Synthesis of discipline-based education research in physics

Jennifer L. Docktor[1] and José P. Mestre[2,*]

[1]*Department of Physics, University of Wisconsin-La Crosse, 1725 State Street, La Crosse, Wisconsin 54601, USA*

[2]*Department of Physics, University of Illinois, 1110 West Green Street, Urbana, Illinois 61801, USA and Department of Educational Psychology, University of Illinois, 1310 South 6th Street, Champaign, Illinois 61820, USA*

(Received 8 August 2013; published 16 September 2014)

This paper presents a comprehensive synthesis of physics education research at the undergraduate level. It is based on work originally commissioned by the National Academies. Six topical areas are covered: (1) conceptual understanding, (2) problem solving, (3) curriculum and instruction, (4) assessment, (5) cognitive psychology, and (6) attitudes and beliefs about teaching and learning. Each topical section includes sample research questions, theoretical frameworks, common research methodologies, a summary of key findings, strengths and limitations of the research, and areas for future study. Supplemental material proposes promising future directions in physics education research.

DOI: 10.1103/PhysRevSTPER.10.020119 PACS numbers: 01.40.Ha, 01.40.Fk, 01.40.gb

I. INTRODUCTION

This paper synthesizes physics education research (PER) at the undergraduate level, and is based on a paper that was commissioned by the National Research Council to inform a study on the status, contributions, and future directions of discipline-based education research (DBER)—a comprehensive examination of the research on learning and teaching in physics and astronomy, the biological sciences, chemistry, engineering, and the geosciences at the undergraduate level [1,2]. PER is a relatively new field that is about 40 years old, yet it is relatively more mature than its sister fields in biology, chemistry, engineering, astronomy, and geosciences education research. Although much is known about physics teaching and learning, much remains to be learned. This paper discusses some of what the PER field has come to understand about learners, learning, and instruction in six general topical areas described herein.

A. Topical areas covered and organization

Given the breadth and scope of PER to date, we organize this synthesis around six topical areas that capture most of the past research in physics education: conceptual understanding, problem solving, curriculum and instruction, assessment, cognitive psychology, and attitudes and beliefs about learning and teaching. To ensure consistency in the presentation and to aid the DBER committee in its charge, each of the six topical areas is organized under the following sections: research questions; theoretical framework; methodology, data collection or sources and data analysis; findings; strengths and limitations; areas for future studies; references. In this paper, the final section on continuing and future directions of physics education research has been removed and placed in the Supplemental Material [3]. In addition, the references have been compiled at the end of the paper rather than individually by section. Because of the cross-cutting nature of some articles, some that were included in a particular section could have just as easily been included in another section; we highlight the specific features of articles as they pertain to the section's emphasis. Although we did not place any restrictions on the dates of the research studies covered, the great majority of the studies cited are within the past 20 years and were done in the United States. The original commissioned paper included published studies up through October of 2010, and this revised paper includes studies published through May of 2013. In addition to the six topical areas, a summary and our conclusions are presented in Sec. VIII.

Equally important to stating what this paper is covering is stating what has been left out. The commissioned paper had a specific focus on "empirical research on undergraduate teaching and learning in the sciences" as outlined by the criteria set by the National Academies [1,2]. Therefore, the following areas of research have not been included in this review: precollege physics education research (e.g., research on high school physics teaching and learning) and research related to physics teacher preparation or physics teacher curricula. We also made a decision to exclude "how to" articles describing research analyses (e.g., ways of analyzing video interviews of students) which are pertinent for the community of physics education researchers but do not have a direct impact on undergraduate physics education. The coverage herein is extensive,

*mestre@illinois.edu

although not exhaustive; whenever multiple articles were available on the same or similar topics, we selected representative articles rather than including all. For areas of research that have been around longer in PER (e.g., conceptual understanding and problem solving), there exist review articles (which we cite), thereby helping to reduce the length of those sections. Emerging areas of PER, on the other hand, lack review articles and hence those sections may be slightly longer for adequate coverage.

It is somewhat surprising to us that this is the first PER review article appearing in *Physical Review Special Topics—Physics Education Research* (PRST-PER), given the large body of work that exists in PER and that review articles are among the types of articles solicited in the "About…" description of the journal. Although we have attempted to present a balanced view, it is unavoidable that the contents herein will reflect to some extent the views and biases of the authors. We hope to see more review articles in PRST-PER in the future. We apologize if we have left out a particular researcher's favorite work—if we did, it was not intentional.

