between epistemic phases. Cross-model comparisons functioned as a key evaluative resource, in which simpler models served as benchmarks for judging more complex ones. Instructor prompts played a significant role: moves that narrowed the modeling scope tended to stabilize coherence, whereas prompts that reopened assumptions often increased epistemic uncertainty and temporarily disrupted alignment that produced measurable drops in DMC. These findings provide a time-resolved account of collaborative modeling dynamics and offer methodological and instructional insights for supporting modeling-rich physics learning.

DOI: [10.1103/44fz-2y3x](https://doi.org/10.1103/44fz-2y3x)**I. INTRODUCTION**

Over the last few decades, as modeling has been foregrounded as a key scientific practice [1–6], policy documents and research have called for integrating real-world problems into science curricula and engaging students in science and engineering practices in authentic, ill-structured contexts [7–12]. These activities can help reveal knowledge gaps and create opportunities for students to frame these gaps as questions they can address [13] and have also been shown to promote more inclusive participation across diverse learners [14,15].

Mathematical modeling is foregrounded in science and mathematics education as a central epistemic practice that students should learn to harness to make sense of complex phenomena. Mathematical modeling of real-world phenomena is particularly challenging for students since it

requires articulating assumptions, interpreting the real-world meaning of mathematical results, and evaluating their validity [4–6,8,9,16–18]. In physics, this involves the extensive coordination of mathematical representations with physical meaning. Specifically, equations and graphs do not only encode relationships in physics; they are objects of reasoning and justification that undergo constant revision as students blend mathematical reasoning and manipulation with the constraints of the physical situation they are modeling [19–23]. Real-world physics problems can be modeled and solved at various degrees of abstraction and approximation, depending on the context and the goal involved. Thus, engaging students in the problem solving of realistic real-world problems leverages learners' consideration of the applicability of different principle-based accounts such as energy conservation, Newtonian dynamics, or friction and drag models [24–26]. The demands are amplified when these problems are intentionally ill structured [27–29], where constraints and solution paths are underspecified. Such problems require students to iteratively revise or extend their models while keeping covarying quantities and changing conditions in mind [29–35].

Numerous multiphase conceptualizations of modeling have been put forward that describe how solvers transition

^{*}Contact author: tomis@campus.technion.ac.il[†]Contact author: skapon@technion.ac.il

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.

between activities such as understanding, mathematizing, interpreting, and evaluating [5,16,17,36]. In physics education, these accounts have mainly been developed in theoretical reflections, historical analyses, or expert-centered perspectives [1,4–6,37–41]. Only a few studies have traced learners' modeling processes *in situ*, turn by turn [32]. Some works in mathematics education have analyzed how participants move across modeling phases over time [42,43]. However, these studies often represent modeling trajectories at the group level rather than attending to individual contributions. In addition, they tend not to examine which physical models or principles the participants invoke, how these are revised, or whether group members remain aligned on the same model and assumptions as the interaction unfolds. This distinction is especially important in physics, where modeling relies on explicit assumptions and physical principles to account for mechanisms in the real world [44,45]. As a result, there is little empirical documentation of how learners coordinate epistemic modeling phases together with the selection and refinement of explicit physical models during authentic activity [46].

In classroom settings, modeling in problem solving is frequently collaborative. Collaboration can support complex problem solving by pooling complementary knowledge and sustaining engagement in challenging tasks. However, groups also often struggle to maintain shared understanding over time [47–50]. Research on collaborative learning provides powerful accounts of how groups build common ground and converge on shared goals and representations [47,51,52]. Many assessment approaches operationalize “quality of collaboration” by evaluating social-interactive indicators or functional discourse moves [53–56]. However, these approaches do not typically track whether collaborators remain aligned on the substantive *content* of their reasoning in physics and specifically whether they are operating within the same underlying physical model and assumptions as the work unfolds. This gap is consequential in modeling-rich tasks, where students may appear interactionally coordinated while reasoning with incompatible models that only yield partial convergence and result in stalled progress.

Finally, although scaffolding and instructional support have been widely studied in inquiry and problem-based environments [27,57–60], fewer studies have connected specific in-the-moment instructor prompts to time-resolved shifts in students' epistemic progression and content-level alignment during physics problem solving [32]. Understanding how such prompts stabilize, redirect, or reopen modeling work requires analytic approaches that can link instructor-group interactions to the evolving dynamics of collaborative reasoning.

The present study addresses these gaps by examining the authentic discourse of collaborative layered modeling by preservice physics teachers working on an ill-structured

real-world mechanics problem [26,33]. In what follows, we present the conceptual framework for the study (Sec. II). Drawing on research on mathematical modeling, collaborative problem solving, and instructional scaffolding, we define the constructs that guided the analysis: epistemic modeling phases, operational physical models, and *discourse model coherence* (DMC). We then use these constructs to articulate the research questions (Sec. III), operationalize them in the method section (Sec. IV), and use them to interpret the findings (Sec. V). In the discussion (Sec. VI), we relate the results to the literature, and conclude with implications for research and instruction (Sec. VII).

II. CONCEPTUAL FRAMEWORK

The conceptual framework of the study draws on three areas of scholarship: mathematical modeling, collaborative problem solving, and instructional scaffolding. In what follows, we review the literature and define the theoretical constructs that guided the analysis.

A. Mathematical modeling

The role of mathematics in physics has often been described as a representational system through which physical quantities, relationships, and constraints are expressed, negotiated, and justified [21,24,25,37,39,46,61–63]. From this perspective, although knowing the relevant mathematics is a prerequisite for physics inquiry and problem solving, it is not enough; physics students need to be cognizant of the fact that mathematical forms carry physical meaning and learn to use them to make sense of real-world phenomena. Examples include interpreting symbols as quantities with units, coordinating dimensional relationships, and using physical constraints to guide algebraic work [5,19,46,64,65].

Many physics educators argue that although mathematics and physics are acknowledged and taught as separate disciplines, productive physics reasoning and discovery emerge from their cognitive *blend*: mathematical operations are guided by physical sensemaking, and physical interpretations are refined through mathematical structure [19,20,22,23,62,66]. This is particularly crucial when students construct, interpret, or evaluate equations while engaged in mathematical modeling in physics, especially when the problem in question is complex or underspecified.

Physics education [4–6] and mathematics education [16,18,36,67] studies tend to use different terminology and emphasize different aspects of the modeling process, but there is a general consensus that solving a modeling problem typically requires learners to (i) identify the physical quantities and relationships involved; (ii) construct a mathematical representation of these relationships; (iii) process this representation to derive a solution; and (iv) interpret and evaluate the solution in context.

TABLE I. Comparison of three conceptualizations of the modeling process as a cyclic, multiphase process.

Blum and Leiß [16] (pp. 225–226)	Kaiser and Stender [17] (p. 278)	Redish and Kuo [5] (p. 568)
<i>Understanding</i> “The text has to be read and the problem situation has to be understood by the problem solver.”	<i>Understand and simplify</i> “The given real problem is simplified in order to build a real model of the situation, amongst other things many assumptions have to be made, and central influencing factors have to be identified.”	-
<i>Simplifying/structuring</i> “The situation has to be simplified, structured and made more precise, leading to a real model of the situation.”		- (Implicitly mentioned, without a distinct labeling) “We ((physicists)) identify measurables—variables and parameters that we can quantify.”
<i>Mathematizing</i> “Mathematization transforms this real model into a mathematical model.”	<i>Mathematize</i> “To create a mathematical model the real model has to be translated into mathematics.”	<i>Modeling</i> “Mapping these measurements into mathematical symbols.”
<i>Working mathematically</i> “Mathematical tools are activated.”	<i>Mathematical work</i> “Mathematical results are worked out by using mathematics.”	<i>Processing</i> “Process the math to solve problems.”
<i>Interpreting</i> “This result has to be interpreted in the real world.”	<i>Interpret</i> “Interpreting the mathematical results leads to real results.”	<i>Interpreting:</i> “What the result means back in the physical system.”
<i>Validating</i> “Validation of this result: Is it reasonable? Is the accuracy appropriate?”	<i>Validate</i> “((The results)) have to be validated within the real situation and/or within the real model.”	<i>Evaluating</i> “Evaluate whether the result supports our original choice of model when compared to observations.”

Our analysis was informed by three key conceptualizations of the modeling process [5,16,17], compared in detail in Table I. Although these conceptualizations share similarities, they differ in structure and emphasis, particularly in the first and last phases. Note that all three are based on theoretical reflections; by contrast, we studied the *in vivo* modeling process that implemented both top-down (informed by these conceptualizations) and bottom-up (informed by the content of the *in vivo* modeling discourse) analyses. Thus, the specific *epistemic phases of the modeling process* that were identified in this study reflect an amalgamated adaptation of the conceptualizations in Table I and our data. In the Method section, we provide definitions for each epistemic phase, its operationalization into discursive actions, and an illustrative example.

Although modeling rarely unfolds as a strictly linear sequence [2,6], the modeling phases in Table I can be seen as moving from “earlier” to “later” stages in a typical modeling sequence [5,16,17,36]. The analysis made use of this ordinal structure to characterize the group’s overall progression over the course of the modeling process.

Finally, many researchers in physics and mathematics education have described the epistemic process of mathematical

modeling as cyclic [2,5,6,16–18,26,33,36,42]. When learners engage with real-world, ill-structured problems with no single correct solution [27,28,68], learners can progressively add layers of complexity to their models by integrating new assumptions, variables, or presumed physical effects. Each iteration can build on previous ones, thus furthering insights and supporting comparisons of alternative models and solution paths for the same phenomenon.

B. Collaborative problem-solving

1. The dynamics of the joint problem space

Although the *epistemic* dimension of modeling has elicited extensive scholarly attention [3–6,16–18,21,25,37–39,42,43,63,67,69–71], the current study also highlights the *social* dimension of modeling that takes place when collaborating with others. Collaborative problem-solving is widely recognized as a key competence, but many learners struggle in this setting [50]. Collaboration can support complex problem solving by allowing group members to contribute complementary knowledge and skills [10,48,72] and helping them sustain engagement in challenging, ill-structured tasks [47,49,68].

At the foundational level, research in the learning sciences has characterized productive collaboration as requiring *convergence*: an iterative process of “displaying, confirming, and repairing” meaning such that participants come to share what they are doing and why [47]. A key object of this convergence is a shared conception of the problem and possible solution paths [47,49,51,73]. Roschelle and Teasley termed this shared conception the *joint problem space* (JPS): a dynamically constructed, jointly maintained structure that integrates goals, representations of the current problem state, and possible next actions [51].

More broadly, accounts of collaborative learning emphasize that groups manage convergence and divergence dynamically over time, by balancing coordination with adaptability as ideas evolve [74,75], often within a dialogic space where participants negotiate what counts as progress and justification [51,76,77]. This interactional work has also been described as “grounding” [52,73,78]. In complex, modeling-rich tasks, these shared conceptions are difficult to sustain [79]: students can appear responsive to one another while implicitly relying on different physical models or incompatible assumptions that can impede the group’s progress. In other words, the JPS in modeling activity is not only a shared topic or goal but also a model-based interpretive frame that must be maintained (and sometimes renegotiated) as assumptions and representations shift.

Despite the centrality of the JPS construct, relatively little work has operationalized it in a way that is time resolved; that is, as an evolving trajectory that can be traced moment by moment across extended collaborative work. This limitation is especially consequential for modeling-rich physics problem solving, where shifts in assumptions or representations can reconfigure the group’s problem space. Our approach responds to this gap by visualizing a modeling-relevant slice of the JPS over time, linking turn-level contributions to group-level trajectories of alignment and divergence.

2. Measures of the quality of collaboration

Much research on collaboration has operationalized “quality of collaboration” via social-interactional dimensions (e.g., sustaining mutual understanding, dialogue management, reaching consensus, task division, time management, and reciprocal interaction), often in computer-supported settings [53,54,56]. Related work has provided fine-grained accounts of collaborative discourse moves that support knowledge construction; for example, by coding question and statement types and their complexity in problem-based learning groups [55,60] or identifying epistemic moves by which students advance sensemaking in physics problem solving [80]. However, these approaches have typically emphasized either (i) the interactional-functional role of moves (what an utterance does) or (ii) general epistemic practices [81], without directly tracking whether the epistemic content of the discourse remains coordinated; in the current study, whether participants are operating within

the same physical model. Accordingly, we treated “epistemic” in two complementary senses: epistemic function (the phase in the modeling process) and epistemic content (the operational physical model in use). This distinction allowed us to examine not only whether an utterance advanced the activity (e.g., mathematizing and evaluating) but also what model-based assumptions made it meaningful.

