Knowledge integration in student learning of Newton's third law: Addressing the action-reaction language and the implied causality

Lei Bao* and Joseph C. Fritchman

Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA

(Received 11 January 2021; accepted 2 August 2021; published 7 September 2021)

Newton's third law is one of the most important concepts learned early in introductory mechanics courses; however, ample studies have documented a wide range of students' misconceptions and fragmented understandings of this concept that are difficult to change through traditional instruction. This research develops a conceptual framework model to investigate students' understanding of Newton's third law through the knowledge integration perspective. The conceptual framework is established with a central idea emphasizing forces as quantitative measures of physical interactions instead of using the common action-reaction language. Guided by the conceptual framework, assessment and interview results reveal that students' concepts of Newton's third law are fragmented without deep understanding. Specifically, three main issues within students' understanding have been identified: (i) students have a disconnect between time order of events and causal reasoning, (ii) students rely on a memorized equal-and-opposite rule to identify interaction forces, and (iii) students directly link action-reaction language to the belief in a causal relation between the interaction forces. The framework is then applied to develop a new instruction intervention that explicitly targets the central idea of Newton's third law. Results from pre- and post-testing show that the intervention is effective in helping students develop more integrated understandings of Newton's third law. Overall, this study shows the potential benefits of applying the conceptual framework method to model student knowledge structures and guide assessment and instruction for promoting knowledge integration and deep learning.

DOI: 10.1103/PhysRevPhysEducRes.17.020116

I. INTRODUCTION

A primary goal in introductory science education is for students to develop deep understandings of essential scientific concepts, yet many students fail to achieve this after traditional instruction [1–5]. As shown by research, students may perform well on typical textbook problems with familiar contexts; however, when faced with unfamiliar contexts or problems requiring deeper understanding, they are likely to rely on pattern matching and memorized equations [2,3,6,7]. These tendencies exhibit the known characteristics of novice knowledge structures where knowledge is often locally clustered with links connecting familiar contexts [8–13]. This form of knowledge organization leads to novices employing problem-solving strategies that focus on memorized processes and solutions cued by surface features [10,14,15], which constrains novices' applications of a concept to contexts similar to those encountered through normal coursework, leaving them unable to transfer in novel situations. Meanwhile, experts' knowledge structures appear as integrated, hierarchically arranged networks built around a few core principles [12,16,17]. Their coherent knowledge organization enlists robust, far-reaching links connecting all elements, including surface features and elements deep in the abstract domain [8,10–13], and enables meaningful thinking which allows experts the ability to apply concepts across different domains and unfamiliar contexts [11,14,18].

Therefore, instruction with a goal of transitioning novices to experts should focus on developing coherent knowledge structures by fostering student abilities to build connections among new and existing ideas, which is a key process emphasized in the knowledge integration perspective of learning and instruction [17,19–22]. Practically, instruction aimed for knowledge integration focuses on the process of establishing organization within a student's knowledge structure through anchoring the structure around a central idea that serves as a core conceptual node for establishing a fully connected hierarchical network of knowledge. In the knowledge integration perspective, expert-level learners are able to use the central idea to solve problems with a wide range of contexts by connecting surface features and principles to the central idea and by determining the optimal strategies for applying the concept [17,19,20].

*Corresponding author.
bao.15@osu.edu

As an operational tool to explicitly model students' knowledge structures and measure the levels of knowledge integration achieved, the conceptual framework model was developed in previous studies [21–24], which can provide a modeling framework to illustrate the differences in knowledge structures between novices and experts by eliciting the existing connections that give rise to the range of students' alternative conceptions. Within a conceptual framework, a learner's ideas and connections are activated by contextual features. Experts link the activated ideas and related conceptual components to form specific reasoning pathways through the central idea, which extend to a fully integrated knowledge structure. The expert approach links a wide range of situations to the central idea, which can meaningfully and efficiently address complex problems in different contexts. Novices, however, often bypass the central idea and form locally memorized links between equations or algorithms and problems' surface-level features. The novice approach may produce quick and correct results in some limited cases, but it often fails to transfer to problem solving in novel situations.

Once the conceptual framework for a concept is established, it can be used as the basis for developing assessment tools, which make emphasis on probing students' knowledge structures and explicitly mapping their conceptual pathways to reveal their levels of knowledge integration. The assessment results can then inform instruction to emphasize specific conceptual pathways and connections during teaching to promote knowledge integration.

Recent studies developed conceptual frameworks for a number of physics topics, which have guided the creation of assessment questions that probe students' reasoning pathways and levels of knowledge integration [21,23,24]. These assessments contain a mixture of typical questions, which can be solved with memorization strategies, and atypical questions, which require integrated thinking involving the central idea. Usually, typical questions are designed with familiar contexts students encounter in coursework, while atypical questions are often designed with deep level conceptual understandings and using unfamiliar contexts not commonly seen during coursework. By comparing students' performances on typical and atypical questions, students' levels of knowledge integration can be empirically modeled and determined [21,24]. The assessments, combined with the underlying conceptual framework and results of student interviews, were shown to reveal students' reasoning pathways within their knowledge structures [21,23,24].

Furthermore, conceptual frameworks can aid in designing teaching interventions to improve knowledge integration by developing the essential missing links in students' knowledge structures. Ample research has demonstrated the importance of using knowledge-integration approaches in instruction to help form integrated knowledge structures and achieve deep conceptual understanding [17,25]. In the recent study on the conceptual framework of force and motion [23], an intervention, which explicitly introduced the central idea while connecting it directly to applications in problem solving, was shown to be effective in promoting knowledge integration.

Following the lead of prior studies of conceptual framework in force and motion, momentum, and light interference [21,23,24], this research examines student understanding on Newton's third law, which is a fundamental concept in physics. In the literature, there exist a large number of studies focusing on identification of the misconceptions and difficulties related to Newton's third law [26–38]. However, studies have shown that many misconceptions are still prevalent after instruction and may even be exacerbated, in part, due to possible misleading representations in textbooks and lectures [30,31].

Building upon the previous studies, this research develops a conceptual framework model for Newton's third law and applies it to design an assessment as well as a teaching intervention. This leads to three studies to be conducted in this research:

- Develop a conceptual framework model for Newton's third law to analyze student difficulties through the knowledge integration perspective.
- Develop a multiple-choice assessment based on the conceptual framework model, which uses typical and atypical questions to evaluate students' levels of knowledge integration and deep understanding.
- Conduct a controlled study to investigate the extent to which a conceptual-framework-based teaching intervention may improve student knowledge integration in learning Newton's third law when compared to the traditional instruction.

II. STUDY 1: DEVELOPMENT OF CONCEPTUAL FRAMEWORK FOR NEWTON'S THIRD LAW

The most critical component of a conceptual framework is the central idea of a concept, which is identified based on experts' views and provides the core explanatory mechanisms and premises for establishing the fundamental and causal relations underpinning the concept. Additionally, related contextual variables, interactive relations, and reasoning processes are identified to form the nodes and connections of the conceptual framework. Through these processes, student difficulties in understanding the concept are carefully reviewed so that the related elements and reasoning pathways existing within students' knowledge structures can be well represented with the conceptual framework.

A. Difficulties with Newton's third law

Research in physics education over the last several decades has documented extensive student difficulties and naive beliefs or misconceptions with Newtonian mechanics, which are persistent even after instruction [26,28–30,39–42]. In particular, studies on Newton's third law (N3L) reveal a range of naive beliefs, even among trained educators [28,29,31,34,35]. Many of these beliefs exist prior to entering a physics classroom as students tend to form generalizations about how the world works from an early age. Over time, these conceptions are reinforced continually and become core components of the students' knowledge structures, which can strongly resist changes to the scientific conceptions [40,42].

Most studies on student understanding of N3L focus on identifying students' misconceptions and assessing how students respond to new instruction [26–36,43]. Common examples of misconceptions include the application of a dominance principle where the faster, more massive, or acting object applies a greater force than the other object, difficulty identifying forces of a N3L pair, thinking action and reaction forces must balance each other, thinking forces are properties of objects rather than observed measures of an interaction, and assigning causal relationships to objects and forces. Each of these points towards a fundamental misunderstanding of the nature of forces, which seems to persist even after instruction [31].

Additionally, these misconceptions exist in comparable situations absent of formal physics descriptions. Studies on psychology students revealed, most notably, a launching effect or causal asymmetry similar to the dominance principle students apply in physics [37,38,44–48], in which the observers universally attributed cause to the "acting" or "dominant" object. Further research has shown that causal asymmetry is prevalent in large portions of the general population, even in early childhood development and has been measured with children less than a year old [49–52]. Thus, by the time students encounter Newton's laws, they have years of observations reinforcing their intuitive physics misconceptions. Absent of formal physics descriptions, this demonstrates an existing bias that students may possess prior to, or concurrent with, physics learning. Ideally, the bias should be minimized through the learning of N3L. However, after instruction, students still demonstrated sensitivity to irrelevant information, such as whether kinetic energy is conserved, and systematic biases, such as believing objects with higher initial speed or mass will apply stronger forces [47,53–57].

From the literature, it is clear that student understanding of the concept of causal reasoning plays a strong role in student understanding and learning N3L. However, the expert view of causality is commonly debated by philosophers and physicists [46,58]; therefore, it is important to define causality as used here. This study follows the parallel research of Chen *et al.* [59] in defining causality based on three relations necessary to form complete causal reasoning: time, covariation, and mechanism. A complete understanding of causal reasoning requires a cohesive understanding of all three elements.

Remedying these misconceptions is further hindered by the methods of teaching and language used. Common textbook and verbal explanations of N3L include statements such as "*for every action there is an equal and opposite reaction,*" which have been noted as flawed for at least the past eight decades [30,31]. The language itself may imply the existence of a cause (the action) and effect (the reaction) and does not emphasize the role of interaction. Textbooks have generally improved their discussions of N3L with statements such as "*We can now recognize force as an interaction between objects rather than as some 'thing' with an independent existence of its own*" [60], but this and similar explanations are often glossed over during lecture and are not a focal point of instruction. Furthermore, most textbooks still rely on the action-reaction language which may directly lead to a belief in causality and further difficulties associated with the dominance principle.

Recognizing that causal reasoning is essential to knowledge formation in physics [61], a small set of studies propose using causal reasoning as a basis for physics teaching [62–64]. Hung and Jonassen [65] studied covariation and mechanism-based instruction in physics education and found that the mechanism-based approach was more effective in improving students' conceptual understanding. Chen *et al.* [59] more deeply probe the role of causality in understanding N3L.

From the literature and review of current instruction on N3L, it appears that the N3L concept is intuitively difficult for students who may hold many naturally developed misconceptions. Meanwhile, the presentation in the traditional curriculum also lacks the necessary emphasis on the core mechanism of the concept, and the use of action-reaction language can be inherently misleading. Therefore, it is important to establish a conceptual framework model for N3L to aid assessment and instruction by emphasizing the correct central idea of the N3L concept.

B. Building the conceptual framework

From the commonly quoted action-reaction description of N3L, it is apparent that this definition lacks a clear description of the mechanistic nature of N3L or its basic properties, and can often lead to student difficulties [28,30,31]. The action-reaction language itself implies a time separation between action and reaction and can often be interpreted as a causal relation. However, this aspect is rarely, if ever, explicitly discussed in teaching, which leaves students to potentially draw their own conclusions on causality in N3L interactions based on colloquial interpretations of the words used rather than a concrete understanding of causality in a physics sense. Despite the issues, textbooks still commonly use a similar, yet more complete version: e.g., "*Every force occurs as one member of an action-reaction pair of forces*" [60]. Although explanations in the text have emphasized that these forces only occur in an interaction, the concept of interaction is not further repeated and enhanced in the text. Instead, the text proceeds to focus on the action-reaction and equal and opposite aspects of the interaction forces. For example, Knight [60] explicitly mentions the need for an interaction pair of forces yet still calls them an action-reaction pair. In addition, there is no explanation on the reason for using the term interaction, and the text makes very little additional emphasis on applying this idea in problem solving. The typical focus of most examples and homework is solely on the equal and opposite aspect.