B. Possible uses of this synthesis

We envision multiple uses of this synthesis. First, it captures a good cross section of PER, and as such it is a good resource for physics faculty, discipline-based education researchers, and, in particular, PER faculty, postdoctoral candidates, and graduate students; the contents herein provide an excellent resource for those interested in an overview of the PER field at the postsecondary level, and would be useful in teaching a graduate seminar on PER. Second, it serves as a historical account of the field, taking stock in where PER has been and where it currently is. Finally, it provides a perspective of the status of other discipline-based education research relative to physics.

II. CONCEPTUAL UNDERSTANDING

One of the earliest and most widely studied areas in physics education research is students' conceptual understanding. Starting in the 1970s, as researchers and instructors became increasingly aware of the difficulties students had in grasping fairly fundamental concepts in physics (e.g., that contact forces do not exert forces at a distance; that interacting bodies exert equal and opposite forces on each other), investigations into the cause of those difficulties became common. Over time these conceptual difficulties have been given several labels, including misconceptions, naive conceptions, and alternative conceptions. In this review we have chosen to use the term *misconceptions* but acknowledge that some researchers may have a preference for other terms.

Some would argue that misconceptions research marked the beginning of modern-day physics education research. Early work consisted of identifying and documenting common student misconceptions [4,5], and there were entire conferences devoted to student misconceptions in the Science, Technology, Engineering and Mathematics (STEM) disciplines [6,7] with thick proceedings emerging from them. Many of these studies also included the development of instructional strategies and curricula to revise students' thinking to be in alignment with appropriate scientific explanations, a process referred to as *conceptual change*. Instructional strategies and curricula are described Sec. IV, and the development of conceptual exams is described in Sec. V.

Recent studies on students' conceptual understanding have broadened the focus from common difficulties to generating theories and explanations on the nature and origin of students' ideas, and describing how those ideas change over time.

A. Research questions

The research questions investigated under the conceptual change generally fall into the following categories.

1. Identifying common misconceptions

What learning difficulties do students possess that are conceptual in nature? What are the most common misconceptions that interfere with the learning of scientific concepts? Much work has gone into documenting preconceptions that students bring into physics classes prior to instruction and identifying which of those are misconceptions that are in conflict with current scientific concepts (see, e.g., Refs. [8,9]). Although many of these studies address topics in mechanics (e.g., kinematics and dynamics), there have also been many studies conducted in electricity and magnetism, light and optics, thermal physics, and a few in modern physics. For a list of approximately 115 studies related to misconceptions in physics, see a resource letter [10].

In addition, many investigations explore whether or not misconceptions persist following instruction, which in turn provide insights into the type of instruction that impacts students' conceptual understanding. This body of work has generated numerous carefully documented studies (see, e.g., Refs. [11,12]) of the role of misconceptions in students' reasoning and learning, as well as several inventories to assess conceptual understanding in several physics domains that will be discussed further in the Sec. V [13–17].

2. Describing the architecture of conceptual structure

What is the nature of scientific concepts in memory? What changes when conceptual change takes place? Another thrust in research on students' conceptual understanding attempts to describe the cognitive architecture of conceptual knowledge (i.e., how conceptual knowledge is structured in memory). This body of work has generated lively debate among those proposing different cognitive architectures [18–23], and some have combined different types of cognitive architectures to explain student reasoning [24]. An interesting emerging line of work in misconceptions uses functional magnetic resonance imaging (fMRI) to investigate the nature of conceptual change; preliminary findings suggest that even when students know the right answers (i.e., when they have supposedly overcome their misconceptions) brain activation suggests that many students may still hold the misconception in memory yet suppress it [25]. Understanding how concepts form in memory and how they are used to reason provides useful insights in devising instructional interventions to help students adopt and use scientific concepts.

3. Developing and evaluating instructional strategies to address students' misconceptions

What instructional interventions are most effective for helping students overcome stubborn misconceptions? Since a primary goal of science instruction is to teach students the major concepts in each discipline as well as how to apply those concepts to solve problems, considerable effort has gone into research to design, evaluate, and refine curricular interventions to target students' stubborn misconceptions. This line of research builds upon the established catalog of common misconceptions in the research literature and uses a cyclical process to devise instructional strategies. Most instructional techniques begin by making students aware of their misconceptions (e.g., by demonstrating to students inconsistencies in their own reasoning across contexts—see discussion of conceptual change below), and then guiding students through a series of activities or reasoning exercises to reshape their concepts to accommodate scientific concepts (see, e.g., Refs. [21,26–31]). Other techniques guide students toward adopting better scientific models via teaching interviews [32].