To examine how coherently the group members coordinated their problem solving, we drew on Brigid Barron’s notion of *relatedness* as a measure of collaborative discourse [49]. Barron assessed the extent to which solution proposals introduced by group members were connected to the immediately preceding line of discussion. Specifically, she coded each proposal as *directly related* when it addressed the same subproblem or conceptual step that the group had been discussing in the preceding turns or as *not directly related* when it shifted to a different subproblem or was otherwise misaligned with the ongoing discourse. To compare groups, she calculated an index of group relatedness, which was defined as the proportion of directly related proposals out of all the proposals coded in each group. The rationale for this measure was that collaborative progress depends not only on the correctness of ideas (a result she established in the first part of her study) but also on whether group members introduce them at times when others can integrate them into a shared (predefined) problem space. Barron showed that the level of relatedness could distinguish productive from unproductive collaborative interactions. Here, we extended and adapted Barron’s notion of relatedness to capture the dynamics of collaborative modeling discourse.

Our students were working on an ill-structured, authentic physics problem that was intentionally not prepared into subproblems. During the layered modeling process, the group employed different operational physics models. The term *operational physics model* refers to the adapted canonical model that was temporarily generated for a particular context and temporal focus, while employing specific assumptions. We coded the operational model that was used or implicitly referred to in each discourse turn. Barron coded the relatedness of *new* idea proposals, whereas we were interested in the moment-to-moment alignment around an operational physical model, since we considered it to be an indicator of local discourse coherence because maintaining a shared model (and its underlying assumptions and principles) is a key coordination demand in collaborative modeling.

We suggest the term *discourse model coherence* (DMC) to capture these dynamics. DMC is the extent to which interlocutors are aligned around the same operational physical model at a given moment. In this study, we used DMC as a way to make content-level coordination visible over time; i.e., whether students’ contributions remained interpretable within a shared, model-based framing of the problem. This focus on alignment around a shared model is compatible with accounts of collaboration that emphasize

maintaining a *joint problem space* and a “shared understanding” of the task [47,51,52,73,78,79,82,83].

Although Barron used relatedness as a summative measure to compare between groups, we were interested in examining the dynamic process of collaborative modeling within groups. Hence, we examined DMC as a function of time. This enabled us to capture fluctuations in coherence across different stages of the activity and to examine how instructional interventions influenced these dynamics.

C. Instructional scaffolding of problem solving

Collaborative modeling in learning contexts can also be influenced by interactions with instructors. Scaffolding strategies have been examined in inquiry and problem-based environments [27,57–60], but less is known about the ways in which real-time instructional prompts influence the direction and coherence of group modeling in ill-structured physics problems. Studies have rarely empirically connected specific instructor moves to time-resolved shifts in (i) the group’s underlying model, (ii) their epistemic progression, and (iii) their moment-to-moment alignment around that model. By relating instructor-group interactions to changes in these three dimensions, we could characterize how such interventions reshaped the evolving JPS.

Instructor moves can also narrow the modeling space to stabilize progress or reopen assumptions to encourage model refinement [57]. Work on ill-structured problem solving has emphasized that learners may benefit from support at multiple levels, including cognitive, metacognitive, and epistemic cognition [84], and that instructors’ prompts can improve performance when they help students represent the problem, justify choices, and monitor progress [27].

Saye and Brush distinguished between hard scaffolds, which are anticipated, preplanned supports, and soft scaffolds; namely, contingent, in-the-moment supports that require ongoing diagnosis and responsiveness [58,59]. Studies of problem-based and inquiry learning have similarly argued for attending to both forms of scaffolding [60,85], and have illustrated how structured tools can stabilize group sensemaking, whereas facilitator questions can prompt monitoring and reflection. From this perspective, instructor interventions in collaborative modeling may be impactful because they function as soft scaffolds that can reshape the group’s JPS by redirecting attention, making warrants explicit, or pushing for justification at moments of uncertainty. Accordingly, we examined how instructors’ in-the-moment prompts related to shifts in the group’s model choice, phase movement, and DMC over time.

These prompts often take the form of temporarily structuring or problematizing aspects of students’ work in ways that support progress without eliminating their agency [27,58–60,86,87]. These scaffolds include question prompts, requests for clarification, or suggestions for alternative representations, and are considered to be particularly important when learners grapple with open, real-

world tasks that have no single correct solution [28,38,68]. In collaborative modeling, these interventions can influence the direction of the modeling trajectory (i.e., how to progress to or within an epistemic phase) and the coherence of the collaborative discourse around a shared model, by helping students articulate assumptions and align representations or by disrupting an emerging consensus.

III. RESEARCH QUESTIONS

In the context of a group of three preservice physics teachers working on an ill-structured mechanics problem, the following research questions were examined:

1. How were the epistemic phases of the modeling process realized, how did they unfold in the collaborative discourse, and how was this progression related to the complexity of the physical model applied by the group?
2. How did the *discourse model coherence* (DMC) of the group vary over time, and what did this reveal about the group’s collaborative reasoning?
3. How did interactions with the instructors shape the group’s collaborative modeling process?

IV. METHOD

A. Data

The data for this study were collected during a mandatory course for preservice physics teachers at the Technion, taught by the authors. We designed a workshop (180 min) on ways to engage high-school students in using the physics they learn in school to solve real-world ill-structured problems [27,28,68,88]. The first implementation of the workshop took place in March 2024. During the first 90 min, the students ($N = 15$) worked in small groups (2–3 participants) to collaboratively solve the problem. During the next 90 min, they discussed their solutions and pedagogical considerations concerning the implementation of this type of activity in high school. The problem the students worked on, and the scaffolds they received, are described in detail elsewhere [26]. The problem is presented in Fig. 1. The activity was explicitly designed to elicit collaborative sensemaking [89] and a variety of solution paths with varying complexity and accuracy [33]. To solve the problem, the students had to use what they had learned in mechanics, as well as search the Internet and/or consult with AI agents on information they had not learned or knew (drag force, vehicle dimensions, etc.).

The analysis below centers on the first 90 min of the workshop. After a short introduction, the students ($N = 16$) tried to collaboratively solve the problem for ~80 min in 6 groups of 2–3 students. Camcorders recorded the work of four groups, their computer screens, and the students’ written notes. Both authors facilitated the activity. In what follows, we refer to the facilitators by role rather than by name: the first author is denoted instructor 1 and the second author as

Group task

Below is a high-context physical problem that you need to solve as a group. You have 90 minutes, at the end of which you must submit a reasoned solution that details and explains your considerations. Note the following:

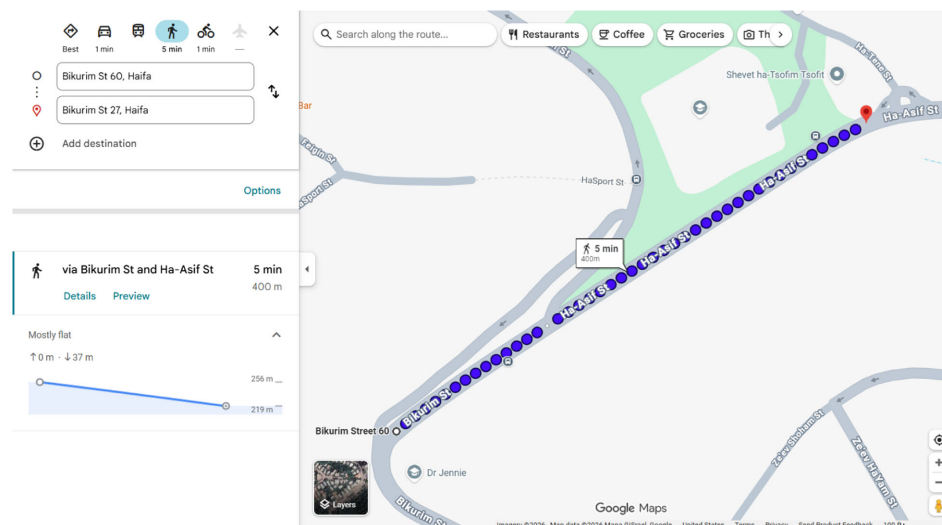
- Several groups will be asked to present and explain their solution path and considerations, and to answer questions from colleagues and teaching staff. Therefore, it is important to document your solution in an orderly manner. You can do so in any way you choose – you don't have to prepare a presentation.
- Each group will submit its documented solution to the submission box on the course website.
- You can use any external material you choose (and the teaching staff of the course). However, the solution you ultimately submit should be complete, with detailed physical considerations and calculations.

The problem

Nurit is a physicist who wants to become a presenter on the local version of the TV program “The MythBusters”. The task she undertakes is to investigate the difference between a bicycle and a car (in neutral), that is sliding downhill without the brakes on. For this experiment she chooses a straight section of the road that connects point A and point B (see map) and begins to slide downhill from rest.

Estimate Nurit's speed at the bottom of the slope when she goes downhill on a bike, and when she goes downhill in a car. Fully explain your considerations. You can use any auxiliary materials you choose.

If you have time left, discuss in your group: What is the importance of these assessments? Do they have meaning in the real world?



A [Link](#) to the directions in Google Maps.

FIG. 1. The group task (originally appeared in Ref. [89]).

instructor 2. Given that this was the first implementation of the activity, and because we were also collecting data for the study, two members of the physics education research group also provided their assistance: Tal Peer circulated with the authors between groups, and Asnat Zohar coordinated the cameras. Tal Peer is hereafter denoted as instructor 3.

We present an analysis of the work of a group of three students (pseudonyms): Rami—an experienced industrial engineer who was considering a second career in teaching, Olga—an immigrant with a Ph.D. in mathematics who was studying toward a teaching certificate, and Eran—a fourth-year undergraduate in electrical engineering who

was taking the course as an elective. This group was chosen for in-depth analysis because it included more than two students, participation was relatively balanced, and the students had complementary disciplinary expertise. They also demonstrated sustained, visible engagement with the task, thus making them well suited for tracing collaborative epistemic engagement during the modeling process.

B. Analysis

The group's full 75-min interaction was transcribed in Hebrew by the first author. Each utterance was treated as a

discursive turn and assigned a sequential turn number in chronological order. All turns were time stamped and supplemented with annotations on gestures, intonation, and prosody, as well as photographs and screen-captures of the students' written notes and online work. In cases where multiple participants took part in a single action (e.g., shared reading or simultaneous tool use), the turn was duplicated and attributed separately to each participant. The analysis was conducted on the original Hebrew transcript. Selected excerpts were translated into English for illustration.

1. Coding operational models and epistemic phases

The specific operational models referenced in each turn (e.g., Newton's second law in inclined plane, with drag force and terminal velocity) were identified according to the following criteria: (a) an *explicit verbal reference* to core model entities or assumptions, such as gravity, drag, friction, or conservation of energy (e.g., "I think it's the air resistance," or "The drag coefficient is 0.3"); (b) an *implicit verbal reference* to the model, inferred from the immediate discourse context (e.g., "I got v squared," said moments after "This is the kinetic energy," suggesting the use of mechanical energy principles); (c) *nonverbal activity* related to a specific model, such as searching online for the drag equation, drawing force diagrams with or without drag, or writing equations that reflected the conservation of mechanical energy. Some turns referenced more than one model and thus were attributed more than one model code. Utterances unrelated to any model, such as off-topic remarks or generic data searches, were labeled as "no specific model." Of the turns, 20% were independently coded for the model code by two members of the physics education research group, yielding 93% agreement. Disagreements were discussed and resolved, leading to refinements in the model coding guidelines.

The coding scheme of the *epistemic phases* of the modeling cycle (second code) was developed through both top-down analysis informed by central conceptualizations of the modeling cycle (Table I; [5,16,17]) and

bottom-up inductive refinements based on the documented authentic modeling discourse using iterative comparative methods [90]. When the definitions of the epistemic phases had stabilized, 20% of the data were independently coded by two members of our research group, yielding 72% agreement. Disagreements informed further refinement of the coding scheme, after which the full dataset was recoded by the first author according to the revised scheme. Table II presents the definitions of the epistemic phases and their operationalization into indicative actions that could be identified in the students' discourse.

2. Quantifying the epistemic progression of the modeling process

The entire transcript was coded turn by turn for the operational model (e.g., Newton's second law in an inclined plane, with drag force and terminal velocity) and the epistemic phase that were referenced (see Table II) in each turn. To follow the temporal progression, the beginning timestamp of each turn was tabulated. A Python code was developed to plot the epistemic phase against time, in a representation that also presented the operational model that was referenced in each turn and the speaker. The plot provides a holistic representation of the detailed progression of the collaborative modeling process as manifested in the JPS.

In addition, a linear regression analysis was applied to the time segments of the epistemic phases. Specifically, we treated the epistemic phase as a discrete, ordinal variable plotted against time, and then fitted a linear trendline to model its overall trajectory. The *slope* of the regression line was interpreted as the group's coarse rate of progression throughout the epistemic phases of the modeling cycle. The coefficient of determination ($0 \leq R^2 \leq 1$) was used to quantify the extent to which the phase progression was well described by a linear trend (i.e., its degree of linearity). Higher R^2 values indicated a more nearly linear, steady progression across phases, whereas lower R^2 values indicated greater deviation from a linear progression.