Additionally, students, who are unable to recognize that N3L must involve interactions between two objects, are then often unable to identify which forces constitute a pair of interaction (or action-reaction) forces. For example, many students focus on a memorized equal-and-opposite rule where students choose any pair of forces which happen to be equal in magnitude and opposite in direction and call them the interaction pair. In situations such as a box on a horizontal surface, this rule is often applied by students to the weight and the normal force exerted on the box. This type of understanding reveals a fragmented knowledge structure relying on memorized rules. In summary, the traditional definition of N3L emphasizes features of the interaction forces but lacks the mechanistic explanation of the origin of the forces, leading to the focus on the "action and reaction" and "equal and opposite" language. This understanding lacks the deeper conceptual foundation which is the key issue to be addressed by the conceptual framework model with the central idea of N3L.

Following reviews of expert views, student difficulties, and concerns in current N3L teaching, the central idea for N3L is identified, which states: "*An interaction between two entities can be observed in terms of a pair of symmetric (i.e., equal and opposite) forces*." In addition, it is emphasized in instruction that all forces are interaction forces, and the two terms are used interchangeably to represent the same concept of interaction forces. Here, it is important to note the unidirectional order of this definition: the observed measures of the interaction are the symmetric forces and not the other way around. The reverse pathway is the commonly memorized rule that uses the "equal and opposite" feature to determine interaction pairs. Reasoning with the central idea then leads to the following elaborated properties which can be readily derived:

- Forces are observed measures of an interaction between two objects. As a result, the forces always occur in pairs acting on two interacting objects, respectively, and can be called a pair of interaction forces.
- Interaction forces must occur at the same time.
- Neither of the two forces causes the other force; the origin of a pair of interaction forces is the interaction.
- Interaction forces must both be the same type of forces (i.e., a pair of gravitational forces, frictional forces, normal forces, etc.)
- Interaction forces are symmetric such that they have an equal magnitude but act in opposite directions on the two interacting objects (never on the same object).

Building off the central idea to include different operational rules and reasoning processes from experts and novices, the conceptual framework of N3L is developed and shown in Fig. 1. The top layer contains the central idea with links to the basic properties. The bottom layer contains contextual features and variables, which are commonly used as part of an N3L problem. These include surface details about the objects involved such as mass, size, velocity, or acceleration. Additionally, the context of interaction typically includes collision, push, force at a distance, or an object at rest on top of another object. The middle layers contain the intermediate reasoning processes and operational rules and procedures, which connect to the central idea in experts' knowledge structures or are locally linked through incorrect reasoning between contexts and responses by novice students. The common applications of student reasoning, such as defining the action and reaction forces based on the dominance principle or focusing on the equal and opposite properties of N3L, are often carried out by students with local links between contexts and responses through some of the naïve type intermediate processes. The conceptual framework forms the backbone for analyzing how learners reason using N3L by combining these layers and the task goals, with arrows connecting different contextual, conceptual, and outcome components representing the possible reasoning pathways of learners. Solid arrows represent experts' conceptual pathways, while the dashed-line arrows represent pathways of novices.

To achieve a deep understanding of N3L, students must be able to identify which forces are interaction forces; on which objects each of these forces act; the magnitude, type, direction, and timing of interaction forces; and what causes these forces. For experts, the central idea serves as a central node that connects to all the different properties and reasoning processes, forming an integrated knowledge structure. Therefore, when solving N3L questions, experts recognize and activate the central idea as the guiding principle in their analysis to identify relevant variables and problem-solving approaches. On the other hand, novices make weak, local connections between the different layers, forming fragmented knowledge structures. They tend to rely on memorized solutions or matching context variables with equations based the surface features of problems without deep understanding.

In N3L problems, students often encounter scenarios where one of the two interacting forces has a dominant surface feature (such as a larger mass or a higher velocity), which is irrelevant to the properties of the interaction forces based on the N3L central idea. However, novices lack the understanding of the central idea and often place focus on the dominant surface features, which leads to difficulties in answering these questions, especially when comparing

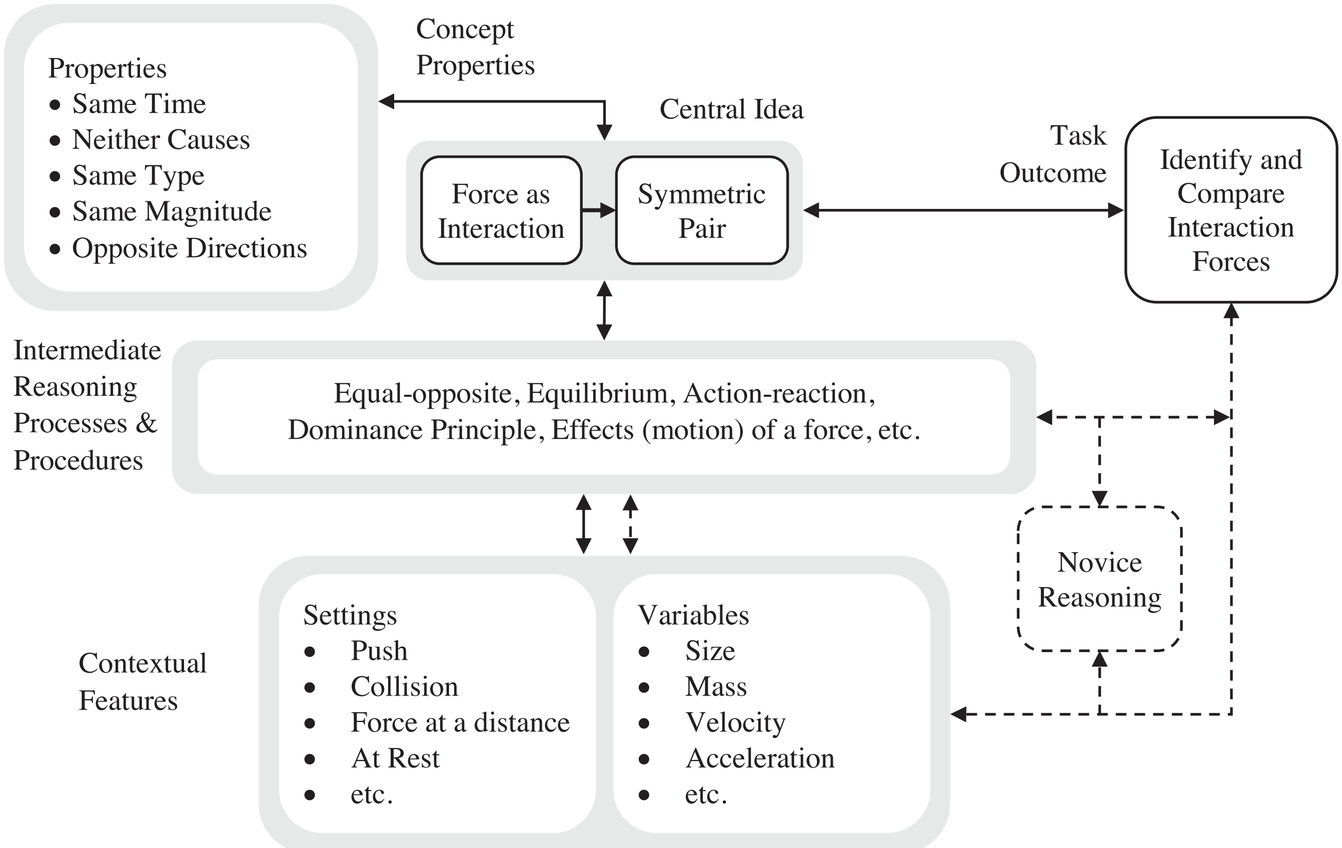

FIG. 1. Conceptual framework for Newton's third law. Solid arrows represent experts' conceptual pathways while dashed-line arrows represent novices' reasoning.

magnitudes of forces [28,37]. The incorrect reasoning on magnitude can be mapped in the conceptual framework with a pathway from the related surface features through the intermediate dominance-based reasoning to the task outcome, which in this case is the thinking that the dominant object applies a larger force than the secondary object. Often a similar difficulty occurs when students must determine causality, in which case the dominant features are often used to determine the causal relation between the interaction forces. If the students have encountered the scenario previously, they may have a memorized rule within their intermediate reasoning and processes, which can be activated and lead to the memorized solution without further processing.

Additional student difficulties in N3L occur when students encounter two forces that are equal and opposite but not necessarily an interaction pair [31,66]. This often manifests itself with students applying an equal-and-opposite rule: because two forces are equal and opposite, then they must be the action and reaction forces in N3L. Students tend to connect any given forces to this rule. For example, in the case of a stationary box in a horizontal surface, the gravitational force and the normal force applied on the box are indeed equal and opposite. As a result, students are likely to consider these two forces as an interaction pair. This type of student reasoning reveals the typical pathway of connecting surface features to a memorized rule without distinction, which then leads to incorrect outcomes. This equal-and-opposite rule is also related to the equilibrium rule that many students hold [31], where students believe that N3L requires an object to be in equilibrium. If students perceive that an object will move in a scenario, then they may assume that N3L does not apply. The conceptual framework can map this type of pathway from surface features of the question to an intermediate reasoning focusing on what motion occurs after the forces.

The novice students' pathways generally focus on the levels of contexts and specific rules in the conceptual framework, with minimal connections to the central idea. Commonly, the interaction idea is completely absent from these students' reasoning. As students advance to intermediate and expert levels, they develop more integrated knowledge structures, expanding connections to elements of the central idea. The conceptual framework can then aid in showing these differences between novice, intermediate, and expert students by comparing their knowledge structures, conceptual pathways, and problem-solving behaviors.

C. Levels of knowledge integration within student knowledge structures

Using the conceptual framework, student misconceptions and difficulties documented in the literature can be represented and interpreted with thinking pathways of specific learning dynamics and states for analysis of students' levels of knowledge integration. In previous research, a number of developmental levels that encompass the range of student knowledge states were identified through large scale testing and interview studies on light interference and momentum [21,24]. Differences between the levels were found to be directly related to performance and reasoning on typical and atypical questions used in the assessments. The levels defined in this study are based on conceptual framework theory and years of teaching experiences. Summaries of the three levels of students and their problem-solving behaviors in N3L are discussed below:

Novice level.—The knowledge structures of novice students are typically fragmented with only local connections among surface features as the means to solve problems, leading to poor performance on both typical and atypical questions. These students base their answers on intuitive understanding of the real world and memorization of learned examples and often use contextual variables to cue their memory of likely answers without meaningful reasoning connecting to other related conceptual components or ideas. These novices can then only correctly answer a limited set of familiar questions, which they previously encountered. Some of these students struggle on even simple typical questions because they have not established either the central idea or memorized links between typical questions and results. These students often directly link surface features to their responses without meaningful reasoning or are relegated to guessing the answer when no memorized examples are applicable.

When working with typical contexts, such as questions asking students to compare the magnitude of forces in an interaction, novice students generally rely on previously memorized results or their intuitively developed naïve understandings such as those based on the dominance principle [43,67,68]. These students often focus on dominant surfaces features, such as greater mass or velocity, which students directly link to the outcome that the dominant object applies the greater force and/or causes the forces. However, when students apply a dominance principle, they likely do not link the scenario to the basic concept of N3L and fail to make connections to the N3L central idea.

On the other hand, when working with atypical questions, such as asking students to identify pairs of interaction forces, the novice students often demonstrate further weaknesses and fragmentation within their knowledge structures. These students typically place an overreliance on the equal-and-opposite rule where any two forces that are equal in magnitude and opposite in direction are likely identified as an interaction pair [31]. This rule is easily memorized and can be directly related to common statements of N3L. Many of the difficulties that novice students encounter with N3L can be traced back to this memorized rule. Students at this level simply do not have a connected knowledge structure built to reason these types of questions.

Additionally, these students do not connect other important properties of N3L to their knowledge structures. Frequently the conceptual components of time, type, and cause may be missing because they are never explicitly discussed in traditional instruction. Students who do relate the cause component to N3L may directly connect it to the action or dominant features and attribute such features as the cause of the interaction forces. These naïve views are often strongly embedded within students' experiences, which can lead to substantial difficulties in learning N3L.