B. Theoretical framework

There are three main theoretical viewpoints about conceptual understanding (including conceptual development and conceptual change) in the science education community, as summarized below.

1. Naive theories or misconceptions view

This view contends that as students gain knowledge about the world (either through formal schooling or informally), they build "naive theories" about how the physical world works, and that often these naive theories contain misconceptions that contradict scientific concepts [4,5,33] (see Ref. [34] for a review). Hence, students do not come to class as "blank slates" upon which instructors can write appropriate scientific concepts [35]. For example, children observe leaves fluttering down from tree branches and rocks thrown from a bridge onto a stream below, and from these and multiple similar observations construct the notion that heavy objects fall faster than light objects. Misconceptions are viewed as stable entities that are used to reason about similar but varied contexts (e.g., a sheet of paper falling from a table to the floor falls slower than a set of keys falling to the floor, both observations fitting into "heavy objects fall faster than light objects" misconception). Misconceptions have three qualities: (a) they interfere with scientific conceptions that teachers attempt to teach in science classes, (b) they are deeply seated due to the time and effort that students spent constructing them, and they make sense to the student since they do explain many observations (it is difficult to "shut off" air resistance and observe a leaf and a rock falling at the same rate, thereby allowing direct verification of the physicist's view that all objects fall at the same rate), and (c) they are resistant to change.

Some researchers make distinctions between "weak" and "strong" restructuring of misconceptions. For example, young children's eventual conception of the meaning of "alive" is considered strong restructuring [36]; prior to correctly conceptualizing "alive," children believe that to be alive means being active (people and animals are alive by this criterion, but not lichen), or being real or seen (Carey conveys an incident whereby her 4-year-old daughter once proclaimed that it is funny that statues are dead even though you can see them, but Grampa's dead because we cannot see him any more [36]). Other distinctions include intentional versus nonintentional conceptual change, where the former is "…characterized by goal-directed and conscious initiation and regulation of cognitive, metacognitive, and motivational processes to bring about a change in knowledge" [37].

In the misconceptions view it is generally accepted that some degree of dissatisfaction is needed for someone to replace a misconception with a more scientifically appropriate form. As Carey argues, "Without doubt, the process of disequilibration is an important part of the process of conceptual change" [36]. Almost three decades ago Posner and co-workers [33,38] formulated a theory of conceptual change with four components that needed to be present for an individual to abandon a misconception in favor of a scientific concept: (1) dissatisfaction with a current concept, (2) intelligibility of the new concept, (3) initial plausibility of the new concept, and (4) usefulness of the new concept in reasoning and making predictions about phenomena. Strike and Posner [39] have since revised their initial views somewhat to include issues related to the learner's "conceptual ecology" as well as developmental and interactionist considerations.

2. Knowledge in pieces or resources view

Another view of students' knowledge and conceptual change offers a different architecture of concepts. According to the "knowledge in pieces" view [40–42], students' knowledge consists of smaller grain-size pieces that are not necessarily precompiled into larger concepts. Students activate one to several pieces of knowledge in response to context and reason with them on the fly. For example, individuals of all ages [43] often state that it is hot during summer because Earth and the Sun are in closer proximity than during winter. This is interpreted by misconception advocates as a misconception that likely formed from thinking about the path of Earth around the Sun as an exaggerated ellipse. The pieces view is that perhaps the individual generated this explanation on the spot by searching memory for some reasonable explanation, and coming up with the knowledge piece "closer means stronger"; for example, moving closer to a fire makes the heat one feels radiating from it stronger, and moving closer to the stage in a rock concert makes the music louder [44]. The pieces view is similar to Hammer's "resources" view [44], where resources are small units of beliefs or thought whose "correctness" depends on the context in which they are applied, or to Minstrell's [45] "facets."