TABLE II. Coding the epistemic phases of the modeling process.

Label	Definition	Indicative actions	Examples
Understanding	Forming an initial idea of the target phenomenon and/or problem definition and recognizing the problem setup.	- Looking at or citing the problem definition. - Referring to what was read. - Rephrasing the problem in their own words. - Taking stock of the data and information given. - Relating the problem (or certain elements in it) to their personal experience.	"Wait, but what do we actually know, what data do we have?" "Let me read: 'Provide your best estimate ...' ((mumbles))." "The length is 400 m and ... the drop is 38." "I still don't get it—why should there be a difference between a car and a bike?"

(Table continued)

TABLE II. (*Continued*)

Label	Definition	Indicative actions	Examples
Identifying	Identifying critical components of the problem and their relations, to create an idealized version of the problem.	• Making assumptions. • Specifying conditions following assumptions. • Listing variables, parameters, and constants. • Estimating/collecting additional data. • Drawing sketches corresponding to the problem (including, e.g., free body diagrams). • Referring to physical entities not explicitly mentioned in the problem statement. • Describing actions, properties, or relations between the physical entities. • Describing a chain of events.	“Presumably there’s no friction since they’re on wheels.” “The question is—what are the forces here?” “In air resistance, the mass matters.” “The car’s mass and force are both much bigger.”
Mathematizing	Representing the model mathematically using symbols, equations, graphs, tables, etc.	• Writing mathematical representations of ideas. • Speaking in terms of symbols, operations, or relationships. • Searching for existing equations that relate variables.	“I’m writing formulas.” “Potential energy is mass times height, right?” “What’s the formula for air resistance?” “Let’s derive an expression for v and plug in carefully.” “We’ve got force ... $mg \sin \theta$.”
Processing	Employing mathematical manipulations of mathematical representations to obtain a mathematical result.	• Using/speaking about algebraic manipulations or calculus-based operations. • Making inferences and deductions <i>without</i> reference to nonmathematical knowledge. • Applying various mathematical operations (e.g., substituting, rounding, and partitioning).	“To get velocity, we need to plug in the time.” “I calculated the slope angle.” “We need to take the square root.” “I want to calculate. ((pulls out calculator))”
Interpreting	Explaining the physical meaning of a mathematical expression or result.	• Referencing the physical meaning of a symbol rather than its mathematical representation. • Considering the units of a mathematical expression or variable. • Unit conversion. • Referring to the expression in the physical context of the question. • Considering whether the obtained result answers the question posed.	“That’s the speed in meters per second.” “Let’s convert it to kilometers per hour.” “Is that the square of the velocity?” “The car would be going 100 km/h at the bottom of the slope.”
Evaluating	Examining the reasonableness of a solution, intermediate expression, or result.	• Comparing an obtained result to personal knowledge/intuition. • Comparing the result to a previous simplified model results. • Checking extreme/special cases/limits. • Discussing the implications of the idealization of the phenomena. • Carrying out dimensional analyses to verify a formula/expression.	“We should verify that all these parameters match in terms of units.” “Does this speed seem reasonable?” “I did this on a skateboard once—I didn’t hit 100.” “It doesn’t make sense!” “What was the speed we got earlier without drag?” “I think this just means the car doesn’t get to its max speed in the given time.”

3. Measuring the DMC over time

We examined the transcript turn by turn and, inspired by Barron's work [49], differentiated between two types of adjacent turn pairs (i.e., turns 1 and 2, 2 and 3, 3 and 4, etc.):

- *Related pairs*: adjacent turn pairs in which both speakers referred (in speech or action) to the same operational model.
- *Unrelated pairs*: adjacent turn pairs in which the speakers referenced different models, or both turns were coded as “No specific model.”

“Same operational model” was determined from the model code assigned to each turn (including implicit references and model-relevant actions). When turns had multiple model codes, a pair was treated as related if the two turns shared at least one operational model code. Pairs in which both turns were coded as “No specific model” were treated as unrelated because they did not reflect alignment around any enacted operational model.

We then divided the transcript into time intervals of (i) independent group work, in which the group worked without interacting with an instructor, and (ii) instructor-group interaction, in which the group discussed their work with an instructor.

We assessed the *discourse model coherence* (DMC) only in the independent time intervals. The DMC of a time interval was quantified as the ratio of the number of related pairs in this interval to the total number of pairs ($0 \leq \text{DMC} \leq 1$).

This coding strategy was implemented in a python code that allowed us to generate a dynamic, time-resolved profile of model coherence across the whole session (i.e., the DMC vs time). Peaks in coherence indicated moments of tightly

coordinated work around a shared model, whereas dips suggested fragmentation due to competing ideas, miscommunication, or a shift toward model-detached brainstorming. In this analysis, these peaks and dips were used as descriptive indicators to support the interpretation of the group's trajectory and to locate episodes for a fine-grained qualitative analysis, rather than as stand-alone measures of the quality of collaboration.

4. Evaluating instructors' interventions effect

The instructors' interventions varied in length and content but were generally brief and took the form of scaffolded questioning or redirection. We compared the DMC before and after each time interval in which an instructor interacted with the group. Because the DMC was defined as a proportion, we treated it as a binomial and used two-tailed Z tests for differences in proportions to test whether the pre-post changes were statistically significant. To ensure Z test validity, we confirmed that all expected values ($n \cdot p$, with $p = 0.5$) exceeded 5 [91]. Instructor interventions that resulted in a statistically significant change in the DMC (either increasing or decreasing the coherence) were further examined qualitatively to determine how different forms of scaffolding supported or disrupted coherence.

V. FINDINGS

A. Epistemic progression

The group's epistemic progression was traced turn by turn. Each turn was coded for the model operationally used in the turn and the modeling phase it reflected. Five distinct models were identified in the group discussion. Table III presents the underlying assumptions and

TABLE III. Assumptions and principles underlying each model discussed in the group.

Model	Assumptions	Principles
A	<i>Influencing entities:</i> • Gravity.	• Newton's laws. • Kinematics in constant acceleration.
B	<i>Neglections:</i> • Friction with the road. • Air drag. • Objects' dimensions. • Wheel rotation.	• Mechanical energy conservation.
C	<i>Influencing entities:</i> • Gravity. • Air drag. <i>Neglections:</i> • Kinetic friction with the road.	• Newton's laws. • Drag equation (fluid dynamics). <i>Special case:</i> • Motion in terminal velocity ($a = 0$).
D	• Wheel rotation.	• Newton's laws. • Drag equation (fluid dynamics). <i>General case (varying velocity)</i>
E	<i>Influencing entities:</i> • Wheel rotation.	(was not developed in physics terms)

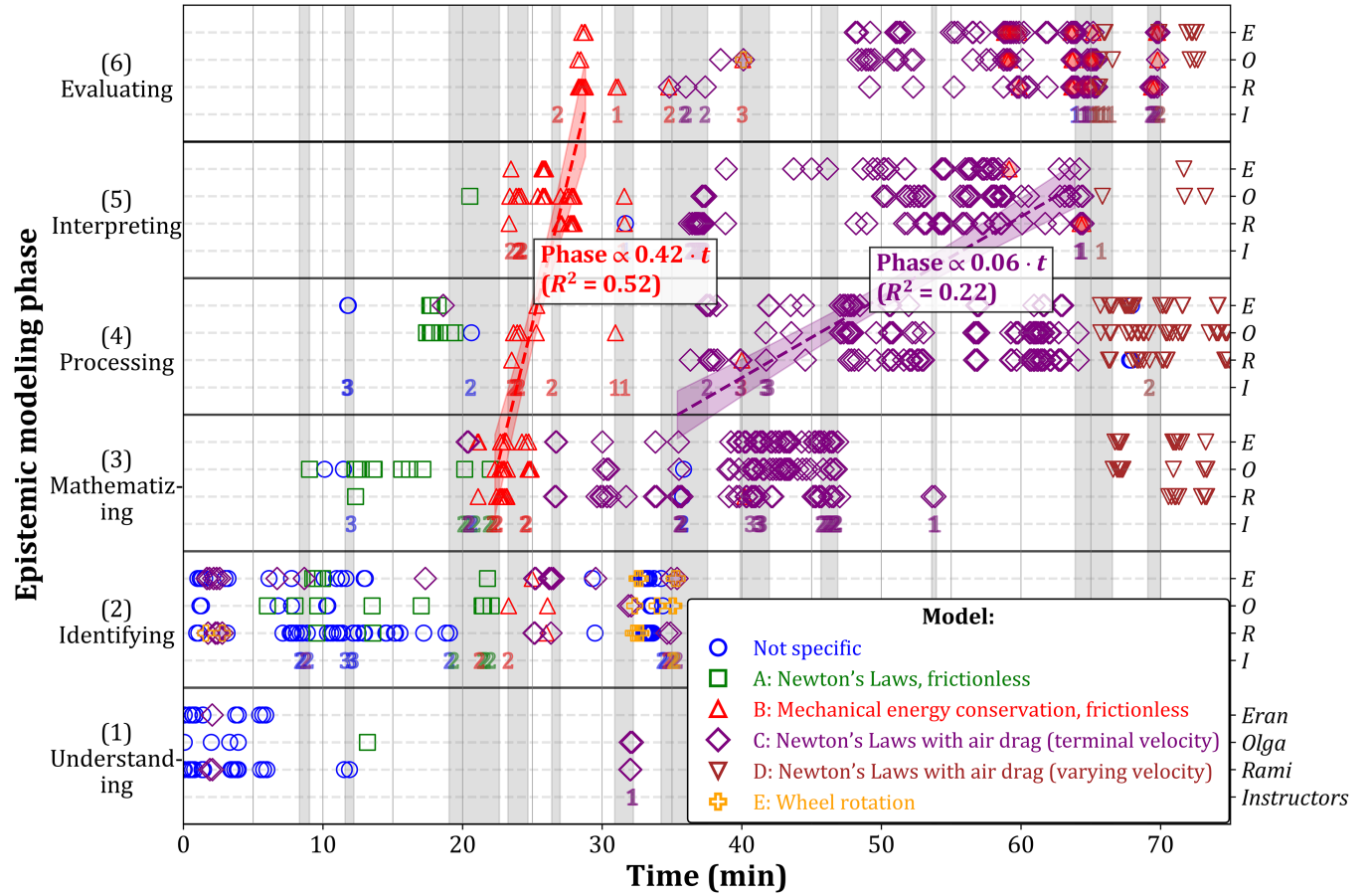


FIG. 2. Group progression across epistemic modeling phases (Table II) over time. models A–E are defined in Table III. Regression lines show the phase progression for models B and C, with shaded confidence intervals.

principles of each model. The operational definitions of the six modeling phases (understanding, identifying, mathematizing, processing, interpreting, and evaluating) appear in Table II. Within the timeframe of the activity, the group did not have time to develop and explore model E.

1. Macrolevel: Linear epistemic progression

Figure 2 visualizes the group's epistemic progression. Each point in the figure represents a turn. The horizontal axis represents the time of the turn (in minutes) that had elapsed from the beginning of the group work. The vertical axis represents the modeling phase manifested by the turn. Students' turns are indicated by colored shapes, and individual students are shown on different horizontal lines (Eran—top, Olga—middle, Rami—bottom). The instructors' turns are marked by numbers: 1—instructor 1 (first author), 2—instructor 2 (second author), and 3—instructor 3. Students' turns are indicated by different colors and shapes to designate the model referenced in that turn (models A–E, Table III). For example, blue circles represent turns that did not reference any specific model. Striped gray areas indicate time periods in which one of the instructors interacted with the group.

Figure 2 reveals two clear modeling cycles in which the group discussions progressed sequentially along the modeling phases, from *mathematizing* (phase 3) to *evaluating* (phase 6). The linear regression trends for each cycle are overlaid, with the slope and coefficient of determination (R^2) values and confidence intervals depicted as semitransparent bands. We defined the start of the cycle as the point in time when at least two of the three participants converged on mathematizing an operational physical model (e.g., adopting a potential-energy approach or the drag equation). The cycle ended when they agreed that work on the chosen physical model had ended (for example, see Excerpt 2 below).

The first trend (red) corresponded to model B, in which the group employed the principle of conservation of mechanical energy while neglecting friction and air drag. The second trend (purple) represented model C, which employed Newton's second law of motion by taking both gravity and air drag as influencing forces into account, but only considering the case of terminal velocity. This equilibrium case simplified the mathematical solution since it eliminated the need to solve a differential equation.