Intermediate level.—Students at this level can engage in a deeper and moderately more extended level of reasoning to develop more connected understanding with the contextual variables than the novice students; however, these students still tend to rely on memorized examples and procedures to aid their problem solving but to a lesser extent than the novice students. The increased integration within their knowledge structures allows students to think about the central idea in limited common textbooklike problems. However, with only weak understanding of the central idea, students often fail on the atypical questions. Students at this level usually exhibit diverse, rich behaviors that can lead to multiple sublevels.

In the N3L conceptual framework, students generally at this level would have greater success when asked to compare magnitudes of forces in an interaction. In a range of question scenarios, they would be able to directly link the problem to the central idea without heavy reliance on memorization or focus on surface features. Weaker students at this level link the surface features directly to the "symmetric forces" property, while more advanced students link first to the central idea (i.e., first to "interaction pair" and then to symmetric forces) before reaching their final responses.

While reasoning in typical questions is improved, the lack of a solid understanding of the central idea often leads to wide-ranging difficulties on the atypical questions. Some of these students would attempt to use elements of the central idea but often apply them inconsistently. Many of these students still rely on the memorized equal-and-opposite rule. However, the more advanced students at this level reveal less reliance on the equal-and-opposite rule and begin to follow thought processes similar to those of the experts, especially on typical questions. This can result in a smaller gap between performances on typical and atypical questions.

As will be shown in a later section of this paper, intermediate students still do not necessarily integrate time, type, and cause into their N3L knowledge structures, but they generally perform well on time and type questions. However, most of these students do not completely understand causality in interactions, and causal reasoning is still often linked directly to dominant surface features within their knowledge structures. Overall, the intermediate students have more developed and integrated knowledge structures than novices. This allows a more consistent application of the central idea in typical questions and begins to allow the use of the central idea in limited atypical questions. Further improvement in reasoning on atypical questions would allow students to approach expertlike understanding, as described next.

Expert-like: Students with expert-like understanding can apply the central idea when answering both typical and atypical questions, signifying a robust, well-networked knowledge structure. This allows them to relate contextual variables to the central idea, along with many intermediate processes and related concepts, to form a comprehensive web of connections that can address a wide range of familiar and novel contexts.

For N3L, this means that these students can recognize that forces occur as the measures of interactions between two objects and that these forces are observed to be symmetric and occur simultaneously. Furthermore, they recognize that the interaction forces are always the same type and occur on two distinct but interacting objects with the same magnitude and opposite directions, and neither force can be the cause of the other force. This allows the student to use the central idea in virtually any scenario. Being able to apply the central idea of a concept uniformly across multiple typical and atypical contexts is the hallmark of an integrated knowledge structure and diverges from the novices' fragmented structures where reasoning pathways are only locally connected within limited contexts and with little transferability.

In summary, these three levels show a general progression of knowledge integration and reveal how common misconceptions manifest themselves using the N3L conceptual framework. Assigning students to these levels would usually require careful interviews to determine reasoning patterns and then matching elements of these patterns to the conceptual framework. To facilitate practical assessment, the next study examines the development of a multiple-choice assessment which will be able to quantitatively measure students' levels of knowledge integration.

III. STUDY 2: DEVELOPMENT OF CONCEPTUAL FRAMEWORK ASSESSMENT FOR NEWTON'S THIRD LAW

Recent studies have demonstrated the effectiveness of conceptual-framework-based assessments on measuring knowledge integration and deep understanding [21,23,24]. These assessments take advantage of the differences in problem-solving strategies students employ when solving typical and atypical problems that can be directly linked to levels of knowledge integration. In this study, an assessment containing 14 multiple choice questions was developed to probe knowledge integration in N3L. Using this assessment, quantitative data were collected to analyze the general categories of students' knowledge structures. In addition, interviews were conducted to explore students' thought processes and reasoning pathways for determining the fine-grained levels of knowledge integration.

A. Designing the conceptual framework-based assessment

Designing an assessment based on a conceptual framework starts with identifying multiple contexts suitable to probe connections among different conceptual elements and the central idea of the concept. These contexts are in the form of a mixture of both typical and atypical questions for the level of instruction targeted. Typical questions are those that are commonly used in instruction as examples, demonstrations, or homework problems. On the other hand, atypical questions are designed to engage the central idea, which often make use of unfamiliar contexts that students rarely encounter within the traditional curriculum. For example, a N3L question involving an electric force between two stationary charged particles would be atypical in the normal first semester introductory mechanics courses, which rarely mention electric forces. However, this question would be typical for second semester (or later) courses after students have been exposed to these forces. Furthermore, towards the properties of the concept to be probed, most typical N3L questions ask students to compare the magnitudes of two forces in a specific interaction. Meanwhile, atypical questions may ask for the cause of forces or ask to identify interaction pairs under more complex conditions. Finally, it must be kept in mind that the identification of a context as typical or atypical is not universal; the background of the population determines the typicality, which may change depending on the material taught to the students at a specific institution. In this study, the typicality is defined based on the question contexts considered typical for U.S. students enrolled in a first semester calculus-based physics course.

Using the conceptual framework for N3L, a test containing 14 multiple-choice questions was developed to assess students' understanding of the N3L concept and determine their levels of knowledge integration (see the Supplemental Material [69] for the test). The set of typical questions were adapted from a Newton's third law survey by Bao, Hogg, and Zollman [43] and questions commonly encountered in introductory physics courses. The atypical questions were created to further probe student understanding of the additional N3L properties that are not commonly addressed in current instruction. The contexts of the questions were designed around six properties of N3L:

1. Magnitude questions ask students to compare the magnitudes of two interaction forces. These are the

most common questions students encounter involving N3L and are therefore generally considered typical.

2. Interaction pair questions ask students to identify the two forces constituting an interaction pair (or action-reaction pair). These questions can be considered typical in simple cases but are considered atypical with the questions in this study, because the question design elicits interference with the equal-and-opposite rule and is then innately more difficult without careful application of the central idea.
3. Cause questions ask students to assess if either of the indicated forces cause the other. These are atypical in that only rarely do lectures or textbooks mention causality in relation to N3L.
4. Time questions ask students if either of the indicated forces occurs before the other force. This property was never directly featured in coursework and was initially considered atypical from the instructor's perspective. However, as indicated through interviews, real-life experience with forces often leads to many students realizing that forces in these scenarios occur at the same time. Therefore, these questions were included in the typical category based on students' pre-conception, since most students had intuitively developed a correct understanding of this property.
5. Direction questions ask students the relative directions of the two interaction forces. These are considered typical as the statement of "opposite directions" is directly given in all textbook versions of the N3L definition.
6. Type questions ask students if the forces are of the same type. While students generally have minimal explicit instruction on types of forces in current curriculum, these appear typical because students often only encounter a single kind of force in most N3L questions.

These six conceptual properties were blended within four N3L scenarios to form the 14 questions. The four scenarios are discussed below and summarized in Table I:

1. Collision between two soccer players running with equal speeds but unequal masses: Typical questions ask students to compare magnitude and timing of the forces exerted by each player on the other, while an atypical question asks students to compare the cause of each force.
2. Book at rest on a horizontal table asking to compare the weight of the book with the supporting force applied by the table on the book: Typical questions ask students to compare the magnitude of each force, while an atypical question compares the cause of each. Additionally, a second atypical question asks students if the two forces are an interaction (or action-reaction) pair, which usually has poor performance even after instruction. This specific interaction pair question was chosen because it cues the equal-and-opposite rule and is generally much more difficult than in other question scenarios. In addition, for all classes involved in this study, weight was explicitly defined as the gravitational force. However, students may still interpret weight as the force pushing down on the floor, which is considered as a misconception held by students. The pushing-down force was clearly defined during instruction as the normal force exerted by the object on the floor. It is also noted that in certain textbooks weight may be defined as the result of weighing an object (see examples in Morrison [70]). In such cases, the assessment questions need to be adjusted to explicitly compare the gravitational force and the supporting force.
3. Skater at rest pushing a second skater at rest: Typical questions ask students to compare the magnitude and timing of the force applied by each skater, while an atypical question asks students about the causal relation between the two forces.
4. Two bar magnets repelling each other while sitting on a table: Typical questions ask students to compare the magnitude, direction, and type of force exerted by each magnet, while the atypical question asks students to compare the cause of each force.

TABLE I. Question contexts defined by scenario and N3L properties probed.

Question scenarios	Properties
Collision with soccer players	Magnitude, time, cause
Book sitting on a table	Magnitude, cause, type, pair
Skater pushing the other skater	Magnitude, time, cause
Bar magnet force at a distance	Magnitude, direction, type, cause

The assessment was created carefully avoiding the action-reaction language (except in the book question) which might cue students' memorization of this language, which may interfere their reasoning in comparing magnitudes or attributing causality of the interaction forces. As such, no question explicitly asked students to identify forces as the action or reaction forces. However, the connections between the action-reaction language and other properties, such as magnitude and cause, were examined qualitatively during interviews.

Looking at the complete assessment, the typical magnitude property and atypical cause property were probed for each scenario, while the other properties were only probed in a subset of the scenarios. This mix of typical and atypical questions allows for the measurement of the levels of knowledge integration discussed in study 1, a process demonstrated in previous research [21,24].

B. Student testing and interviews

Pretests and post-tests using the developed assessment were given to students enrolled in three sections of a first semester calculus-based physics course at a large public Midwestern state university. Traditional instruction was used in two sections taught by one professor (control group) and a conceptual-framework-based intervention was used in the third section taught by a second professor (intervention group). Intervention methods and results are detailed in study 3.

A total of 476 students took the pretest during the first week of the course, with 164 students in the intervention group and 312 students in the control group. These students were predominantly first-year engineering majors. The N3L content was taught during the third week of the course, and a post-test was conducted three weeks after the instruction on N3L. A total of 153 students from the intervention group and 307 students from the control group took the post-test.

Four versions of tests were used in this study. Version A consisted of all questions and was used in the intervention group due to its smaller sample size. Versions B, C, and D were shorter versions, each containing questions on three out of the four scenarios, which were designed to accommodate for time constraints on testing. These short versions were randomly given to students in the control group. Because of the larger sample size of the control group, the randomized use of multiple short versions was able to produce statistically equivalent assessment on the different properties of the concept, which is a method shown to be practical in previous studies [71,72]. Pairwise t tests show no significant differences in scores to questions across the three versions used in the control group ($p > 0.2$). Among the short versions, version B omitted the collision scenario, version C omitted the skater scenario, and version D omitted the magnet scenario. Most properties were covered in multiple scenarios such that all students encountered questions related to the magnitude, cause, time, and type of force. Additionally, the book scenario was contained in all versions so that every student would encounter the interaction pair question. However, the direction question was only used in the magnet scenario leading to a smaller, but sufficiently large, number of the control group students encountering this property.

Interviews of undergraduate students were conducted after the post-test with 20 volunteers from both the intervention ($N_{\text{int}} = 14$) and control groups ($N_{\text{con}} = 6$). The student volunteers represented the whole ability distribution (5 expert level, 11 intermediate level, and 4 novice level). The purpose of the interviews was to identify the reasoning patterns students used to answer questions and to determine which links in the conceptual framework were being used. During the interviews, students were asked to review the assessment and explain their answers in think-aloud mode. In addition to the assessment questions, students were asked to describe their understanding of N3L, identify the interaction forces, and to declare which forces were the “action” and “reaction” forces in each scenario. Thoughts on connections between the action-reaction language and other properties of N3L were explicitly probed through these interviews.

Additionally, 22 physics graduate students at the same university volunteered to be interviewed about their understanding of N3L. They were asked to first describe and explain N3L before being presented with each of the four scenarios sequentially. Each student was asked to describe forces acting between the objects in each scenario. Specifically, they were asked to identify the interaction forces; the action and reaction forces; the magnitude, direction, timing, and type of forces; and the cause of each force. At the conclusion of each interview, students were asked to summarize their understanding of N3L, especially regarding action or reaction forces and the cause of each force. For all interviews, each session lasted approximately 20 min, and students were each given a gift card as a reward for their participation.