Proponents of the pieces view argue that there are two major difficulties with the misconceptions view that are not present with the pieces view [40,42,44,46–48]. The first is that a misconception is thought of as a compiled cognitive structure that an individual can employ to reason about a situation, which implies that the individual should show consistency in applying the misconception across similar contexts. Interviews with students as well as assessments reveal that slight changes in physics contexts can lead to radically different answers, reasoning patterns, and/or explanations [20,49,50], suggesting that novice knowledge is highly context dependent and not stable as a misconceptions view would suggest. The second difficulty is that, although misconceptions are constructed over time from experiences and observations, and therefore constructivist in nature, the misconceptions view does not provide an account of how misconceptions eventually evolve into correct scientific concepts. A recent publication, however, argues that the two camps are not as far apart as some might think [51].

3. Ontological categories view

A third perspective of novice knowledge and conceptual change is the "ontological categories" view. This view, attributed to Chi and co-workers [18,19,23,52,53] and building on previous work by Kiel [54], argues that students' naive conceptions are due to miscategorizing knowledge and experiences into inappropriate ontological categories, where ontological categories are loosely defined as the sorts of things there are in the world. Examples of ontological categories are material things like objects, temporal things like events, and processes like the emergent phenomenon that gives rise to such things as flocking behavior in birds. The ontological categories view argues that students sort knowledge into distinct and stable ontological categories and that many of the difficulties in student understanding are due to categorizing processes such as heat and electric current into the matter or "things" category. This would imply that the way to help students overcome misconceptions is to have them change their miscategorized knowledge into the appropriate ontological categories, but misconceptions are difficult for students to dislodge precisely because it is hard for students to recategorize across different ontological categories. For example, students tend to think of forces as things rather than as the interaction between two objects, and hence talk of forces as being used up as gasoline is used up in a car. One well-known example of this is students discussing the forces on a coin thrown vertically up in the air—they mention "the force of the hand" as one of the forces "possessed" by the coin while it is rising, a force which gets used up when the coin reaches the top of its trajectory, after which the force of gravity takes over to make the coin fall [4].

Recently, this view has been challenged [21]. Gupta and collaborators argue that both experts (in talks delivered to other experts in their profession) and novices (when discussing ideas with their peers) are able to traverse between ontological categories without confusion in order to explain complex phenomena. For example, Gupta *et al.* cite physics journal articles in which physicists mix ontological categories to build a coherent argument, such as discussing a pulse emerging from a sample and having a distinct peak as if it were a thing. In addition, they argue that novices do not show stable ontological categories, showing examples of novices speaking about electric current as a thing and one hour after instruction using much more sophisticated processlike descriptions of current. Finally, they also argue that crossing ontological categories is extremely common in how we communicate (e.g., phrases such as anger being "bottled up," or having a cold that cannot "be shaken," treating emotion and sickness as if they were things.)

C. Methodology (data collection or sources and data analysis)

1. Contexts

Research on identifying and documenting misconceptions is done either by administering assessments designed to probe students' views in contexts where concepts need to be applied or discussed or by conducting clinical interviews with students about a context in which persistent conceptual errors seem to be prevalent. Clinical interviews has been the methodology most used in probing students' conceptual architecture in memory; during clinical interviews, which typically last 1 hour, a researcher probes a student's conceptual understanding in a target topic through a series of interviewer-led questions—questions that often are open ended and guided by the student's responses to previous questions. Because interview studies result in large amounts of data that are difficult to analyze and interpret, studies using this methodology rely on small numbers of students. Studies evaluating the efficacy of interventions for helping students overcome misconceptions are "engineered" (i.e., devised somewhat by trial and error based on best guesses from research and teaching experience) using a cyclic process where initial best guesses are made in designing an initial intervention, then trying the intervention and evaluating its effectiveness, and then revising and starting the cycle again until learning outcomes are achieved.

2. Participants

Misconceptions research in STEM has been conducted with students of all ages and in various contexts such as large undergraduate introductory courses, as well as high school physics courses. Although not discussed in this synthesis, misconceptions on physical science concepts have also been studied with middle school and elementary school children. Recently, there have been an increasing number of studies exploring conceptual understanding among upper-division physics students as well as graduate students (see, e.g., Refs. [55,56]).