The slope of model C (purple) was noticeably less steep than the red one, reflecting a slower progression—approximately one modeling phase every 16.7 min in the

drag-inclusive model, compared to a modeling phase every 2.4 min in the frictionless model B (red trend). This difference can be attributed to the higher complexity of model C compared to model B. Model C included the air drag force that depends on the object's velocity ($F \propto v^2$), resulting in a differential equation which the students first analyzed qualitatively and then realized they could use the terminal velocity to solve a simpler equation in this case.

2. Narrating pathways of epistemic progression

As shown in Fig. 2, during the first 10 min of the discussion, the group engaged primarily in the first two epistemic phases of the modeling cycle: *understanding* and *identifying*. During this time, the students explored the problem setting by brainstorming physical entities (gravity, air friction, and the number of wheels) that might be relevant to determining the velocity of each vehicle at the bottom of the slope. The dominant color of the dots during this time was blue, indicating that no specific model was referenced in the vast majority of the turns during this time. Excerpt 1, drawn from the first 2 min, illustrates the nature of this discussion. The time stamps on the left can be used to follow the points in Fig. 2.

Excerpt 1

0:00:30 1. Rami: Do we need to look for the height difference on the web?
 (...)
 0:00:51 2. Eran: The length is 400 m and ... the drop is 38 m.
 0:00:57 3. Rami: Really? That's it? 38 m? Okay, that's quite a bit.
 0:01:02 4. Eran: It's pretty steep.
 0:01:06 5. Rami: Yeah.

As shown in Excerpt 1, Rami and Eran spent the first minutes of the session finding the relevant information (Rami, turn 1), extracting it from the Google map (Eran, turn 2), and making sense of it (turns 3–5).

Toward the end of the first 10 min, green points start to appear in Fig. 2. These points persisted until shortly after 20 min into the group work. They represent Olga's attempts to explicitly consider model A, which she derived for the car using Newton's second law in the case where the influencing force on the (constant) acceleration is gravity (see Table III). She was the leader in terms of the *mathematization* and *processing* of the model, and only occasionally consulted with Rami and Eran. Figure 2 also depicts occasional purple dots. Eran was working in parallel. He thought that air drag had to be an influential entity (i.e., he was in the *identifying* phases of the modeling cycle) and searched the Internet for a mathematical model to describe it. Hence, he began working on model C (purple) and engaged in both *identifying* and *mathematizing* (phases 2 and 3). Rami was busy searching the Internet for other data, such as the

weight and dimensions of a typical car or bicycle without referencing a particular model (blue, phase 2). Hence, during this time, the collective discussion did not reflect any convergence in terms of the models used.

In minutes 23–29, after an instructional prompt (see below), the group *collectively* began working on model B—a frictionless descent using conservation of mechanical energy. Model B enabled quick mathematization and processing, leading to a first estimate of the vehicles' final speed $v = \sqrt{2gh} \approx 27$ m/s for a car and a bicycle (independent of the vehicle's mass). During this segment (minutes 23–29), the group advanced through the epistemic phases of the modeling cycle relatively efficiently: it took them only 7 min to go from *mathematizing* (phase 3) to *evaluating* (phase 6). Excerpt 2 illustrates part of the evaluating phase toward the end of this cycle.

Excerpt 2

0:26:59 1. Rami: Let's convert it ((the velocity magnitude)) to kilometers per hour.
 (...)
 0:27:27 2. Olga: This way, right? Oh, it's the other way around.
 0:27:35 3. Rami: The other way around. It's a hundred kilometers an hour, 98.24.
 (...)
 0:28:15 4. Olga: It doesn't make sense, I think. Does it?
 0:28:18 5. Rami: One hundred kilometers an hour? Why, it's not that much. 400 m to get to 100 km an hour it could be, couldn't it?
 0:28:28 6. Olga: I don't know(?)
 0:28:30 7. Rami: ((To Eran)) What do you say, a hundred kilometers an hour, 98, do you go along?
 0:28:32 8. Eran: Umm ...
 0:28:38 9. Rami: Let's say, when I did it on my skateboard, I don't recall doing 98 km an hour, but —
 0:28:44 10. Eran: But you said ((before)) you didn't get to the bottom.
 0:28:46 11. Rami: Right, right. ((Laughs.))

Converting the velocity units from m/s to km/h (turns 1–3) was a way of *interpreting* (epistemic phase 5) the meaning of the mathematical result in a real-life context, where velocity is often measured in km/h. These efforts allowed the group members to *evaluate* the reasonableness of the mathematical solution based on their real-life experiences (phase 6) in the remaining turns. Turns 4–11 illustrate a rich sensemaking discussion, in which the students drew heavily on their prior experiences to make sense of their mathematical solution. This part of the discussion was highly responsive. For example, when Rami (turn 9) recalled descents on his skateboard that achieved much lower speed, Eran (turn 10) reminded him of the bounds on that experience (“you didn't get to the bottom”), which Rami jovially accepted (turn 11).

In minutes 30–31, the group members worked on constructing a mathematical model (i.e., *mathematizing*, phase 3) that incorporated the air drag (purple, model C). Then, after another instructor’s intervention, they considered other influences to get a better estimation of the speeds the vehicles would attain at the bottom of the slope. Rami wondered if there was a torque involved, and Olga mentioned the difference in the type and the number of wheels on a car compared to a bicycle (yellow points, minutes 32–35, indicating model E). At the same time, Eran suggested that they should test the problem physically in both a car and a bicycle on-site to “see what the actual speed would be” (0:33:12, blue, no specific model).

This discussion was interrupted by another instructor’s intervention in minute 35 (gray bar, Fig. 2). After this intervention, the group refocused on the drag-inclusive model (purple, model C). Initially, in the *mathematizing* phase, based on Newton’s second law, they derived a (differential) equation of motion: $ma = mg \sin \theta - kv^2$, where $k = \frac{1}{2} C_d A \rho$. Here C_d is the drag coefficient (based on the object’s shape), A is the cross-section of the object, and ρ is the density of the air. At this point, however, Rami pointed out that the equation depended on the vehicles’ velocity, making it more difficult to solve. Guided by instructor 3 (Tal Peer), the group realized that under these conditions, the vehicles might reach a terminal velocity when the acceleration reached zero, $a = 0$. By solving the equation in the simplified case of no acceleration, they found that the terminal velocity depended on the mass of the vehicle: $v = \sqrt{2mg \sin \theta / C_d A \rho}$. This move reflected a path from a general-case mathematical model (varying velocity and acceleration) to a *reduced* mathematical model (constant velocity) through physically informed simplifications, thereby generating a result through mathematical processing.

3. Microlevel: Fluctuating epistemic progression

While an overview of Fig. 2 showed an almost linear progression over the course of the epistemic phases of the modeling cycle as predicted by conceptualizations of the modeling processes, as predicted by conceptualizations of the modeling processes, a high-time-resolution view of Fig. 2 complicates these conceptualizations by revealing frequent oscillations between epistemic phases. This is demonstrated by Excerpt 3, which took place after finding the parametric expression for the terminal velocity of a vehicle (model C).

Excerpt 3

-
- 0:48:22 1. Rami: What did we say the air density was?
 0:48:28 2. Eran: Let’s see. In kilograms per cubic meter, it’s 1.2.
 0:48:35 3. Olga: What are the units for these coefficients? I want to check the units.
-

All three students were working on the same parametric expression, but they were not engaged in the same epistemic phase. Rami (turn 1) asked for a numerical value to substitute into a parameter in the equation and was thus engaged in *processing* (phase 4). However, Eran’s response in turn 2 not only provided a mere numerical value of 1.2 (phase 4) but also specified the measurement units (“kilograms per cubic meter”). Hence, he *interpreted* (phase 5) the meaning of this parameter in real terms. Rami’s question drew Olga’s attention to the parametric equation. In turn 3, she wanted to confirm that the units in the equation were consistent; hence, at the epistemic level, she engaged in *evaluating* (phase 6).

The calculation of the actual value of the terminal velocity for both the car and the bicycle took about 30 min, during which there were frequent shifts between epistemic phases. These fluctuations, which are visible throughout Fig. 2, underscore the iterative, non-linear nature of authentic modeling work and reflect the *blended* use of mathematics- and physics-based reasoning in modeling phenomena [19,20,22,23,66]. Notably, the lower coefficient of determination for model C ($R^2 = 0.22$) compared to model B ($R^2 = 0.52$) indicates that the phase progression in the more complex model deviated more from a linear trend, consistent with the more variable, back-and-forth phase fluctuations observed in the time-resolved plot.

4. Tracing the process of evaluation

Excerpt 4 started after the group calculated the terminal velocity of the car as 51 m/s according to model C.

Excerpt 4

-
- 0:58:45 1. Rami: That (51 m/s) doesn’t make sense!
 0:58:46 2. Olga: That doesn’t make sense!! ((pointing at the result))
 (...)
 0:58:54 3. Eran: Wait a minute, what was the result without the ... before, in the regular calculation? How much was it?
 (...)
 0:59:03 4. Olga: 97.2.
 0:59:09 5. Eran: That’s in kilometers per hour. How much was it in meters per second? ((moves his finger over Olga’s calculation sheet and looks for the result, then replies to himself)) 27.
 0:59:13 6. Olga: 27.
-

Rami (turn 1) and Olga (turn 2) immediately sensed that a velocity of 51 m/s was too high. Eran (turn 3) suggested a way to evaluate the plausibility of their solution by comparing it to the upper-limit velocity they calculated in the frictionless case (model B). Olga provided the value (turn 4), but not in the same units, which was not

comparable. Eran (turn 5) glanced through her documented calculations to find the value in meters per second, which she then confirmed (turn 6).

The students' calculation for the car's terminal velocity was, in fact, correct. What they overlooked at this point was that the car might not reach its terminal velocity within the finite length of the road [33]. The subsequent cross-model comparison illustrated in Excerpt 4 (seen in Fig. 2 by the overlay of purple diamonds and red triangles denoting models B and C simultaneously), led the group to the conclusion that the value of the velocity that they calculated was unreasonable since it violated the conservation of energy. However, at this stage, they were unable to identify the problem in their calculations.

When they did not find any mistakes in their calculation for the terminal velocity of the car, they decided to try to calculate the terminal velocity for the bicycle. Excerpt 5 started when Olga, who was in charge of the calculations, presented the result.

Excerpt 5

- 1:03:08 1. Olga: 15.87 m a second ... that's 54 km an hour.
(...)
1:03:35 2. Eran: Okay. And that's LESS than the result we
found in the ...
1:03:38 3. Rami: Much less!
(...)
1:03:42 4. Olga: Yeah, yeah, yeah ((smiles and nods.))
(...)
1:03:44 5. Eran: So it'll ((the bicycle)) get to that speed!
(...)
1:03:48 6. Olga: Wow, wow, wow! ((Smiles and leans
back.))
(...)
1:03:51 7. Rami: We did it. They should give us the Nobel
Prize!

As turns 2–4 show, Eran, Rami, and Olga immediately noticed that the bicycle's terminal velocity (15.87 m/s, according to model C) was less than the upper limit velocity when mechanical energy was conserved (~ 27 m/s according to model B). Eran (turn 5) articulated the meaning of the result, namely, that the bicycle would reach its terminal velocity but the car would not. Olga and Rami's enthusiastic expressions of satisfaction in turns 6–7 reflected their sense of achievement: they felt that they had reached an insight they trusted and could explain and were proud of having done so.

In the final stage (after minute 64), after instructor 1's intervention, the group attempted to return to the full differential formulation of the drag-inclusive model (model D): $ma = mg \sin \theta - kv^2$. This model implementation is indicated in brown in Fig. 2. Olga began formulating an analytical solution with Eran's help, identifying $a = dv/dt$ and drafting a differential equation

with its general solution, but the lesson ended before they could fully solve it.

B. Dynamics of the group's coherence in the modeling process

Figure 3 displays the dynamics of the proportion of related pairs, as measured by the *discourse model coherence* (DMC, see Sec. IV), over time. The height of each blue bar in the graph represents the value of the DMC, and its width indicates the time interval over which it was calculated. Overall, a general trend of increasing coherence emerged, from 0.17 at the beginning of the session to 1.00 in later segments. Periods of interactions with the instructors are shaded gray and annotated by instructor labels (I_1 —first author, I_2 —second author, I_3 —Tal Peer).

Figure 3 also reveals that the DMC changed after the instructors' interventions. Two-tailed Z tests showed five significant changes in the DMC: (i) after instructor 2's second visit ($t = 22.6$ min), the DMC *rose* sharply from 0.44 to 1.00 ($Z = 3.50, p < 0.001$); (ii) after Instructor 1's first visit ($t = 30.9$ min), the DMC *dropped* from 0.91 to 0.29 ($Z = -5.14, p < 0.001$); (iii) after Instructor 2's fifth visit ($t = 34.2$ min), the DMC *rose* again from 0.29 to 1.00 ($Z = 4.93, p < 0.001$); (iv) after instructor 1's third visit ($t = 63.9$ min), the coherence *dropped* moderately from 0.99 to 0.82 ($Z = -4.91, p < 0.001$); and (v) instructor 2's final brief interaction ($t = 69.1$ min) coincided with a comparable rise back to 1.00 ($Z = 3.36, p < 0.001$).