Statistical analysis was performed using t tests to compare students’ mean scores on each dimension between the levels of knowledge integration. This helps determine the differences and similarities in conceptual understanding between different levels of knowledge integration.

C. Study 2 results

As shown in previous research, students’ levels of knowledge integration can be revealed from the gap between scores on typical and atypical questions [21,24]. In this study, the typical questions probe students’ understanding on magnitude, time, type, and direction, while the atypical questions measure interaction pair and causes of interaction forces. Figure 2 shows the score distribution for the property dimensions, with the time, type, and direction combined due to similar performances. Histogram of frequency of total score is displayed in the background to show the distribution of students across the different performance levels. These results are calculated with pre- and post-test data from all students to increase the sample size so that a more stable distribution can be obtained.

As shown in Fig. 2, scores on typical and atypical questions are similarly low for all students with low total scores (score < 0.35), indicating a novice level understanding that leads to poor performances on both types of questions. As the total score increases to low-to-medium range (score $= 0.35 \sim 0.60$), a gap is more pronounced, suggesting that students at this range have begun to perform well on typical questions using memorization but without establishing a deep understanding. For mid-to-high total score (score $= 0.60 \sim 0.85$), the typical question performance is near mastery and students’ performance on atypical questions starts to show significant advancement. This indicates that students have developed partially

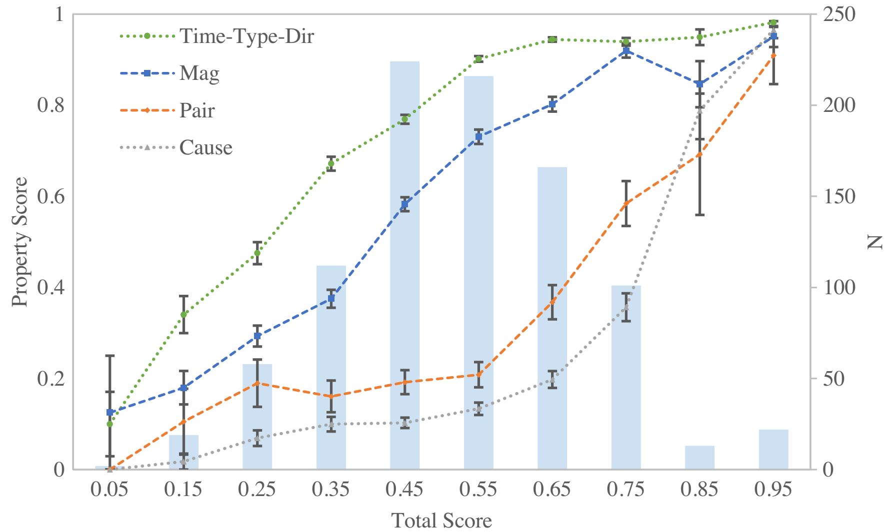

FIG. 2. Plot of property subscores across total scores (with error bars denoting standard error) for all students in this study. Frequency of total score distribution is shown as a bar chart in the background, with absolute count of students falling into each range shown on the right axis. The results are based on pre- and post-test data from all students.

integrated knowledge structures that allow them to apply their knowledge in some unfamiliar situations. Finally, the highest scoring students demonstrate no difference between their scores on typical and atypical questions, which implies that they have achieved a deep understanding with a well-integrated knowledge structure. These patterns of scores on different questions reveal a general proregression of student knowledge integration. To gain insight into the actual reasoning pathways of students at different performance levels, interview results are analyzed and discussed next along with the assessment outcomes.

From Fig. 2, it is clear that the time-type-direction properties resulted in the highest scores, while the cause property resulted in the lowest scores. These two dimensions, except for direction of forces, were not taught as part of normal physics courses and instead reflected intuitive understanding of the students. Specifically, the time and cause properties revealed a disconnect between their understanding of causality and the fundamental aspect of time order of causal events. Hence, students attributed cause to one of the forces even though they answered that the two forces occurred at the same time. When interviewed, most undergraduate students either did not directly link, or only weakly linked, time and cause, responding with comments such as, "*I don't think one could happen before the other*," or "*each happens at the same time…but I think the cause must be the action force*." When asked if cause and effect could occur at the same time, most students responded that they can. A few students with higher total scores recognized the condition between cause and time in their responses: "*Collisions happen at the same time so one [force] can't cause the other" and "they've got to happen at the same time…but I thought cause and effect had to be different times…but one of these has to cause the other, right?*" Overall, it is clear that students generally did not correctly link cause and time within their understanding of causality in the context of N3L.

All undergraduate students interviewed described N3L in a form similar to "for every action force there is an equal and opposite reaction force," and explained it as "if I push on something, it will push back." Only three students explicitly mentioned the concept of interaction between the two objects. When further asked if generally one of the forces caused the other, most students linked action directly to cause and reaction directly to effect. This demonstrates the connection that students have between the action-reaction language and cause and effect. Half of these students then suggested that the action force would be stronger than the reaction force, indicating the possible thinking of the dominance principle. The belief in an action force being the cause and having a larger magnitude appears to be most pronounced with students at lower levels of knowledge integration. Furthermore, the use of the equal-and-opposite rule has been found to be applied by most students with novice and intermediate levels of knowledge integration. These will be discussed with regards to each of the knowledge integration levels below.

Moving forward, comparison of the changes in score gaps on typical and atypical questions to students' total scores allowed a detailed exploration into levels of knowledge integration. Study 1 suggested the existence of three general levels, but here the intermediate level can be split into an upper and a lower category based on behaviors noted in the interviews and differences between the typical and atypical question performances. A total of four knowledge integration levels can be categorized, which are shown

TABLE II. Mean property scores per knowledge integration level. Standard errors are given in brackets. The results are based on pre- and post-test data from all students.

Knowledge integration level	Total score	N	Time-type-direction		Magnitude		Pair		Cause	
Novice	0.00–0.35	80	0.42	(0.03)	0.26	(0.02)	0.16	(0.04)	0.05	(0.01)
Intermediate-lower	0.35–0.60	467	0.77	(0.01)	0.56	(0.01)	0.19	(0.02)	0.11	(0.01)
Intermediate-upper	0.60–0.85	271	0.93	(0.01)	0.85	(0.01)	0.45	(0.03)	0.26	(0.02)
Expertlike	0.85–1.00	20	0.96	(0.02)	0.88	(0.04)	0.75	(0.10)	0.89	(0.04)

in Table II. The frequency of rule usage and performance of students within each property in the different levels are summarized in Table III.

Novice level (total score <0.35).—Students performed poorly on both typical and atypical questions. Magnitude, interaction pair, and cause properties all resulted in much lower mean scores for these students. However, students performed moderately better on time, type, and direction. When answering questions, these students relied heavily on memorized rules or related real-world intuition. Additionally, these students exhibited little to no use of the central idea, and instead directly related elements of the surface features to their responses. Students who exhibited this behavior had similar thoughts to interview excerpts shown below:

> *Student A: (response to soccer collision questions) "The heavier player, I think, applies the bigger force because $F = ma$ so bigger mass must mean a bigger force. Maybe if he pushes hard enough, he doesn't get pushed back. Uhh…I'm not sure what the cause is but I guess the bigger person must cause the collision so that causes the forces. I'd guess they'd happen at the same time." (response to book on table questions) "The force on the book by the table and the force of gravity of the weight of the book have to be equal and opposite or else…I think…the book would fall through the table maybe. Since they're equal and opposite, they must be the action and reaction forces, probably."*
> *Student B: (response to the skater questions) "I think we're supposed to use Newton's third law but I feel like Amy will push harder than the other person. No wait, they must be the same because we talked about this in class. They are equal and opposite. … Amy causes the other force because she pushes first. The forces still happen at the same time…well, maybe…there might be a super tiny time between the forces but I don't know. I know Amy does the action force for sure…the reaction force just kinda happens then and might be Jane falling down…so uhm, the pair of forces would be Amy pushing and gravity?"*

As also shown from the interviews, when thinking about the magnitudes of interaction forces, students at this level often relied on the dominance principle, considering the force by a larger body or the "action" party being stronger than the other force. Student A related the mass of the heavier player directly to the larger force through the known $F = ma$ equation but without understanding conceptually how it would apply in this situation. Meanwhile, student B demonstrated that while the student's own reasoning might lead to an incorrect magnitude response, vaguely remembering what was taught in class was able to produce the correct response in this typical context. Both students were inconsistent in their reasoning on the magnitude questions and focused on dominant features or attempting to use memorized solutions.

On questions to identify the interaction pairs, both student A and B exhibited minimal ability in the book question and encountered similar difficulties in other scenarios. They frequently thought interaction forces would act on the same object (e.g., gravity and normal force in the book question). One of the most common incorrect links these students held is the memorized equal-and-opposite rule, which was applied universally by these students in all contexts. However, not all novice students held this belief;

TABLE III. Summary of students' use of rules and performance on N3L properties separated by the four levels of knowledge integration. For rule use frequency, students at each level are categorized as using the rule always, often, sometimes, or rarely. Performance on N3L properties is then categorized as low, medium, or high. Rules summarized here include use of the equal-and-opposite rule, directly relating action-reaction to cause, and reliance on a dominant feature to determine magnitude.

	Rule use frequency			Performance			
Knowledge integration level	Equal-opposite	Action-reaction	Dominance principle	Causal	Pair	Magnitude	Time-type-dir
Novice	Always	Always	Always	Low	Low	Low	Medium
Intermediate-lower	Often	Always	Sometimes	Low	Low	Medium	High
Intermediate-upper	Sometimes	Often	Rarely	Medium	Medium	High	High
Expertlike	Rarely	Rarely	Rarely	High	High	High	High

for example, student B did not demonstrate even a rudimentary understanding of the interaction pair in the pushing question. Instead, the student chose two forces acting on the same person including the gravity, which was not even mentioned in the question. Another student thought the interaction forces in the collision question were the sum of the force of collision acting on each person and each person's weight.

Multiple additional students at this level also directly stated that "*the action force causes the reaction force*." Each of these students explained that the action force is stronger and that "*the stronger force produces (causes) a weaker force back at it*." Two students explained that Newton's third law only applies when the forces are "*equal and opposite*" but not "*when one of the forces is clearly stronger…like when a big truck crashes and crushes a car or when Amy pushes that other skater*."

Although the novice students showed reasonable performance on time and type questions, there is no evidence showing that these students considered them related to N3L. In follow-up discussions, most students demonstrated confusion, asking "*why are you asking about timing and force type? The other questions seem like you're asking about Newton's third law*." Additionally, they would often seem uncertain on their answers to these questions with responses such as, "*I guess they'd push at the same time or one might move before the other could push and then how would…they wouldn't touch and then…they couldn't push?*" or "*What does 'type' of force mean…they're both running into each other so that seems like the same type or whatever*." The novice students offered a variety of reasonings, but each seemed to rely on real world experience and not any element of N3L.

Based on the results from assessment and interviews, it is clear that novice students' knowledge is fragmented with local connections among surface features leading to the reliance on memorized rules and solutions as the means to solve problems. This leads to poor performances on both typical and atypical questions, especially when memorized rules, such as the equal-and-opposite rule, are applied universally leading to incorrect responses.

Overall, the novice students demonstrated weak performances on most typical and atypical questions, except for the time-type-direction questions, which showed moderate performances. The poor performance can be linked to students' fragmented knowledge structures and minimal understanding on the central idea. Answers they chose generally fell into categories of memorization of rules and problem solutions, or guesswork.

Intermediate level (total score between 0.35 and 0.85).—These students exhibited a range of behaviors, but all show significantly enhanced performances when compared to novices. The diverse student behaviors in this level allowed a finer grouping into a *lower intermediate level* (total score between 0.35 and 0.60) and an *upper intermediate level* (total score between 0.60 and 0.85).