These oscillations, as annotated in the figure, show that the instructor interventions had immediate and measurable productive or disruptive effects on the group's DMC. In the following, we explore these interventions qualitatively in terms of the interactional features that shaped their outcomes.

C. How the instructors' moves shifted coherence

As shown in the gray striped areas in Figs. 2 and 3, the instructors interacted with the group frequently and spent a total of 19 min with them during the 75-min session. These visits affected the group's discourse and the *discourse model coherence* (DMC) in particular. To unpack the dynamic shifts in this coherence, we *qualitatively* examined five selected instructor interventions, each of which was statistically associated with a significant *quantitative* change in the group's DMC. Two of these, namely, instructor 1's first and instructor 2's fifth interventions are analyzed in depth, as they occurred in sequence and had contrasting, pronounced effects on the DMC. The remaining three are summarized more briefly to contextualize broader patterns of facilitative impact.

Immediately before instructor 1's first visit (27.0–30.9 min), the group had just finished evaluating the results of model B (mechanical energy conservation), and Rami and Eran were discussing how to incorporate

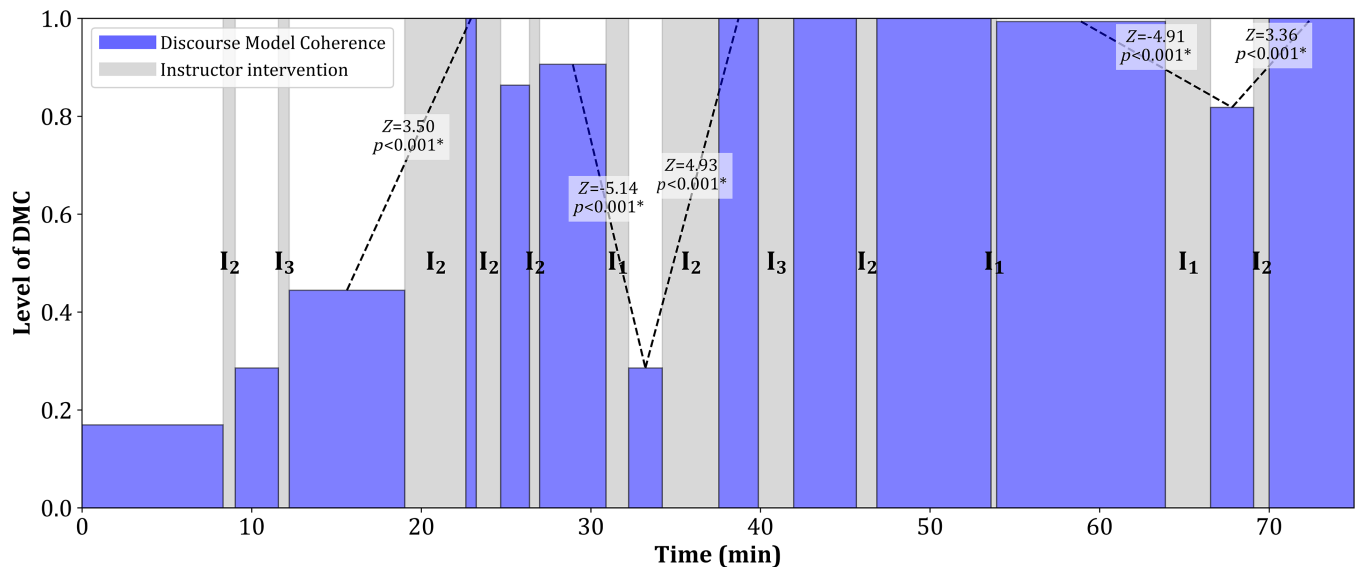


FIG. 3. The level of the group’s discourse model coherence (DMC) over time (blue bars). Statistically significant changes in coherence levels (based on two-tailed Z tests) are annotated. Instructors’ interventions appear in light gray, with labels indicating the instructor (I₁, I₂, I₃).

the air drag into their calculations (model C). Excerpt 6 began with Rami’s statement about finding a formula for the air drag by consulting with an AI chatbot (Microsoft Copilot).

Excerpt 6

- 0:29:38 1. Rami: I wrote in the chat: “What’s the formula for air resistance.” Let’s see what it gives me. (...) Aerodynamic drag of a car ... Oh! There you go! ((Points to his screen))
- 0:30:01 2. Eran: Okay. ((Looks into Rami’s screen, smiles))
- 0:30:03 3. Rami: Olga, here’s a formula. ((Points to his screen and tilts it toward Olga))
- 0:30:06 4. Olga: What’s that? ((Looks into Rami’s screen)) Oh, right! I found it too.
- 0:30:21 5. Rami: Oh, really? You don’t say. ((Looks into Olga’s tablet))
- 0:30:28 6. Olga: The formula. Okay, now I’ll—
- 0:30:35 7. Rami: So, we need ... The air drag force is, okay ... the drag coefficient depends on the car’s shape, okay, we’ll find that. The cross-section we already know. Air density—whatever, we’ll find it. Okay.

As can be seen in Excerpt 6, the group worked jointly on constructing a model that incorporated air drag. This was reflected in the relatively high DMC (0.91) in this time segment. At this point, toward the end of minute 31, instructor 1 visited the group (30.9–32.2 min). At first, they talked about the previous calculation, in which they obtained the vehicles’ speed while neglecting the vehicles’ dimensions and external frictional forces. Then, Rami

began explaining their current efforts that required car and bicycle specifications (Excerpt 7).

Excerpt 7

- 0:31:40 1. Rami: We have the dimensions and weights of a car and a bicycle.
- 0:31:41 2. I₁: And what are you going to do with them?
- 0:31:43 3. Rami: Now I am looking at air drag, maybe we’ll delve into that. (...) We found the formula.
- 0:31:53 4. Olga: There is also friction here, it depends on the weight.
- 0:31:59 5. Rami: Now we need to find out what differences there are.
- 0:32:01 6. Olga: Yeah.
- 0:32:03 7. Rami: We’re not sure what ...
- 0:32:05 8. Olga: If it’s not due to friction, what would be the difference?
- 0:32:07 9. I₁: That’s exactly the question. Maybe there isn’t any difference.
- 0:32:10 10. Olga: ((Giggles.))

Instructor 1’s reaction in turn 2 prompted the group to engage in deeper reasoning (“what are you going to do with them?”). Rami’s response in turn 3 elaborated on their intention to use the air drag force formula to calculate the velocity. However, Olga’s statement in turn 4 showed that the group was not entirely aligned on the same model: whereas Rami was focused on air drag, Olga was also thinking about (surface) friction. Then, in turns 5–8, both Rami and Olga framed the group’s problem as being able to understand what caused the difference between the car’s and the bicycle’s velocity. In turn 7, Rami explicitly expressed the group’s confusion (“We’re not sure what ...”). In response, in turn 9,

instructor 1 reinforced the group's framing ("That's exactly the question."). However, instead of guiding them on how to address this problem, he implied that they should work it out themselves, which was further amplified by his leaving the group after this turn. The statement "maybe there isn't any difference" added to the confusion. Excerpt 8 begins immediately after the end of Excerpt 7.

Excerpt 8

0:32:14 1. Rami: There's ... I don't know, yeah, friction, I don't know. Is there any torque involved here?"

(...)

0:32:35 2. Eran: Maybe kinetic friction, but ...

0:32:42 3. Rami: Once you're rolling, you're rolling.

(...)

0:32:50 4. Eran: The question is if we have any way to evaluate the magnitude of the ... Can we go there? 90 min should be enough time to drive out there and check.

((Giggles))

((Group members keep talking about testing the velocity by going down the actual hill for ~1 min)).

In this interval (32.2–34.2 min), the coherence dropped markedly (0.29). Rather than advancing the drag model (model C), the group's discourse was disjointed and affected by a variety of influences, including torque/wheel rotation effects (turns 1 and 3), kinetic friction (turn 2), and an on-site test (turn 4 and its subsequent turns), without considering any of them in depth. This detour reflected creative engagement and out-of-the-box thinking, in which the discussion shifted from a shared modeling trajectory to improvisational brainstorming, which resulted in a drop in coherence. At this point, without any coordination between the instructors, instructor 2 visited the group (34.2–37.6 min). First, she briefly recapped their earlier simplification ("Last time I saw you were neglecting ..."). Then, she asked them to describe their current modeling focus (Excerpt 9, turn 1).

Excerpt 9

0:34:52 1. I₂: So what influence are you going to check now?

0:34:54 2. Rami: I'm looking at the air.

0:34:56 3. Eran: Yeah, air friction.

0:34:57 4. I₂: Okay.

0:34:59 5. Olga: I have a feeling it's the wheels, maybe something like that.

0:35:01 6. I₂: So, you can check both of them.

(...)

0:35:08 7. Olga: This also ((influences the speed))

0:35:10 8. Eran: The wheels, the engine, all that.

0:35:13 9. I₂: So why don't you decide arbitrarily—first examine one thing, and then the other. Right?

0:35:22 10. Eran: ((Humming in agreement.))

In turns 2 and 3, Rami and Eran signaled alignment on the same model: finding a way to represent the air friction (drag—model C). In turn 5, Olga suggested that the differences in the vehicles' wheels might lead to differences in their end-of-slope velocities (model E). In response (turn 6), instructor 2 reframed the discussion around competing influences (air drag vs wheel rotation). When Olga and Eran implicitly noted the complexity of handling multiple factors (turns 7–8), instructor 2 steered them toward a systematic examination of each influence separately ("first examine one thing, and then the other"), which Eran endorsed.

After Excerpt 9, instructor 2 remained with the group for roughly 2 min longer and provided targeted scaffolding. She confirmed they had a usable expression for air drag, and asked them to clarify the meaning of each term, and to verify dimensional consistency (unit balance). This brief guidance helped the group realign around the drag model. In the subsequent group-alone interval (37.6–39.9 min), the group quickly converged on developing the mathematical model for estimating the vehicles' speed when taking the drag force into account (model C). Excerpt 10 is drawn from this interval.

Excerpt 10

0:38:00 1. Rami: I need a drag coefficient. ((To Eran)) Could you find me a drag coefficient? One for the car, one for the bike.

0:38:11 2. Eran: Here, Toyota Corolla—0.3.

0:38:15 3. Rami: 0.3. And now, ((to Eran)) can you find me one for a motorcycle or something?

0:38:29 4. Olga: ((Verifying formula units)) Newton is kilogram times meter per second squared ... but wait ...

0:38:50 5. Rami: ((To Eran)) By the way, do you know the units of this coefficient? Does it have any?

0:38:53 6. Eran: Umm ...

0:39:00 7. Olga: ((Drawing a force diagram.)) What's the direction of this ((air drag)) force, like this?

0:39:05 8. Rami: ((Looks at Olga's diagram.)) Yeah.

0:39:10 9. Olga: Oh, so we've got ... ((Keeps drawing and writing on her force diagram.))

0:39:38 10. Rami: Well, I'm not sure how I'm gonna use it ((the drag force formula)) because it's ... velocity dependent, right?

0:39:43 11. Eran: Exactly, it affects the whole thing.

This excerpt shows the group's refocusing on the drag-inclusive model to estimate the vehicles' speed. All the turns in this interval dealt with this model (i.e., DMC = 1.00), although individual members of the group tackled different aspects of the modeling work. Rami and Eran searched for the drag coefficients of a Toyota Corolla and a bike (turns 1–3), while Olga checked the units of the drag-force formula (turn 4). When seeing her work, Rami realized that he did not know the units of the drag

coefficient and asked Eran (turn 5), who was not sure (turn 6). Meanwhile, Olga was busy with something else: she drew a force diagram and wanted to be sure that the drag-force direction was correct, so she asked for Rami's confirmation (turn 7). Rami confirmed (turn 8), and Olga continued (turn 9). At this point, Rami (turn 10) flagged a challenge by noting (correctly) that the drag force formula was velocity-dependent, implicitly indicating that solving a differential equation would be required. Eran (turn 11) affirmed the additional complexity.

Overall, instructor 2's intervention temporarily focused all the group members on the same model (i.e., DMC = 1). This move facilitated a responsive interaction around different aspects of *the same* model (Model C): turns 1–6 centered on units, turns 7–9 concerned the force diagram, and turns 10–11 focused on the mathematical formulation, which helped move the group's reasoning forward.

1. Other interventions

Earlier in the session (19.0–22.6 min), the group was individually involved in different models: Olga was using constant-acceleration kinematics (model A), Eran was exploring air resistance (model C), and Rami was searching for typical car dimensions, which collectively resulted in low coherence (DMC = 0.44). Instructor 2's clarifying prompts ("What assumptions are you making?," "What else can you use besides force?") redirected them toward a simplified energy-conservation model (model B), after which they aligned (DMC = 1.00).

Later in the session, between 63.9 and 66.6 min (right after Excerpt 5), Instructor 1 again raised doubts, this time about the terminal velocity result of Model C (Table III). He questioned the group's confident claim that the bicycle would reach its terminal speed, whereas the car would not ("How do you know that the bicycle would reach it and the car wouldn't?"). Through discussion, he then asked whether the terminal velocity could function as an upper-bound for the speed, effectively pointing to the need to reiterate the modeling process with either a gentler approximation or by solving the full differential equation ($ma = mg \sin \theta - kv^2$) (model D, Table III). While Olga and Eran pursued this new challenge, Rami showed signs of disengagement by joking and describing Eran and Olga's mathematical reasoning as "a secret language." Though the group shifted to a more complex model (model D, Table III), the divergence in participation led to a temporary drop in DMC from 0.99 to 0.82.

Instructor 2's brief final visit, at 69.1–70.0 min, served as both a diagnosis and affirmation. She prompted reflection ("So what does that mean?"), confirmed earlier insights, and nudged the group to consider whether the predicted terminal velocity could actually be reached. This cross-instructor convergence may also have mattered: when the two instructors independently suggested that the group proceed in the same direction, the group appeared more

confident about how to proceed. In this case, Rami reengaged by asking his partners for explanations and clarifying steps. The group restored full participation and alignment around solving the problem within model D, with the DMC increasing from 0.82 to 1.00.

VI. DISCUSSION

This study investigated how a small group of preservice physics teachers collaboratively engaged in solving an ill-structured physics problem. The findings illustrate how the learners engaged in layered modeling progressed iteratively through the different epistemic phases of the modeling process, the ways in which the coherence of their discourse in terms of the in-the-moment operational model evolved, and how instructor interventions influenced this process.

The physics education literature includes descriptive and prescriptive accounts of modeling [1,4–6] as well as studies of experts' thinking and application of models [38,39]. However, studies that track how mathematical modeling unfolds in action as a learning activity are still rare. Our analysis traced the *in vivo* collaborative modeling process turn by turn, examined the epistemic phases and operational models employed in each turn, and its *discourse model coherence*. By combining turn-level coding of epistemic function (phase) and content (operational model) with time-resolved measures of coherence (DMC), we also provided a way to visualize how collaborative modeling unfolds and evolves in the JPS [51] (Figs. 2 and 3). The analysis uncovered how the group co-constructed and repeatedly updated their JPS, including a shared conception of "what the problem is," which physical model was currently in play, and what counted as progress within it.

A. Macrolevel linear progression along the epistemic phases of the modeling process

At a macrolevel, the modeling cycle emerged as an overall directional progression across phases, even though the group's work was not strictly linear. In an overview of Fig. 2, the group's work exhibited "sawtooth" cycles, as indicated by the mean-field trendlines, where the trajectory progressed from earlier to later modeling phases. The fact that the regression slopes for the two main cycles were positive provides an empirical justification for why multi-phase cyclic accounts of modeling have a rationale in action [5,6,16,17]. At the same time, the cycles differed quantitatively: the fitted slope for model C (purple) was lower than the fitted slope for model B (red), indicating slower phase progression in the more complex, drag-inclusive model.

B. Microlevel fluctuations between epistemic phases of modeling within the same model

A high-time-resolution view revealed that these local departures took on a systematic form. The group discourse

exhibited *phase fluctuations*: recurring back-and-forth transitions between epistemic modeling phases while remaining within a given physical model. The greatest fluctuations were most prevalent when the group engaged in finding the vehicles' velocities within the drag-inclusive, terminal-velocity model (see purple markers in Fig. 2, especially between minutes 45 and 65). As shown in the findings section, even when the group was ostensibly "processing" the math, members intermittently shifted to interpreting symbols in physical terms and verifying dimensional consistency (e.g., Excerpt 3). These microshifts showed that evaluation was not only a postcalculation stage, as linear conceptualizations of the modeling cycle suggest [5,6,16,17], but rather a recurrent, in-the-moment resource enacted during processing and interpreting, thus extending accounts of evaluation strategies that are often emphasized as end-of-solution checks [30]. These shifts foreground the blended nature of physics modeling, in which learners continuously coordinate formal mathematical work with conceptual physical sensemaking [19,20,23,62,92], thereby assigning meaning to mathematical operations and guiding reasoning and debugging.

These microlevel phase fluctuations can also be related to the macrolevel structure of the phase progression. The coefficient of determination R^2 was lower for model C than for model B, indicating that the phase progression in the more complex model deviated more from a linear trend. Methodologically, this raises the possibility that R^2 (as well as the phase/time slope) could serve as a quantitative indicator of model complexity *in situ*. In the present case, it was not merely a statistical descriptor of the timeline, but an index of how nonlinear the modeling cycle became for these learners when implementing a given model. For instance, the lower R^2 in model C aligned with the group's visible uncertainty and recurrent phase backtracking while working with a more demanding model formulation.

C. Microlevel fluctuations between models within the same epistemic phase of modeling

The second kind of fluctuation involved *model fluctuations*: fluctuations between the different operational physical models that the solvers applied. These transitions served distinct functions that were typically related to the epistemic phase in which they appeared. For example, fluctuating between models in the *identifying* phase (e.g., models A and C between minutes 5 and 10, and models C and E between minutes 30 and 35 in Fig. 2) constituted a "competition" between models that could potentially address the problem (frictionless vs air drag; air drag vs wheel rotation). However, the fluctuation between models in the *evaluating* phase (e.g., models B and C between minutes 55 and 65 in Fig. 2, represented by red-filled, purple-edged markers) corresponded to a "comparison" between the *results* of different models.

This cross-model "sanity-checking" can be read as a specific instantiation of evaluation-as-comparison, but with a layered-modeling twist [26,33]: rather than relying on intuition or empirical checks [30], the group treated a simpler model as a physically meaningful benchmark against which a more complex model's result could be judged. Here, model C's mathematical result was checked against physically plausible bounds from a simpler model (model B), such that the layered approach became part of the group's evaluation work. More broadly, because models are abstractions of reality, using simpler ones to confirm (or question) the plausibility of more complex ones is a key pedagogical and cognitive outcome of the layered approach.

Notably, we showed that when the group transitioned from the frictionless, mechanical-energy-conservation model (model B, red) to incorporating air drag (model C, purple) (see Fig. 2), they restarted the modeling cycle from the third phase (mathematizing) rather than from the first phase (understanding and identifying). The reason was that they had already taken several factors into account that oriented them toward several candidate physical models during the first 20 min of work. This underscores the function of early "brainstorming": even when it manifested as low coherence and in turns with "no specific model," it can be understood as an expansion of the JPS in which candidate models and assumptions are put into play before convergence is achieved [47], preparing the learners to reenter the modeling cycle without starting from scratch when iterating toward more complex models for this ill-structured problem.

D. Physics-informed considerations that simplify mathematical expression

We also found that applying extreme and special case reasoning [92] could help not only in making sense of a mathematical result but also in the process of obtaining such a result. When the students tried to solve the drag-inclusive Newton's equation for the vehicle going down the slope, they encountered a snag: the equation of motion was differential, because the acceleration depended on the velocity (see Excerpt 10). Guided by one of the instructors, the students realized that for the terminal velocity, the acceleration was zero, and the equation could easily be solved, and would yield a solution for the (terminal) velocity of the vehicle. In terms of the modeling phases, we can think of this sequence as adding an additional (optional) step *between* mathematizing and processing that consisted of simplifying the "full" mathematical model to a "reduced" one by adding physically informed assumptions and only then processing the math. This reduction step reflects another blended math-physics reasoning mechanism, in which physical constraints helped simplify the mathematical formulation. Seen through this lens, the terminal-velocity move was not merely a computational

shortcut but an idealizing, approximating step in which mathematics functioned to construct a tractable model of a physical situation [46].

This also speaks to the *complexity* and to some extent also the *dynamicity* of the task [29]. Not only did a and v covary in the drag-inclusive equation $ma = mg \sin \theta - kv^2$, but different models foregrounded different assumptions, variables, and solution paths. As the group moved between these models, their representation of the problem was repeatedly reconfigured. The group's eventual attempt to return to the full differential formulation (model D) demonstrates how the evolving problem representation was a source of complexity and clarifies why a *layered modeling* approach can be productive in such tasks [26,33]. If they had started with the full differential model, their conceptual traction and phase progression might have been hindered, and they would have missed out on the later cross-validation enabled by comparing results to simpler models.

E. Time-resolved coherence and the role of instructor prompts

Methodologically, this study suggests robust ways to examine collaboration in real time and to operationalize the JPS as a time-resolved modeling trajectory. Whereas previous studies have compared groups' related discussions in a summative manner [49], we showed how discourse coherence can be tracked within a group in a time-resolved way. This dynamic view showed that coherence is not a single outcome but rather fluctuating "islands" of alignment and divergence, consisting of episodes of metastability in which the group traded off coordination with adaptability during a complex collaboration [74]. This perspective is also congruent with broader accounts of cognitive *convergence* in collaborative learning [47,75] by specifying *what* is converging (the operational physical model) and by rendering convergence/divergence visible at fine temporal grain.

With respect to epistemic phases of the modeling process, the DMC pattern in Fig. 3 showed that in the first 20 min, in which the group was mainly busy understanding the problem and identifying key components, the DMC was relatively low, reflecting divergent explorations of possible physical models. Although at first glance this might appear undesirable, in this case, it seemed productive: the group members explored multiple approaches before discussing them and deciding which path to follow first (especially after a constructive interaction with instructor 2 around 20 min into the activity). This early exploratory phase also later enabled them to employ a more complex drag-inclusive physical model without having to rethink the problem from scratch. Notably, during this stage, the group members often appeared aligned in the epistemic modeling phase (mainly in the Understanding and Identifying epistemic

phases) even when they were not aligned on a single physical model (as indicated by the low DMC). This suggests that coherence in epistemic function and coherence in model content may constitute distinct dimensions of collaborative modeling. Future research could shed light on the possible complementary nature of these two aspects of collaboration.

Instructor interventions were shown to shift the model coherence dynamics in contrasting ways. Interventions that clarified assumptions, narrowed scope, or prompted explicit model articulation tended to restore or strengthen alignment around a shared modeling direction, supporting coordinated progress within a model. Interpreted through Reiser's account of scaffolding mechanisms [57], such moves functioned primarily as structuring supports by temporarily reducing the degrees of freedom (e.g., by foregrounding an influence or a next step) to help the learners maintain a coordinated modeling trajectory. This interpretation is also consistent with broader accounts of scaffolding in ill-structured problem solving that emphasize cognitive, metacognitive, and epistemic support [84], and the distinction between contingent "soft" scaffolds enacted in-the-moment vs preplanned "hard" scaffolds [58,59].

By contrast, interventions that relaunched questions about model validity or reinjected doubts at times when the group had begun to stabilize were associated with significant drops in DMC, even when they raised scientifically legitimate concerns. In Reiser's terms [57], these moves can be read as *problematizing* supports that deliberately surface unresolved assumptions and the limitations of a current model, thereby increasing epistemic pressure and, at least locally, disrupting coordination. Relaunching epistemic uncertainty sometimes broadened the modeling space, pinpointing other candidate influences (e.g., wheel rotation), kindled out-of-the-box ideas, such as driving to the physical venue and carrying out an actual experiment within the activity time frame (Excerpt 8), and at later stages encouraged the group to attempt a more sophisticated model formulation (model D, which accounted for varying velocity). This framing highlights a central tension in scaffolding complex work: structuring supports can stabilize shared direction and raise coherence, whereas problematizing supports can lower short-term coherence. Nevertheless, when the timing of problematizing is well attuned to students' understanding, problematizing enables deeper reengagement with model validity and refinement, cultivating epistemic uncertainty as a driving force for learning [93–96].

F. Why *in vivo* group discourse matters

In both physics and mathematics education, many studies have examined students' in-the-moment reasoning during think-aloud interviews [67,80,97–99] rather than during authentic work in group interactions. Although valuable, interview-based methods can shape the interactional frame

and learners' displayed reasoning, and they entail continuous interviewer influence on what becomes salient in the moment [100,101]. As Czochoer herself noted [67], she interacted with students' work "by posing reflective questions which challenged the students' models and assumptions" (p. 102) and explicitly called for examining how such interventions shape the observed transitions. Our setting complements this work by tracing reasoning in a more naturalistic small-group space, where the learners worked independently for extended stretches and instructor visits were intermittent, thus enabling a time-resolved analysis of how such interactions relate to changes in the epistemic phase of modeling, the operational model that was chosen, and the coherence of the discussion.