As shown in Table II, students in the lower intermediate level demonstrated superior performance on the typical magnitude ($S_{\text{mag}} = 0.56$) and time-type-direction questions ($S_{\text{ttd}} = 0.77$) when compared to novices (0.26 and 0.42, respectively) [$t_{\text{mag}}(545) = 10.32$, $p < 0.001$, $d = 1.25$; $t_{\text{ttd}}(545) = 13.29$, $p < 0.001$, $d = 1.61$]. Little or no differences were noted on the atypical interaction pair ($S_{\text{int}} = 0.19$) or cause questions ($S_{\text{cause}} = 0.11$). Generally, students in this level exhibited a mixture of memorization of rules and solutions and evidence of some limited reasoning using the central idea.

Often these students demonstrated inconsistent reasoning depending on the contexts of the questions. For example, student C was able to easily respond that the magnitudes were equal in the soccer questions, but believed that the person actively pushing would apply a greater force in the skater questions:

> *Student C: (response to soccer questions) "The forces have to be equal and opposite…they always have to be because that's what Newton's third law said…"*
> *(response to skater questions) "Amy definitely applies the greater force I think. She has to since she's…uhh, maybe…the only one acting. She feels a little force back but Jane feels more."*

The question scenario then directly affected reasoning, with student C demonstrating at least partial reasoning using the central idea on the collision question but a reliance on surface features on the skater question. Overall, the lower intermediate students' reasoning on magnitude questions was significantly improved when compared to novices, with some students being able to answer all magnitude questions.

Reasoning on the atypical interaction pair questions demonstrated a range of varied abilities, with some of these students attempting to use elements of the central idea. However, applications of the central idea to atypical contexts were generally more haphazard and students were prone to instead relying on intuition based on contextual features:

> *Student D: (response to soccer questions) "The action and reaction forces must be the forces between the people running into each other…I don't know which one would be the action or reaction…they happen at the same time so maybe it doesn't matter which one is the action. Either way, they're equal and opposite so that's all we need for them to be that."*
> *(response to book questions) "The weight and normal force of the table are the action and reaction. The weight has to be the action and causes the reaction…without the book, I think there wouldn't be a normal force. Again they're equal and opposite and that's all Newton's third*

law says we need for the…uhhh…action-reaction pair….or I think my professor called them interaction forces or something but…"

Student D and other similar students demonstrated reliance on the equal-and-opposite rule for identifying the interaction pair in the book question without recognizing that the two interaction forces cannot both act on the same object, similar to novices. During the interview, student D was further presented with a similar question set up to the book question where all forces were explicitly stated, including the normal force applied by the book on the table, and still responded that the weight and normal force would have to be the interaction pair. Further discussion led to the student realizing that if the table was slanted, the forces would no longer be equal and opposite, but the student was unable to identify a pair of interaction forces, stating that "*maybe Newton's third law doesn't happen here then.*"

Students at this level also tended to attribute the action force as the cause of the reaction force. However, these students appeared less likely than novices to directly link the action force to a larger magnitude force, with some stating, "*the action force causes the other force, but action forces don't have to be bigger…at least not always.*" This indicates changes in connections within student knowledge structures with weaker links between the action-reaction language and the dominance principle. However, for these students, the "action" forces are still determined based on surface features of action and some level of dominance, and the "action" force is often believed to cause the corresponding "reaction" force.

Overall, students at the lower intermediate level demonstrated a moderate performance on the typical magnitude questions and a strong performance on the time-type-direction questions. The equal-and opposite rule appears to be heavily used by students, which helps to determine the correct directions of the interaction forces. Meanwhile, students' reasoning for determining the magnitudes of interaction forces is still influenced by the dominance principle, leading to weaker performances on magnitude questions than on the direction questions. It is worth noting that while the performances on magnitude and direction questions could be directly linked to the central idea, it appears that reasoning on time and type questions was intuitive and exists as fragments separate from their understanding of N3L. Nevertheless, the ability shown on the typical questions distinguishes these students from novices. Specifically, students' linking typical question reasoning to some elements of the central idea clearly separates their reasoning abilities from those of novices. However, students' performances on the atypical cause and interaction pair questions were at the same low level as novices, indicating that students at this level were unable to use the central idea when reasoning on atypical questions. This weakness is the key factor separating lower and upper intermediate students, as will be discussed next.

Upper intermediate students (total score between 0.60 and 0.85) then saw continued improvement, when compared to lower intermediate students, in the typical magnitude questions ($S_{\mathrm{mag}} = 0.85$) and time-type-direction questions ($S_{\mathrm{ttd}} = 0.93$) until they reached a level of mastery [$t_{\mathrm{mag}}(736) = 16.63$, $p < 0.001$, $d = 1.27$; $t_{\mathrm{ttd}}(736) = 11.95$, $p < 0.001$, $d = 0.91$]. Additional improvements were measured in the atypical interaction pair ($S_{\mathrm{pair}} = 0.45$) and cause ($S_{\mathrm{cause}} = 0.26$) questions [$t_{\mathrm{pair}}(736) = 8.09$, $p < 0.001$, $d = 0.62$; $t_{\mathrm{cause}}(736) = 8.85$, $p < 0.001$, $d = 0.68$]. From interviews, these students demonstrated further improved reasoning in the typical questions compared to the lower intermediate students. Their reasoning on the atypical properties of the action-reaction language and cause appeared similar to lower intermediate students, but the connections between these and magnitude were mostly nonexistent. The primary difference separating the upper and lower intermediate students was that the upper intermediate students begin to show more reasoning using the central idea in the atypical questions. In contrast, the lower intermediate and novice level students lacked the needed understanding of the central idea when thinking about atypical questions.

However, as shown by interviews, the upper intermediate students still had difficulty with the interaction pair question in the book on the table scenario where students considered the weight of the book and normal force from the table as an interaction pair. What is encouraging is that these students started to reason that "*the forces are equal and opposite because they are the action and reaction forces*" instead of the other way around as novices and lower intermediate students reasoned. This marked a noticeable change in students' thinking pathways in that they focused on the central idea first before its properties, indicating improved reasoning beyond the memorized equal-and-opposite rule. Still, most students at this level were unable to correctly apply the central idea in this question.

Overall, the intermediate students' performance and reasoning clearly separated them from behaviors of novice level students. Specifically, intermediate students started to apply the central idea in typical questions. Their use of the central idea in atypical questions was still inconsistent with strong context dependence on the question scenarios. The key factor discriminating intermediate students from expert-like understanding was the use of the central idea on atypical questions, as described next.

Expertlike (total score >0.85).—These students demonstrated near mastery in all typical and atypical questions with the most notable improvement over intermediate students appearing in the atypical interaction pair ($S_{\mathrm{int}} = 0.75$) and cause ($S_{\mathrm{cause}} = 0.89$) questions, which are significantly better than the upper intermediate students [$t_{\mathrm{pair}}(289) = 2.58$, $p = 0.01$, $d = 0.60$; $t_{\mathrm{cause}}(289) = 9.88$,

$p < 0.001$, $d = 2.29$]. Scores in magnitude ($S_{\mathrm{mag}} = 0.88$) and time-type-direction ($S_{\mathrm{ttd}} = 0.96$) questions were already at a high level of performance similar to the upper intermediate students. The expertlike understanding required a well-integrated knowledge structure such that students were able to consistently apply the central idea when answering typical and atypical questions, which is evident form the interview excerpts shown below.

> *Student E: (response after all questions) "All these questions were the same once I realized I just needed to use Newton's third law. Choosing the action-reaction forces was tough…mainly with the book though…I wanted to say the book and normal force were them, but I remembered, like, we needed the forces on different things but those two both were on the book."*
> *Student F: (response to magnet questions) "I have no idea how magnets work…I thought that stuff was next semester. I guess it's still force and maybe, uhhh…kinda like gravity but not…you know…magnets don't always pull together though. There's lots of forces here though…gravity, friction, normal forces, air resistance maybe. I think you're asking about the magnetic forces which I guess…oh I can use Newton's third law, right? So the forces must be equal and opposite…the left magnet applies a force on the right one and the right on the left one then….so those would be the action and reaction forces."*

From the interviews, students at this level were able to recognize that the magnitudes of interaction forces are equal and that the interaction forces must occur on two distinct objects. Both of these students focused their reasoning on the central idea, even in the atypical interaction pair property or the unfamiliar case of the force of magnets. The students directly related the questions to N3L and then to their answers. The ability to apply the central idea of a concept consistently across multiple typical and atypical contexts is the hallmark of an integrated knowledge structure, which diverges from the novice structure where reasoning pathways connected surface features to memorized procedures and responses.

Additionally, the interviews with the 22 physics graduate students revealed overall strong understanding of N3L but with a few commonalities to undergraduate students who had yet to master the material. Most (20/22) grad students initially used the words action and reaction when stating N3L. Only 10/22 students initially stated that the pair of interaction forces must act on two separate objects as part of their definition. 9/22 explicitly stated that action and reaction forces were inherently different from the action idea commonly associated with the dominance principle, and all students explained that magnitudes of forces were equal in every question. This suggests that the graduate students were minimally influenced by the dominance principle in their reasoning.

However, over half (12/22) grad students still thought that the weight and normal force in the book question were a pair of interaction forces. Of these 12 students, 10 noted that the symmetry of the weight and normal forces meant that they had to be the action-reaction pair, demonstrating a reliance on the equal-and-opposite rule exhibited by many undergraduate students. The results suggest that even graduate students may still lack a complete understanding of the central idea. Specifically, deciding which forces constitute an interaction pair reveals a persistent difficulty such that even some presumable expert-level students still fail to recognize that the interaction forces must occur on two different objects.

Overall, the assessments and interviews revealed widespread and persistent difficulties in students' understanding of N3L. Based on the results from concept test and interviews, a progression of four levels of knowledge integration can be identified, which is shown in Table III along with students' performances and rule usage. The results in Table III suggest that differences in two main aspects of thinking can help distinguish different progression levels: (i) the use of the memorized equal-and-opposite rule, and (ii) the association of action-reaction language with cause and effect and magnitude of the forces.

First, novice level students exhibited a high degree of dependence on the equal-and-opposite rule to determine interaction pairs. This dependence weakened in lower-intermediate students and was minimal in upper-intermediate and expertlike students. For students at the upper intermediate level, this equal-and-opposite rule was often replaced by the correct thinking pathway: because two forces are an interaction pair, then they must be equal and opposite. However, these students were not able to consistently apply this reasoning correctly in problem solving.

Second, novice students directly linked what they deem the action force to the dominant, causal, or stronger force. These connections between the action-reaction language, causality, magnitude, and other surface features represented intuitive understandings that significantly diverge from the central idea. Both the upper and lower intermediate level students still exhibited strong connections between the action-reaction language and causality but did not connect magnitude to the dominant features as strongly as novices. Expertlike students were able to connect all properties through the central idea regardless of the scenario or surface features presented.

In addition, most students, including some expertlike students, demonstrated a disconnect between their understanding of time and causality. They were able to respond that interaction forces occur simultaneously but did not realize that causality and time were in fact related.

With the knowledge integration levels defined in terms of the N3L conceptual framework, the focus of research can shift to how best to promote advancement of students from lower to higher levels of knowledge integration. This is explored using a conceptual framework-based intervention in the third study.

IV. STUDY 3: DEVELOPMENT OF CONCEPTUAL FRAMEWORK INTERVENTION FOR NEWTON'S THIRD LAW

The assessment studies have shown that students difficulties are likely related to their lack of understanding of the central idea of N3L. The results can be used to enhance instruction by explicitly introducing the central idea and emphasizing its connections to other elements of the knowledge structure using specifically designed demonstrations and examples. This intervention focuses on changes to content emphasis under the existing teaching environment, which is easier to manage in large lecture classes than changes to instructional style.

A. The conceptual framework-based intervention

The design of the teaching intervention makes emphasis on explicitly introducing the central idea in instruction. For the intervention group, the modified instruction was only implemented in teaching N3L, and the rest of the instruction was not conceptual framework based. Both the intervention and control classes used traditional lectures with multiple-choice clicker questions. The conceptual framework based intervention on N3L used clicker questions and guided discussions that emphasize the central idea. Specifically, the intervention first emphasizes that forces are caused by an interaction between two objects. Additional emphasis is placed on the interaction itself being the origin of the forces, not one force or object causing the force(s). This relation is particularly important in addressing the implied causality from the language of action-reaction forces. After emphasizing the interaction nature of forces, the second element of the central idea is introduced that these forces, which are measures of an interaction, are observed to be symmetric. The concept of symmetric interaction is then expanded to derive the set of properties that the interaction forces occur in pairs, act on two distinct interacting objects, are the same type of force, occur at the same time, are noncausal to each other, and are equal and opposite. These properties are then re-enforced using demonstrations and examples which require the application of the N3L central idea. This strategy aims to help students build their knowledge structures around the central idea and is summarized in Fig. 3.

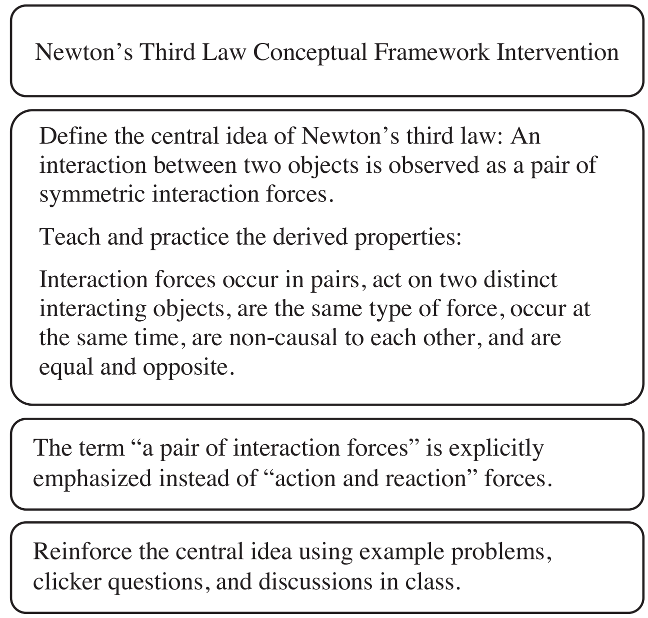

FIG. 3. Steps of the intervention employed in modified N3L lectures.

Key examples used in the intervention include identifying interaction forces in N3L scenarios such as collisions or objects pushing one another. Figure 4 shows two questions with a ball being dropped and moving through the ground. Both questions ask students to compare the force on the ball by the ground to the force on the ground by the ball. The contextual features are varied. In one question, the ball is stopped after traveling a distance into the ground, while in the other question, the ball is still traveling into the ground. In these, and other N3L scenarios, it is important that students learn to focus their attention on the interaction itself instead of the surface features of the problems such as moving or with different masses. Further questions involve asking students to identify what interaction pairs exist in the questions.

After introducing the central idea and connecting it to the different properties, it is important to relate the conceptual framework to the traditional statement of N3L using the action-reaction language. Students are likely to have previous and/or future exposures to this language; therefore, drawing explicit connections between different existing and new knowledge components will further integrate students' understanding and aid in building their comfort and confidence in applying the central idea. These connections explicitly emphasize that labeling forces as action or reaction is a traditional language for interaction and does not imply one force causes the other as the two forces occur simultaneously and do not cause each other.

Participants in this study were the same in study 2, who were students enrolled in three sections of a first semester calculus-based physics course, and the assessment described in study 2 was used for pre-post testing following the procedure described in study 2. The intervention group consisted of one section of the course taught by a professor using the intervention content described here, while the control group consisted of two sections taught by a different professor using traditional content. Both professors had similar experience and style in teaching this same course. Statistical analysis was performed using t tests to compare students' changes in scores on each dimension

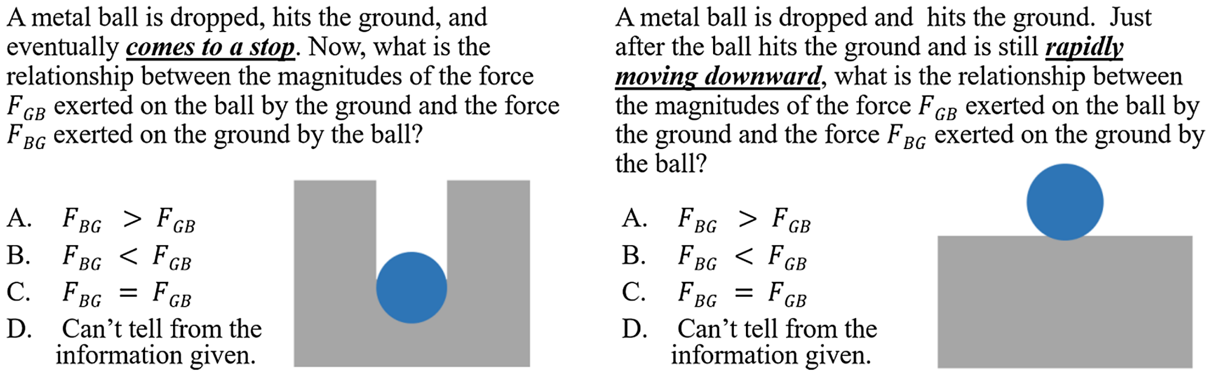

FIG. 4. A sample sequence of clicker questions for N3L used during the N3L intervention.

between control and intervention groups. This helps determine the differences and similarities in changes to conceptual understanding between the two groups.

B. Pretest results

First, a comparison between pre-test groups was performed to determine the homogeneity of the control and intervention groups prior to instruction. Pre-test results revealed minimal differences between the groups (see Table IV). The minor differences were not significant overall or for each of the conceptual properties. Additionally, students' mean scores were high on questions probing time (≥ 0.81), type of force (≥ 0.76), or direction (≥ 0.70). However, students' mean scores were low on questions probing cause of force (≤ 0.19) and identifying an interaction pair of forces (≤ 0.34). This indicates that students have a good level of prior knowledge or intuitive understanding with the time, type, and direction properties.

C. Comparison between pre- and post-test

While the two groups performed similarly on the pretest, their performances diverged after instruction on the post-test as shown in Fig. 5. For both groups, their overall scores improved, but the intervention group's improvement was significantly greater than the control group's with a large *Cohen's d* effect size [$t(458) = 7.47$, $p < 0.001$, $d = 0.74$]. Furthermore, distributions of students at the four knowledge integration levels reveal that intervention group students generally transitioned to higher levels, while there were fewer favorable transitions in the control group (see Table V). Only the intervention group saw a noticeable increase in students classified at the expertlike level, suggesting that it is difficult for students to achieve expertlike understanding without modified instruction.

To get a richer picture on how student understanding changes after instruction, student score changes in individual properties were also analyzed (see Fig. 5 and Table VI). t tests were conducted comparing the pre-post score changes between the intervention and control groups with results shown in Table VI. Differences in score changes were significant for the magnitude, pair, cause, and direction properties but not for the time or type properties which had high pretest scores. The effect sizes were moderate to large for the magnitude, pair, and cause properties but small for the direction property. Discussion of differences within each property will follow.

Figure 5(a) shows that overall, the intervention group had a significantly larger increase of their scores from pretest to post-test than the control group did [$t(458) = 7.47$, $p < 0.001$, $d = 0.74$]. Looking at specific properties, Figure 5(b) shows that a large improvement in the magnitude property was achieved by the intervention group when compared to the control group [$t(458) = 4.16$, $p < 0.001$, $d = 0.41$]. The improvements in the magnitude questions point towards improved reasoning and more integrated knowledge structures. The larger increase in performance in the intervention group suggests that these

TABLE IV. Comparison of pretest scores between the two groups. Differences were not significant for any conceptual property.

Conceptual	Intervention			Control						
Properties	N	Score	SE	N	Score	SE	t	df	p	d
Magnitude	164	0.60	(0.02)	312	0.58	(0.02)	−0.92	474	0.36	−0.09
Pair	164	0.34	(0.04)	312	0.28	(0.03)	−1.42	474	0.16	−0.14
Cause	164	0.19	(0.02)	312	0.16	(0.01)	−1.28	474	0.20	−0.12
Time	164	0.84	(0.02)	312	0.81	(0.02)	−1.21	474	0.23	−0.12
Type	164	0.76	(0.03)	312	0.76	(0.02)	0.19	474	0.85	0.02
Direction	164	0.70	(0.04)	133	0.78	(0.04)	1.57	295	0.12	0.18

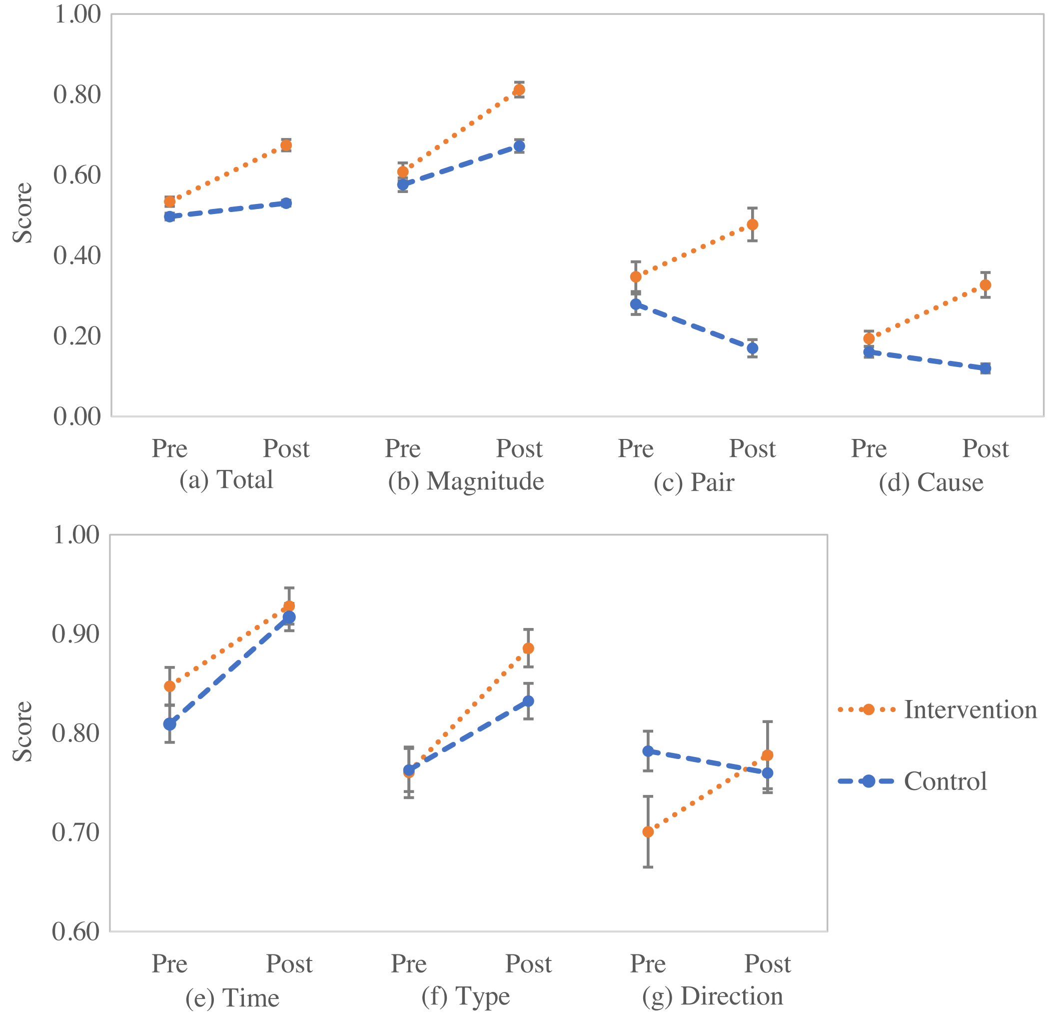

FIG. 5. Pre-post mean scores for the total score and each property: (a) Total, (b) magnitude, (c) cause, (d) time, (e) type, (f) interaction pair, and (g) direction. The error bars represent the standard errors of the means.

students were less influenced by the dominance principle in the typical magnitude questions, which gave students a better position to approach atypical questions.