Some studies have examined the modeling paths of groups but have not attended to the contribution of each individual and the mutual alignment among group members [42,43]. This approach limits what can be said about how the JPS is assembled and renegotiated moment by moment, because the group's discourse convergence and divergence processes are unavailable for scrutiny. Our treatment of group coherence using the DMC construct demonstrated the variance of intragroup content alignment during the modeling process, while connecting the individual unit of analysis with the group level of analysis. Substantively, this highlights that coherence does not require everyone to be doing the same subtask at the same time; rather, it can reflect distributed yet epistemically unified activity [83], in which group members pursue complementary actions while remaining anchored to a shared operational model. In this sense, DMC also extends the classic notions of *grounding* and *shared understanding* [52,73,78] to modeling-rich problems: what must be grounded is not only "what we are talking about," but the model assumptions and principles that give utterances their intended meaning within the *joint problem space* [51].

VII. CONCLUSION AND IMPLICATIONS

This study examined the *in vivo* collaborative mathematical modeling of an ill-structured mechanics problem of a triad of preservice physics teachers. By jointly analyzing the epistemic phase, the operational physical model, and the time-resolved *discourse model coherence* (DMC), we showed how collaborative problem solving unfolds as a coordinated activity around evolving assumptions and solution plans, including how mathematical and physical reasoning are interwoven moment by moment [19,20,66], and how in-the-moment interactions with the instructors shaped this process.

During the collaborative work on the authentic problem, modeling did not unfold linearly; instead, it proceeded through iterative shifts across epistemic phases and candidate operational models, where the ongoing evaluations of interim products shaped what the group did next.

The modeling progressed in a broadly directional multi-phase cycle, albeit with microlevel fluctuations: the learners repeatedly shifted back and forth among various epistemic phases of modeling. Specifically, interpretation and evaluation operated as in-the-moment resources during the mathematical processing rather than only as end-stage checks [30]. Model switching had its own functionality: early transitions supported competition among candidate models, whereas later transitions supported evaluation-as-comparison, in which a simpler model served as a physically meaningful benchmark for judging a more complex one [26,33]. These dynamics are amplified in ill-structured contexts with covarying quantities and changing constraints, which increase the need for continuous interpretation, validation, and simplification during the modeling process [29,31,34,35]. The findings also suggested that brief interactions with instructors can both stabilize and destabilize the group's epistemic coherence, depending on how they restructure or problematize the group's evolving JPS [54].

Methodologically, this study contributes a potent way to visualize the evolving JPS [51] in a time-resolved manner by linking turn-level coding of epistemic function (phase) and epistemic content (operational model) to a group-level coherence index (DMC). This approach bridges the individual and group as units of analysis [83,102,103] and makes visible the within-group convergences and divergences that would have been missed by summative outcomes alone [49,75].

Pedagogically, the analyses underscore both the promise and delicacy involved in facilitating students' engagement in mathematical modeling. Interactions with instructors that clarified assumptions, narrowed scope, or prompted explicit model articulation tended to stabilize the group's shared direction, whereas interactions that reopened model validity tended to disrupt coherence, but when well timed could support deeper refinement. In Reiser's terms [57], these contrasting moves can be understood as *structuring* and *problematizing* supports. Viewed through the lens of *productive uncertainty*, the instructional challenge is therefore not simply to maximize stability or to disrupt it but to assess the nature and level of uncertainty in the group's discussions [27,93]. When the group's uncertainty is too low, this might suggest that they have converged prematurely on an underexamined model. When it is too high, they may lack a sufficiently shared problem space to make progress. In the former case, productive interventions could reopen the work by encouraging the students to articulate assumptions, compare candidate models, or justify why a result should be trusted. In the latter, productive interventions may be more structuring, in that they could help students surface sources of ambiguity, identify areas of agreement and disagreement, and formulate a next step that could propel the investigation forward.

This points to the value of instructional training that helps facilitators recognize signs of shared solution framing, moments of epistemic dissonance, and transitions between conceptual reasoning and mathematization [20,27,57]. For teacher education, this suggests implementing curricular materials organized around recurring pedagogical decision points such as whether to stabilize or reopen the work, what to keep constant and what to leave open, and how to support students in making accountable decisions about the next step in their investigation [93,104]. These materials could include annotated transcripts, video examples, rehearsal activities, and discussion prompts that would help prospective facilitators interpret whether a group is productively exploring alternatives, prematurely closing uncertainty, or becoming uncertain without a workable shared direction.

Finally, this study supports task designs that afford layered modeling approaches, where learners revisit the same situation through a gradual increase in model complexity and compare alternatives in ways that mirror epistemologically authentic inquiry and knowledge-building practices [29,105,106]. In particular, prompting learners to

benchmark results from more complex models against simpler, physically interpretable bounds (e.g., upper-limit sanity checks) can function as a productive evaluation resource [26,33]. Future work should test the generality of these coherence and intervention patterns across topics and settings and develop supports for coherence-sensitive facilitation in teacher education and classroom practice [50].

ACKNOWLEDGMENTS

We would like to express our deep gratitude to Tal Peer for his assistance in facilitating the activity, to Asnat Zohar for her help in the data collection, and to all the members of the Physics Education Research Group at the Technion for their helpful feedback on early versions of this manuscript.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because of legal restrictions preventing unrestricted public distribution. The data are available from the authors upon reasonable request.

-
- [1] D. Hestenes, Toward a modeling theory of physics instruction, *Am. J. Phys.* **55**, 440 (1987).
 - [2] E. Brewster, Modeling theory applied: Modeling instruction in introductory physics, *Am. J. Phys.* **76**, 1155 (2008).
 - [3] J. K. Gilbert and R. Justi, Models of modelling, in *Modelling-Based Teaching in Science Education*, edited by J. K. Gilbert and R. Justi (Springer International Publishing, Cham, 2016), pp. 17–40.
 - [4] O. Uhdén, R. Karam, M. Pietrocola, and G. Pospiech, Modelling mathematical reasoning in physics education, *Sci. & Educ.* **21**, 485 (2012).
 - [5] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, *Sci. & Educ.* **24**, 561 (2015).
 - [6] D. Sands, Modeling as sensemaking: Towards a theory of modelling in physics education, *Eur. J. Phys.* **42**, 064001 (2021).
 - [7] M. Good, E. Marshman, E. Yerushalmi, and C. Singh, The value of using different types of physics problems to help students become proficient problem-solvers, *Phys. Educ.* **59**, 015018 (2024).
 - [8] NGSS Lead States, *Next Generation Science Standards: For States, By States* (National Academies Press, Washington, DC, 2013).
 - [9] A. Razi and G. Zhou, STEM, iSTEM, and STEAM: What is next?, *Int. J. Technol. Des. Educ.* **5**, 1 (2022).
 - [10] N. C. Kawa *et al.*, Training wicked scientists for a world of wicked problems, *Humanit. Soc. Sci. Commun.* **8**, 189 (2021).
 - [11] National Research Council (U.S.), *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (The National Academies Press, Washington, DC, 2012).
 - [12] C. A. Ogilvie, Changes in students' problem-solving strategies in a course that includes context-rich, multifaceted problems, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020102 (2009).
 - [13] M. Williams, The missing curriculum in physics problem-solving education, *Sci. & Educ.* **27**, 299 (2018).
 - [14] Z. Hazari, P. M. Sadler, and R. H. Tai, Gender differences in the high school and affective experiences of introductory college physics students, *Phys. Teach.* **46**, 423 (2008).
 - [15] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, *Sci. Educ.* **91**, 847 (2007).
 - [16] W. Blum and D. Leiß, How do students and teachers deal with modelling problems?, in *Mathematical Modelling* (Elsevier, New York, 2007), pp. 222–231.
 - [17] G. Kaiser and P. Stender, Complex modelling problems in co-operative, self-directed learning environments, in *Teaching Mathematical Modelling: Connecting to Research and Practice*, edited by G. A. Stillman, G. Kaiser, W. Blum, and J. P. Brown (Springer Netherlands, Dordrecht, 2013), pp. 277–293.
 - [18] M. Niss and W. Blum, *The Learning and Teaching of Mathematical Modelling*, 1st ed. (Routledge, London and New York, 2020).

- [19] T. J. Bing and E. F. Redish, The cognitive blending of mathematics and physics knowledge, *AIP Conf. Proc.* **883**, 26 (2007).
- [20] E. Kuo, M. M. Hull, A. Gupta, and A. Elby, How students blend conceptual and formal mathematical reasoning in solving physics problems, *Sci. Educ.* **97**, 32 (2013).
- [21] E. F. Redish and A. Gupta, Making meaning with math in physics: A semantic analysis, in *Physics Community and Cooperation: GIREP-EPEC & PHEC 2009 International Conference*, edited by D. Raine, C. Hurkett, and L. Rogers (Lulu/The Centre for Interdisciplinary Science, University of Leicester, Leicester, UK, 2010), Vol. 1, pp. 244–260.
- [22] S. Van Den Eynde, B. P. Schermerhorn, J. Deprez, M. Goedhart, J. R. Thompson, and M. De Cock, Dynamic conceptual blending analysis to model student reasoning processes while integrating mathematics and physics: A case study in the context of the heat equation, *Phys. Rev. Phys. Educ. Res.* **16**, 010114 (2020).
- [23] B. P. Schermerhorn and J. R. Thompson, Making context explicit in equation construction and interpretation: Symbolic blending, *Phys. Rev. Phys. Educ. Res.* **19**, 020149 (2023).
- [24] M. Niss, Towards a conceptual framework for identifying student difficulties with solving real-world problems in physics, *Latin-Am. J. Phys. Educ.* **6**, 3 (2012).
- [25] E. F. Redish, Using math in physics: 2. Estimation, *Phys. Teach.* **59**, 525 (2021).
- [26] T. Reshef-Israeli and S. Kapon, Engaging students in collaborative layered modelling of a real-world problem, *Phys. Educ.* **60**, 055026 (2025).
- [27] X. Ge and S. M. Land, A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions, *Educ. Technol. Res. Dev.* **52**, 5 (2004).
- [28] N. Shin, D. H. Jonassen, and S. McGee, Predictors of well-structured and ill-structured problem solving in an astronomy simulation, *J. Res. Sci. Teach.* **40**, 6 (2003).
- [29] D. H. Jonassen, How does problem solving vary?, in *Learning to Solve Problems: A Handbook for Designing Problem-Solving Learning Environments* (Routledge, New York, 2011), pp. 1–24.
- [30] A. R. Akinyemi, M. E. Loverude, and J. R. Thompson, Solution evaluation strategies used by first-year physics students, *Phys. Rev. Phys. Educ. Res.* **21**, 020114 (2025).
- [31] C. Zimmerman, A. Olsho, T. I. Smith, P. Eaton, and S. W. Brahmia, Assessing physics quantitative literacy development in algebra-based physics, *Phys. Rev. Phys. Educ. Res.* **21**, 020108 (2025).
- [32] S. Kapon and M. Schvartzter, Guided inquiry into a physics equation, *Cognit. Instr.* **42**, 159 (2024).
- [33] T. Reshef-Israeli and S. Kapon, Layered modeling of a real-world drag-force problem: A pedagogical journey from idealized to realistic physics, *Eur. J. Phys.* **46**, 065701 (2025).
- [34] D. H. Jonassen, Understanding causal relationships in problems, in *Learning to Solve Problems: A Handbook for Designing Problem-Solving Learning Environments* (Routledge, New York, 2011), pp. 267–284.
- [35] D. H. Jonassen, What makes scientific problems difficult?, in *Learning to Solve Complex Scientific Problems*, edited by D. H. Jonassen (Routledge, New York, 2007), pp. 3–23.
- [36] R. Borromeo Ferri, Theoretical and empirical differentiations of phases in the modelling process, *Z. Didakt. Math.* **38**, 86 (2006).
- [37] M. Pietrocola, Mathematics as structural language of physical thought, in *Connecting Research in Physics Education with Teacher Education*, edited by M. Vicentini and E. Sassi (International Commission on Physics Education, 2008), Vol. 2, <https://www.per-central.org/items/detail.cfm?ID=9881>.
- [38] J. J. Clement, Case study of model construction and criticism in expert reasoning*, in *Creative Model Construction in Scientists and Students*, edited by J. J. Clement (Springer Netherlands, Dordrecht, 2008), pp. 67–96.
- [39] R. Karam, Framing the structural role of mathematics in physics lectures: A case study on electromagnetism, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010119 (2014).
- [40] C. C. Silva, The role of models and analogies in the electromagnetic theory: A historical case study, *Sci. & Educ.* **16**, 835 (2007).
- [41] R. D. Tweney, Representing the electromagnetic field: How Maxwell's mathematics empowered Faraday's field theory, *Sci. & Educ.* **20**, 687 (2011).
- [42] L. Albarracin, J. Arleback, E. Civil, and N. Gorgorio, Extending modelling activity diagrams as a tool to characterise mathematical modelling processes, *Math. Enthusiast* **16**, 211 (2019).
- [43] J. B. Årleback, On the use of realistic Fermi problems for introducing mathematical modelling in school, *Math. Enthusiast* **6**, 331 (2009).
- [44] S. Kapon and L. Berland, Epistemic models of sense-making and reasoning, in *The International Handbook of Physics Education Research: Learning Physics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing LLC, Melville, New York, 2023), pp. 12–1–12–22.
- [45] R. Karam and O. Krey, Quod erat demonstrandum: Understanding and explaining equations in physics teacher education, *Sci. & Educ.* **24**, 661 (2015).
- [46] E. Palmgren, T. Kokkonen, and J. Bruun, Roles of mathematics in physics education: A systematic review, *Phys. Rev. Phys. Educ. Res.* **21**, 020602 (2025).
- [47] J. Roschelle, Learning by collaborating: Convergent conceptual change, *J. Learn. Sci.* **2**, 235 (1992).
- [48] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *Am. J. Phys.* **60**, 627 (1992).
- [49] B. Barron, When smart groups fail, *J. Learn. Sci.* **12**, 307 (2003).
- [50] A. C. Graesser, S. M. Fiore, S. Greiff, J. Andrews-Todd, P. W. Foltz, and F. W. Hesse, Advancing the science of collaborative problem solving, *Psychol. Sci. Publ. Interest* **19**, 59 (2018).
- [51] J. Roschelle and S. D. Teasley, The construction of shared knowledge in collaborative problem solving, in *Computer Supported Collaborative Learning*, edited

- by C. O'Malley (Springer Berlin Heidelberg, Berlin, Heidelberg, 1995), pp. 69–97.
- [52] H. H. Clark and S. E. Brennan, Grounding in communication, in *Perspectives on Socially Shared Cognition* (American Psychological Association, Washington, DC, 1991), pp. 127–149.
- [53] A. Meier, H. Spada, and N. Rummel, A rating scheme for assessing the quality of computer-supported collaboration processes, *Comput. Supported Learn.* **2**, 63 (2007).
- [54] N. Rummel, A. Deiglmayr, H. Spada, G. Kahrimanis, and N. Avouris, Analyzing collaborative interactions across domains and settings: An adaptable rating scheme, in *Analyzing Interactions in CSCL*, edited by S. Puntambekar, G. Erkens, and C. Hmelo-Silver (Springer, Boston, MA, 2011), pp. 367–390.
- [55] C. E. Hmelo-Silver, Analyzing collaborative knowledge construction, *Comput. Educ.* **41**, 397 (2003).
- [56] N. Rummel and H. Spada, Instructional support for collaboration in desktop video conference settings: How it can be achieved and assessed, in *Barriers and Biases in Computer-Mediated Knowledge Communication*, edited by R. Bromme, F. W. Hesse, and H. Spada (Springer, Boston, MA, 2005), pp. 59–88.
- [57] B. J. Reiser, Scaffolding complex learning: The mechanisms of structuring and problematizing student work, *J. Learn. Sci.* **13**, 273 (2004).
- [58] T. A. Brush and J. W. Saye, A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment, *J. Interactive Online Learn.* **1**, 1 (2002).
- [59] J. W. Saye and T. Brush, Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments, *Educ. Technol. Res. Dev.* **50**, 77 (2002).
- [60] C. E. Hmelo-Silver and H. S. Barrows, Facilitating collaborative knowledge building, *Cognit. Instr.* **26**, 48 (2008).
- [61] S. Mahajan, Welcome to the back of the envelope, *Am. J. Phys.* **85**, 959 (2017).
- [62] E. F. Redish, Using math in physics: Overview, *Phys. Teach.* **59**, 314 (2021).
- [63] E. Redish, Using math in physics: 4. Toy models, *Phys. Teach.* **59**, 683 (2021).
- [64] E. Palmgren and T. Rasa, Modelling roles of mathematics in physics: Perspectives for physics education, *Sci. & Educ.* **33**, 365 (2024).
- [65] B. Sherin, Common sense clarified: The role of intuitive knowledge in physics problem solving, *J. Res. Sci. Teach.* **43**, 535 (2006).
- [66] M. M. Hull, E. Kuo, A. Gupta, and A. Elby, Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010105 (2013).
- [67] J. A. Czochoer, Introducing modeling transition diagrams as a tool to connect mathematical modeling to mathematical thinking, *Math. Think. Learn.* **18**, 77 (2016).
- [68] V. Shekoyan and E. Etkina, Introducing ill-structured problems in introductory physics recitations, *AIP Conf. Proc.* **951**, 192 (2007).
- [69] W. Blum and R. B. Ferri, Mathematical modelling: Can it be taught and learnt?, *J. Math. Model. Appl.* **1**, 45 (2009).
- [70] J. Cai, G. Kaiser, R. Leikin, and J. B. Årleback, The relationship of mathematical modelling education to problem solving, problem posing, and creativity, in *International Horizons in Mathematics Modelling Education*, edited by T. Ikeda, A. Saeki, V. Geiger, and G. Kaiser (Springer Nature, Switzerland, Cham, 2025), pp. 47–67.
- [71] J. A. Czochoer, S. S. Kularajan, E. Roan, and R. Sigley, Validating a multiple-choice modelling competencies assessment, in *Advancing and Consolidating Mathematical Modelling*, edited by G. Greefrath, S. Carreira, and G. A. Stillman (Springer International Publishing, Cham, 2023), pp. 161–176.
- [72] S. M. Fiore, Interdisciplinarity as teamwork: How the science of teams can inform team science, *Small Group Res.* **39**, 251 (2008).
- [73] P. Dillenbourg and D. Traum, Sharing solutions: Persistence and grounding in multimodal collaborative problem solving, *J. Learn. Sci.* **15**, 121 (2006).
- [74] R. Pacheco-Vega, R. D. Hardin, A. C. C. Coolen, G. E. Marcus, C. Nierstrasz, J. Sulik, S. E. van der Leeuw, P. F. M. J. Verschure, and A. von Schenk, What is collaboration good for?, in *The Nature and Dynamics of Collaboration*, edited by P. F. M. J. Verschure (The MIT Press, Cambridge, MA, 2025), pp. 297–316, [10.7551/mitpress/15533.003.0023](https://doi.org/10.7551/mitpress/15533.003.0023).
- [75] S. D. Teasley, B. Ave, A. Arbor, F. Fischer, A. Weinberger, and K. Stegmann, Cognitive convergence in collaborative learning, in *International Perspectives in the Learning Sciences: Creating a Learning World. Proceedings of the Eighth International Conference of the Learning Sciences—ICLS* (International Society of the Learning Sciences, 2008), Vol. 3, pp. 360–367.
- [76] M. Lipman, The reflective model of educational practice, in *Thinking in Education*, 2nd ed. (Cambridge University Press, Cambridge, England, 2003), pp. 9–27.
- [77] M. Lipman, Thinking in community, in *Thinking in Education*, 2nd ed. (Cambridge University Press, Cambridge, England, 2003) pp. 83–104.
- [78] H. H. Clark and E. F. Schaefer, Contributing to discourse, *Cognit. Sci.* **13**, 259 (1989).
- [79] N. Arthars, L. Markauskaite, and P. Goodyear, Constructing shared understanding of complex interdisciplinary problems: Epistemic games in interdisciplinary teamwork, *J. Learn. Sci.* **33**, 405 (2024).
- [80] T. Ole. B. Odden, and R. S. Russ, Sensemaking epistemic game: A model of student sensemaking processes in introductory physics, *Phys. Rev. Phys. Educ. Res.* **14**, 020122 (2018).
- [81] G. J. Kelly and P. Licona, Epistemic practices and science education, in *History, Philosophy and Science Teaching*, edited by M. R. Matthews (Springer International Publishing, Cham, 2018), pp. 139–165.
- [82] I. Mulder, J. Swaak, and J. Kessels, Assessing group learning and shared understanding in technology-mediated interaction, *J. Educ. Technol. Soc.* **5**, 35 (2002), <https://www.jstor.org/stable/jeductechsoci.5.1.35>.

- [83] E. Hutchins, The distributed cognition perspective on human interaction, in *Roots of Human Sociality*, 1st ed., edited by N. J. Enfield and S. C. Levinson (Routledge, New York, 2020), pp. 375–398.
- [84] K. S. Kitchener, Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing, *Hum. Dev.* **26**, 222 (1983).
- [85] C. E. Hmelo-Silver, R. G. Duncan, and C. A. Chinn, Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006), *Educ. Psychol.* **42**, 99 (2007).
- [86] D. H. Jonassen, How does problem solving vary?, in *Learning to Solve Problems* (Routledge, New York, 2010), pp. 1–24.
- [87] B. J. Reiser, Scaffolding complex learning: The mechanisms of structuring, and problematizing student work, *J. Learn. Sci.*, **13**, 273 (2018).
- [88] N. Geirsdotter Bækkelund, A. Isdal Haugland, J. Sande, and J. O. Rogde Mjånes, Ill-structured problems in capstone courses—Investigating students’ approaches and experience, *Eur. J. Eng. Educ.* **1** (2025).
- [89] T. Reshef-Israeli and S. Kapon, Collaborative sensemaking: The emergence of insights in a distributed knowledge system, present at PER Conf. 2024, Boston, MA, [10.1119/perc.2024.pr.Reshef-Israeli](https://doi.org/10.1119/perc.2024.pr.Reshef-Israeli), pp. 361–366.
- [90] R. Thornberg and K. Charmaz, Grounded theory and theoretical coding, in *The SAGE Handbook of Qualitative Data Analysis* (SAGE Publications, London, 2014), pp. 153–169.
- [91] B. Rosner, *Fundamentals of Biostatistics*, 7th ed. (Brooks/Cole, Cengage Learning, Boston, 2011).
- [92] M. Eichenlaub and E. F. Redish, Blending physical knowledge with mathematical form in physics problem solving, in *Mathematics in Physics Education*, edited by G. Pospiech, M. Michelini, and B.-S. Eylon (Springer International Publishing, Cham, 2019), pp. 127–151.
- [93] E. Manz and E. Suárez, Supporting teachers to negotiate uncertainty for science, students, and teaching, *Sci. Educ.* **102**, 771 (2018).
- [94] J. Watkins, D. Hammer, J. Radoff, L. Z. Jaber, and A. M. Phillips, Positioning as not-understanding: The value of showing uncertainty for engaging in science, *J. Res. Sci. Teach.* **55**, 573 (2018).
- [95] J. Watkins and E. Manz, Characterizing pedagogical decision points in sense-making conversations motivated by scientific uncertainty, *Sci. Educ.* **106**, 1408 (2022).
- [96] Y.-C. Chen, Epistemic uncertainty and the support of productive struggle during scientific modeling for knowledge co-development, *J. Res. Sci. Teach.* **59**, 383 (2022).
- [97] Z. Dulger and F. Ogan-Bekiroglu, Students’ metacognition knowledge and skills during physics problem-solving process, *Phys. Rev. Phys. Educ. Res.* **21**, 020106 (2025).
- [98] E. W. Burkholder, J. K. Miles, T. J. Layden, K. D. Wang, A. V. Fritz, and C. E. Wieman, Template for teaching and assessment of problem solving in introductory physics, *Phys. Rev. Phys. Educ. Res.* **16**, 010123 (2020).
- [99] D. Leiss, J. Plath, and K. Schwippert, Language and mathematics—Key factors influencing the comprehension process in reality-based tasks, *Math. Thinking Learn.* **21**, 131 (2019).
- [100] R. S. Russ, V. R. Lee, and B. L. Sherin, Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction, *Sci. Educ.* **96**, 573 (2012).
- [101] S. Kapon, Unpacking sensemaking, *Sci. Educ.* **101**, 165 (2017).
- [102] P. Dillenbourg, What do you mean by “Collaborative Learning”?, in *Collaborative-Learning: Cognitive and Computational Approaches*, edited by P. Dillenbourg (Elsevier, Oxford, 1999), pp. 1–19.
- [103] G. Stahl, Group cognition as a foundation for the new science of learning, in *New Science of Learning: Cognition, Computers, and Collaboration in Education*, edited by M. S. Khine and I. M. Saleh (Springer, New York, 2010), pp. 23–44.
- [104] E. I. Manz, *Productive Uncertainty in Science Education: Engaging Students in Meaningful Science Practice* (Teachers College Press, New York, NY, 2025).
- [105] C. A. Chinn and B. A. Malhotra, Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks, *Sci. Educ.* **86**, 175 (2002).
- [106] W. A. Sandoval, Understanding students’ practical epistemologies and their influence on learning through inquiry, *Sci. Educ.* **89**, 634 (2005).