Figure 5(c) shows that only the intervention group saw an increase on scores for the atypical N3L interaction pair question. The control group saw scores decreased on the post-test while the intervention group improved, resulting in a significant difference between the intervention and control groups [$t(458) = 5.68$, $p < 0.001$, $d = 0.56$]. The book question on interaction pair was designed with two forces which are equal and opposite but do not form an interaction pair. This design is to probe if students would use the equal-and-opposite rule without meaningful reasoning, which was the primary thinking pathway used by novice students during interviews. Improvement on this atypical question then demonstrates the increased connections to the central idea within students' knowledge structures, rather than the local connections focusing on the memorized rule.

Figure 5(d) shows that only the intervention group saw an increase in score on the cause property while the control group saw a decrease in score from pre- to post-test with t tests showing a large difference between the changes of the two groups [$t(458) = 7.19$, $p < 0.001$, $d = 0.71$]. For the control group students, most incorrect answers followed the expected dominance principle, attributing the dominant features as the causes. In contrast, the intervention group's incorrect answers were mostly "each force causes the other force", which is an incorrect understanding

TABLE V. Percentage of students in each group at each knowledge integration level on the pre- and post-test.

	Intervention		Control	
Knowledge integration level	Pre	Post	Pre	Post
Novice	7.4%	1.4%	18.1%	6.7%
Intermediate-lower	60.1%	31.0%	60.0%	61.9%
Intermediate-upper	31.9%	57.0%	21.1%	30.6%
Expertlike	0.6%	10.6%	0.8%	0.7%

TABLE VI. Comparison of changes in scores (ΔS) between pre- and post-tests for each group. Differences between the groups were significant for the overall score changes and the magnitude, interaction pair, and cause at the $p < 0.001$ level. Moderate to large effect sizes were measured for the magnitude, pair, and cause properties while a small effect size was noted for the direction property. N's are the counts of students answering questions with these properties on the post-test and SE is the pooled standard error between the pre- and post-tests for each group.

	Intervention			Control						
Property	N	ΔS	SE	N	ΔS	SE	t	df	p	d
Magnitude	153	0.21	(0.02)	307	0.10	(0.02)	4.16	458	<0.001	0.41
Pair	153	0.14	(0.04)	307	−0.11	(0.02)	5.68	458	<0.001	0.56
Cause	153	0.14	(0.03)	307	−0.04	(0.01)	7.19	458	<0.001	0.71
Time	153	0.08	(0.02)	307	0.11	(0.02)	−0.90	458	0.37	−0.09
Type	153	0.13	(0.02)	307	0.07	(0.02)	1.86	458	0.06	0.18
Direction	153	0.08	(0.03)	200	−0.02	(0.03)	2.04	351	0.04	0.22
Total	153	0.14	(0.01)	307	0.03	(0.01)	7.47	458	<0.001	0.74

of causality but demonstrates a stronger recognition of the symmetry in the interaction forces. This answer also shows the necessity for implementing explicit formal instruction on causality in teaching physics, because, without such instruction, students will nevertheless develop their own versions of causal relations, which are often incorrect and can lead to more persistent difficulties in their understanding. Score changes on the atypical cause questions further demonstrate that the intervention appears to have improved students' understanding of the central idea regarding the causality aspect more than the traditional instruction.

As shown in Fig. 5(e), students' scores on the time property were high on the pretest and improved to a similar degree for each group, with no significant difference measured in the gains [$t(458) = -0.90$, $p = 0.37$, $d = -0.09$]. However, although students generally were doing well answering the time questions, their understanding of the connection between time and causality was problematic. As revealed from interviews, most students lacked a rudimentary understanding of time order as a necessary condition for establishing causality. For example, students would claim one force being the cause of the other force while recognizing the two forces were occurring simultaneously. Only three interviewed students directly stated that cause and time were related to each other. The results suggest that most students did not recognize the relation between time and causality and its relevance to the concept of N3L.

For the type and direction properties shown in Figs. 5(f) and 5(g), students were doing well on the pretest and the differences between the two groups were on the borderline of being significant, with small effect sizes. The results show that students were able to recognize the directions and types of interaction forces based on previous learning or intuitive understanding. However, many students in interviews did not recognize the relevance of the type of force to N3L, which indicates a lack of understanding of the nature of interaction forces. Therefore, future research is needed to further explore this property and its impact on students' learning of N3L, such as in the book on table question where many students considered the gravitational force and the normal force applied on the book as a pair of interaction forces. As discussed earlier, this line of study will also need to consider the specific of definition of weight used in the instruction.

From Fig. 5, it is clear that students in the intervention group showed significant improvements over their counterparts in the control group on their overall understanding as well as in most individual properties of N3L. This result aids in confirming the efficacy of using conceptual-framework-based instruction to improve student deep understanding and knowledge integration. Greater improvements were noted in the typical magnitude questions, and the atypical interaction pair and cause questions, suggesting a more uniform impact on both typical and atypical questions. This also led to more students in the intervention group being identified with higher levels of knowledge integration than students in the control group, as shown in Table IV. Furthermore, the greater improvement by the intervention group in the magnitude property, which is typically the focus of traditional N3L assessment and teaching, indicates that the intervention is also successful by traditional standards. Overall, this study demonstrates the benefits of expanding current N3L instruction to include elements of the N3L conceptual framework. Without these changes to instruction, it appears that students may lack the understanding of multiple key components of N3L.

V. CONCLUSIONS

In this study, a conceptual framework for Newton's third law was developed and applied to guide assessment and instruction of student learning. Based on assessment and interview results, students were found to fall into four levels of knowledge integration: novice, lower intermediate, upper intermediate, and expertlike. The levels of knowledge integration found here are similar to those found in previous research [21,24,73,74]. The reasoning behaviors of students at different levels reveal a progression of reasoning from a rudimentary surface level to deep understanding, summarized in Table III by their performance on a N3L conceptual question performance and frequency of rule use.

The novice students performed poorly on most questions, with reasoning closely linked to surface features and minimal connections to the central idea that interaction forces are observed measures of an interaction. In problem solving, these students often use memorized fragments of the concept which were directly linked to specific contextual features. Use of the memorized fragments left students unable to answer questions they were less familiar with. In particular, the use of the equal-and-opposite rule led students to think that any two equal and opposite forces would be a pair of interaction forces. These students also directly connected the action-reaction language to their thinking of a dominance principle, where the dominant action force would cause the other, weaker reaction force. Additionally, a strong disconnect was noted in novice students such that they had no knowledge of a relation between time and causality.

Lower intermediate students demonstrated greater performance than novices on typical questions but similarly poor performance on atypical questions. They were able to move beyond simple memorization of solutions to some typical questions and apply some components of the central idea, but these students still failed to apply the central idea to the atypical questions and instead relied on surface features and memorized rules. Specifically, these students applied the equal-and-opposite rule frequently and still connected the action-reaction language to their understanding of causality and magnitude. However, their performance increased in most typical questions suggesting that connections to some components of the central idea had been formed. Still, many of these students were unable to recognize the relation between time and causality, similar to novices.

Upper intermediate students, on the other hand, were able to apply the central idea in most typical and some atypical questions. Compared to novices and lower intermediate students, upper intermediate students demonstrated less reliance on the memorized equal-and-opposite rule when finding the interaction pair. Instead, they often were able to directly apply the central idea; however, they did not exhibit consistent reasoning in identifying interaction forces. Furthermore, these students sometimes still used surface features to name the action and reaction forces and link this directly to causality. As an improvement compared to novices and lower intermediate students, the upper intermediate students did not connect the action force to the magnitude of interaction forces. Additionally, some of these students begin to recognize the connection between timing and causality, but often still considered that the action force would cause the reaction force.

Expertlike students were able to successfully answer most questions with explicit usage of the N3L central idea and rarely demonstrated the use of memorized rules. They would correctly identify the interaction forces and knew these forces were symmetric. Furthermore, all elements of the central idea were understood and strongly connected so that even unfamiliar atypical questions would be properly handled. While some of the students did identify an action and reaction force based on surface features, these were activated as part of prior knowledge, which were then properly addressed and not used in reasoning to answer questions. Additionally, these students recognized the relation between timing and causality.

Graduate students, who would generally be assumed to be experts, demonstrated mostly expert-like knowledge structures, but with some missing the essential interaction portion of the central idea. This led to a subset of the graduate students exhibiting an upper intermediate level behavior with some use of the equal-and-opposite rule.

This progression of student understanding agrees well with the knowledge integration principles where novice students start with weakly connected knowledge structures focused on surface features and begin to construct a stronger network of connections until they reach an expert-like integrated knowledge structure, which is essential for obtaining deep conceptual understanding [17].

The conceptual-framework-based intervention explicitly introduced the central idea to students, that forces are the result of an interaction and that these forces are observed to be symmetric. Implementation of the intervention was found to yield significant improvements over traditional instruction in understanding the typical magnitude and direction questions and the atypical cause and interaction pair questions. Specifically, focus on the interaction nature of forces led to a decrease in students' application of the memorized equal-and-opposite rule and less reliance on surface features for determining causality and magnitude of forces. The development of a deeper conceptual understanding and higher levels of knowledge integration helped the students to improve on both typical and atypical questions. This research demonstrates the potential benefits when instruction is designed to enhance knowledge integration around the central idea of a concept. Consistent with previous studies on knowledge integration interventions [23,25], this study provides further evidence for the value of the conceptual framework model in guiding instruction.

Furthermore, students were shown to have a disconnect in their understanding of causality and time in that many students considered the forces being causal while occurring at the same time. In interviews, students directly attributed causality to what they believed was the action force, hinting at an implication of causality from the action-reaction language used to teach N3L. However, measurement of this link in large scale has not been conducted in this study due to limitations on assessment design, and is explored further in the study by Chen *et al.* [59].

Additional limitations of this study exist in that different professors taught the control and intervention groups, and multiple test versions were used. However, the large effects measured in the intervention study are not explained alone by these limitations and do not provide a serious threat to the internal validity of the study. Nevertheless, in future studies on finer aspects of N3L learning, improved controls in research design need to be carefully considered.

Moving forward, existing research has developed a wealth of curricular materials that address N3L through interactive engagement approaches. In this study, the conceptual framework method was implemented in the traditional lectures with clickers. It will be interesting and important to conduct controlled studies to compare the effectiveness of the conceptual framework method with the existing interactive engagement approaches, which should warrant a series of future studies for different controlled conditions and implementation styles.

In summary, this research expands the previous work on the use of conceptual frameworks to include Newton's third law. Assessments and interviews reveal the utility of the conceptual framework model to categorize students' levels of knowledge integration based on performance on typical and atypical questions. Specifically, levels of knowledge integration demonstrate distinct students' thinking pathways through their reasoning with the memorized equal-and-opposite rule, the action-reaction language, and the relation between time and causality. Meanwhile, intervention methods which focus on explicit introduction and use of the interaction nature of forces lead to increased assessment performance and higher levels of knowledge integration. Overall, the results demonstrate that the conceptual framework approach is promising in guiding assessment of knowledge integration in physics education and can facilitate the development of effective instruction for promoting deep conceptual learning.

ACKNOWLEDGMENTS

The research is supported in part by the National Science Foundation Grant No. DUE-1712238. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

[1] L. Bao and K. Koenig, Physics education research for 21st century learning, Discip. Interdiscip. Sci. Educ. Res. **1**, 2 (2019).

[2] M. Chiu, C. Guo, and D. F. Treagust, Assessing students' conceptual understanding in science: An introduction about a national project in Taiwan, Int. J. Sci. Educ. **29**, 379 (2007).

[3] E. Kim and S.-J. Pak, Students do not overcome conceptual difficulties after solving 1000 traditional problems, Am. J. Phys. **70**, 759 (2002).

[4] National Research Council, *Assessing 21st Century Skills: Summary of a Workshop* (National Academies Press, Washington, DC, 2011).

[5] National Research Council, A framework for K-12 science education: Practices, in *Crosscutting Concepts, and Core Ideas* (National Academies Press, Washington, DC, 2012).

[6] M. Alonso, Problem solving vs. conceptual understanding, Am. J. Phys. **60**, 777 (1992).

[7] D. Stamovlasis, G. Tsaparlis, C. Kamilatos, D. Papaoikonomou, and E. Zarotiadou, Conceptual understanding versus algorithmic problem solving: Further evidence from a national chemistry examination, Chem. Educ. Res. Pract. **6**, 104 (2005).

[8] E. Bagno, B.-S. Eylon, and U. Ganiel, From fragmented knowledge to a knowledge structure: Linking the domains of mechanics and electromagnetism, Am. J. Phys. **68**, S16 (2000).

[9] L. Bao and E. F. Redish, Model analysis: Representing and assessing the dynamics of student learning, Phys. Rev. ST Phys. Educ. Res. **2**, 010103 (2006).

[10] Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, Cogn. Sci. **5**, 121 (1981).

[11] J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, Science **208**, 1335 (1980).

[12] A. H. Schoenfeld and D. J. Herrmann, Problem perception and knowledge structure in expert and novice mathematical problem solvers, J. Exp. Psychol. Learn. Mem. Cogn. **8**, 484 (1982).

[13] J. L. Snyder, An investigation of the knowledge structures of experts, intermediates and novices in physics, Int. J. Sci. Educ. **22**, 979 (2000).

[14] A. B. Champagne, R. F. Gunstone, and L. E. Klopfer, A perspective on the differences between expert and novice performance in solving physics problems, Res. Sci. Educ. **12**, 71 (1982).

[15] Q. Chen, G. Zhu, Q. Liu, J. Han, Z. Fu, and L. Bao, Development of a multiple-choice problem-solving categorization test for assessment of student knowledge structure, Phys. Rev. Phys. Educ. Res. **16**, 020120 (2020).

[16] H. Freyhof, H. Gruber, and A. Ziegler, Expertise and hierarchical knowledge representation in chess, Psychol. Res. **54**, 32 (1992).

[17] M. C. Linn, The knowledge integration perspective on learning and instruction, in *The Cambridge Handbook of the Learning Sciences*, edited by R. Sawyer (Cambridge University Press, Cambridge, England, 2005), pp. 243–264.

[18] J. H. Larkin, Processing information for effective problem solving, Eng. Educ. **70**, 285 (1979), https://eric.ed.gov/?id=EJ213382.

[19] M. Kubsch, J. Nordine, K. Neumann, D. Fortus, and J. Krajcik, Measuring integrated knowledge—A network analytical approach, in *Rethinking Learning in the Digital Age: Making the Learning Sciences Count* (International Society of the Learning Sciences, London, UK, 2018).

[20] H.-S. Lee, O. L. Liu, and M. C. Linn, Validating measurement of knowledge integration in science using multiple-choice and explanation items, Appl. Meas. Educ. **24**, 115 (2011).

[21] R. Dai, J. Fritchman, Q. Liu, Y. Xiao, and L. Bao, Assessment of student understanding on light interference, Phys. Rev. Phys. Educ. Res. **15**, 020134 (2019).

[22] R. Duit and D. F. Treagust, Conceptual change: A powerful framework for improving science teaching and learning, Int. J. Sci. Educ. **25**, 671 (2003).

[23] Y. Nie, Y. Xiao, J. Fritchman, Q. Liu, J. Han, J. Xiong, and L. Bao, Teaching towards knowledge integration in learning force and motion, Int. J. Sci. Educ. **41**, 2271 (2019).

[24] W. Xu, Q. Liu, K. Koenig, J. Fritchman, J. Han, S. Pan, and L. Bao, Assessment of knowledge integration in student learning of momentum, Phys. Rev. Phys. Educ. Res. **16**, 010130 (2020).

[25] J. Nordine, J. Krajcik, and D. Fortus, Transforming energy instruction in middle school to support integrated understanding and future learning, Sci. Educ. **95**, 670 (2011).

[26] D. E. Brown and J. Clement, Misconceptions concerning Newton's law of action and reaction: The underestimated importance of the third law, in *Proceedings of the Second International Seminar: A Misconceptions and Educational Strategies in Science and Mechanics, Vol. 3* (Cornell University, Ithaca, NY, 1987), pp. 39–53.

[27] A. C. Graesser, D. Franceschetti, B. Gholson, and S. Craig, Learning Newtonian physics with conversational agents and interactive simulations, in *Developmental Cognitive Science Goes to School* (Routledge, New York, 2013), p. 157.

[28] C. Hellingman, Newton's third law revisited, Phys. Educ. **27**, 112 (1992).

[29] M. J. Hughes, How I misunderstood Newton's third law, Phys. Teach. **40**, 381 (2002).

[30] G. A. Lindsay, Newton's third law of motion as presented in textbooks of physics, Am. J. Phys. **11**, 319 (1943).

[31] D. J. Low and K. F. Wilson, The role of competing knowledge structures in undermining learning: Newton's second and third laws, Am. J. Phys. **85**, 54 (2017).

[32] A. Savinainen, A. Mäkynen, P. Nieminen, and J. Viiri, An intervention using an interaction diagram for teaching Newton's third law in upper secondary school, Physics Alive, in *Proceedings of the GIREP-EPEC 2011 Conference* (University of Jyväskylä, Jyväskylä, Finland, 2012), pp. 123–128.

[33] A. Savinainen, P. Scott, and J. Viiri, Using a bridging representation and social interactions to foster conceptual change: Designing and evaluating an instructional sequence for Newton's third law, Sci. Educ. **89**, 175 (2005).

[34] S. Stocklmayer, J. P. Rayner, and M. M. Gore, Changing the order of Newton's laws—why & how the third law should be first, Phys. Teach. **50**, 406 (2012).

[35] C. Terry and G. Jones, Alternative frameworks: Newton's third law and conceptual change, Eur. J. Sci. Educ. **8**, 291 (1986).

[36] M. Ventura, V. Shute, and Y. J. Kim, Assessment and learning of qualitative physics in Newton's playground, in *International Conference on Artificial Intelligence in Education* (Springer, New York, 2013), pp. 579–582.

[37] P. A. White, The causal asymmetry, Psychol. Rev. **113**, 132 (2006).

[38] P. A. White, Impressions of force in visual perception of collision events: A test of the causal asymmetry hypothesis, Psychon. Bull. Rev. **14**, 647 (2007).

[39] D. E. Brown, Students' concept of force: The importance of understanding Newton's third law, Phys. Educ. **24**, 353 (1989).

[40] J. Clement, Students' preconceptions in introductory mechanics, Am. J. Phys. **50**, 66 (1982).

[41] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

[42] I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, Am. J. Phys. **53**, 1043 (1985).

[43] L. Bao, K. Hogg, and D. Zollman, Model analysis of fine structures of student models: An example with Newton's third law, Am. J. Phys. **70**, 766 (2002).

[44] I. E. Gordon, R. H. Day, and E. J. Stecher, Perceived causality occurs with stroboscopic movement of one or both stimulus elements, Perception **19**, 17 (1990).

[45] R. Mayrhofer and M. R. Waldmann, Indicators of causal agency in physical interactions: The role of the prior context, Cognition **132**, 485 (2014).

[46] A. E. Michotte, *The Perception of Causality (TR Miles, Trans.)* (Methuen Co, London, England, 1963).

[47] A. N. Sanborn, V. K. Mansinghka, and T. L. Griffiths, Reconciling intuitive physics and Newtonian mechanics for colliding objects, Psychol. Rev. **120**, 411 (2013).

[48] B. J. Scholl and P. D. Tremoulet, Perceptual causality and animacy, Trends Cognit. Sci. **4**, 299 (2000).

[49] L. Cohen and L. Oakes, How infants perceive a simple causal event, Dev. Psychol. **29**, 421 (1993).

[50] J. F. Kominsky, B. Strickland, A. E. Wertz, C. Elsner, K. Wynn, and F. C. Keil, Categories and constraints in causal perception, Psychol. Sci. **28**, 1649 (2017).

[51] A. M. Leslie and S. Keeble, Do six-month-old infants perceive causality?, Cognition **25**, 265 (1987).

[52] E. Mascalzoni, L. Regolin, G. Vallortigara, and F. Simion, The cradle of causal reasoning: Newborns' preference for physical causality, Devel. Sci. **16**, 327 (2013).

[53] I. E. K. Andersson and S. Runeson, Realism of confidence, modes of apprehension, and variable-use in visual discrimination of relative mass, Ecol. Psychol. **20**, 1 (2008).

[54] D. Gilden and D. Proffitt, Heuristic judgment of mass ratio in two-body collisions, Percept. Psychophysics **56**, 708 (1994).

[55] D. R. Proffitt and D. L. Gilden, Understanding natural dynamics, J. Exp. Psychol. Hum. Percept. Perform. **15**, 384 (1989).

[56] S. Runeson, P. Juslin, and H. Olsson, Visual perception of dynamic properties: Cue heuristics versus direct-perceptual competence, Psychol. Rev. **107**, 525 (2000).

[57] J. T. Todd and W. H. Warren Jr., Visual perception of relative mass in dynamic events, Perception **11**, 325 (1982).

[58] M. Bunge, *Causality and Modern Science* (Dover Publications, Mineola, New York, 2011).

[59] C. Chen, L. Bao, J. C. Fritchman, and H. Ma, Causal reasoning in understanding Newton's third law, Phys. Rev. Phys. Educ. Res. **17**, 010128 (2021).

[60] R. Knight, *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics*, 4th ed., /content/one-dot-com/one-dot-com/us/en/higher-education/program.html.

[61] R. Corrigan and P. Denton, Causal understanding as a developmental primitive, Dev. Policy Rev. **16**, 162 (1996).

[62] D. Psillos and P. Koumaras, Multiple causal modelling of electrical circuits for enhancing knowledge intelligibility, in *Learning Electricity and Electronics with Advanced Educational Technology*, edited by M. Caillot (Springer, Berlin, Heidelberg, 1993), pp. 57–75.

[63] U. Besson, Some features of causal reasoning: Common sense and physics teaching, Res. Sci. Technol. Educ. **22**, 113 (2004).

[64] B. Y. White, J. R. Frederiksen, and K. T. Spoehr, Conceptual models for understanding the behavior of electrical circuits, in *Learning Electricity and Electronics with Advanced Educational Technology*, edited by M. Caillot, NATO ASI Series (Series F: Computer and Systems Sciences) Vol. 115 (Springer, Berlin, Heidelberg, 1993), https://doi.org/10.1007/978-3-662-02878-0_6.

[65] W. Hung and D. H. Jonassen, Conceptual understanding of causal reasoning in physics, Int. J. Sci. Educ. **28**, 1601 (2006).

[66] K. F. Wilson and D. J. Low, "On Second Thoughts…": Changes of mind as an indication of competing knowledge structures, Am. J. Phys. **83**, 802 (2015).

[67] I. A. Halloun and D. Hestenes, Common sense concepts about motion, Am. J. Phys. **53**, 1056 (1985).

[68] D. P. Maloney, Rule-governed approaches to physics—Newton's third law, Phys. Educ. **19**, 37 (1984).

[69] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.17.020116 for conceptual assessment of Newton's third law and the implied causal attributions.

[70] R. C. Morrison, Weight and gravity—the need for consistent definitions, Phys. Teach. **37**, 51 (1999).

[71] J. Han, K. Koenig, L. Cui, J. Fritchman, D. Li, W. Sun, Z. Fu, and L. Bao, Experimental validation of the half-length Force Concept Inventory, Phys. Rev. Phys. Educ. Res. **12**, 020122 (2016).

[72] J. Han, L. Bao, L. Chen, T. Cai, Y. Pi, S. Zhou, Y. Tu, and K. Koenig, Dividing the Force Concept Inventory into two equivalent half-length tests, Phys. Rev. ST Phys. Educ. Res. **11**, 010112 (2015).

[73] O. L. Liu, H.-S. Lee, and M. C. Linn, Measuring knowledge integration: Validation of four-year assessments, J. Res. Sci. Teach. **48**, 1079 (2011).

[74] O. L. Liu, H.-S. Lee, C. Hofstetter, and M. C. Linn, Assessing knowledge integration in science: construct, measures, and evidence, Educ. Assess. **13**, 33 (2008).