3. Data sources and analysis

Data sources for misconception studies come from students' performance in assessment questions or from transcripts of clinical interviews. In cases where an intervention or teaching approach is being evaluated, assessments of misconceptions are administered to students prior to, and following, the intervention and differences between the postscores and prescores are analyzed. Data from interview studies are analyzed or interpreted using "grounded theory" [57], defined as developing a theory by observing the behavior of a group (in this case, students) in concert with the researcher's insights and experiences; any theory emerges from data as opposed to formulating hypotheses which are then tested.

D. Findings

Discussing research findings in conceptual understanding can be a controversial issue since, as discussed above, there are three theoretical perspectives describing the nature of concepts in memory and conceptual change. Misconceptions and ways of overcoming them are central themes according to the misconceptions theoretical view, whereas the knowledge in pieces or resources view is not in agreement with the robustness of misconceptions or with how one might go about overcoming them; proponents of this view would argue that, as typically defined, students do not possess misconceptions, but rather compile knowledge pieces on the spot to reason about phenomena, and thus the misconceptions that emerge for us to observe are highly context dependent. There are also distinct differences between the ontological categories view and the other two in terms of explaining the nature of misconceptions and how to help students overcome them. It is, therefore, important to keep in mind these unresolved disagreements in the research community as some salient findings are presented below.

1. Misconceptions

- Students possess misconceptions that are deeply rooted and difficult to dislodge [34,35]. An abundance of misconceptions have been identified across a wide range of physics topics in undergraduate physics (for a good review prior to 2000, see Ref. [10]). Often, misconceptions seemingly disappear and are replaced with scientific concepts following instruction only to reappear months later.
- Several intervention strategies, largely based on the conceptual change framework of Strike and Posner [38], have proven effective at helping students overcome misconceptions. Some strategies, such as one that has been successfully used by the University of Washington Physics Education Research group for many years, are based on a cyclic process that begins by identifying misconceptions in a physics topic, then designing interventions based on previous research, instructor experiences, or best guesses, then piloting the intervention and evaluating its success with post-tests that measure transfer to related situations, then refining the intervention over other cycles until evidence is obtained that the intervention works. It should be pointed out that even the most successful groups at this type of work (e.g., McDermott, Herron, and Shafer at the University of Washington) will freely admit that devising effective instructional strategies to combat misconceptions is often a slow, painstaking task requiring multiple tries—not unlike engineering a solution to a complex problem. Other approaches, such as Clement's "bridging analogies" [26,58,59], start with "anchoring intuitions," which are strong and correct understandings that students possess, and devising interventions that attempt to bridge from students' correct intuitions to related contexts in which students display misconceptions. Yet another very effective approach [29] uses interactive lecture demonstrations to help students overcome prevalent misconceptions involving Newton's third law (i.e., the belief that heavy or fast moving objects exert larger forces on light or stationary objects when the two interact via collisions) by displaying in real time the force exerted by mutually interacting carts during a collision under different conditions (e.g., heavy cart colliding with a light cart or moving cart colliding with a stationary cart, etc.). After experiencing the interactive lecture demonstrations intervention, students retain appropriate understanding of Newton's third law months afterwards.

- Despite the previous bullet point, designing an effective intervention to help students overcome a particular misconception can be elusive [43,60].

2. Knowledge in pieces or resources

- Students possess knowledge pieces, also referred to as “phenomenological primitives” (or p-prims), or resources, that are marshaled to reason about physics contexts. These pieces of knowledge can be recalled singly or in groups and compiled in different ways in real time in response to different contexts. It is common to observe that contexts that are considered equivalent, perhaps even nearly identical by experts, are viewed as different by students and different knowledge pieces are brought to bear to reason about them. Students’ knowledge is more dynamic in the pieces or resources view than in the misconceptions view. As expertise is developed with time and experience, there is refinement of the knowledge in memory whereby knowledge pieces that repeatedly prove effective to recall and apply to particular contexts are compiled into scientific concepts. That is, repeated rehearsal of knowledge pieces recalled and compiled to deal with similar situations leads to the formation of stable scientific concepts in experts.

3. Ontological categories

- Misconceptions stem from students categorizing scientific ideas into inappropriate categories [51,53]. For example, processes are placed in the things category (e.g., electric current is thought of as “fuel” that is used up in light bulbs).
- Misconceptions are relatively easy to fix when they involve modification within the same category, but are difficult to fix when they involve modifications across categories [23,61].
- Instructional strategies designed to help students overcome misconceptions by recategorizing into appropriate ontological categories have shown promise (see, e.g., Refs. [28,62–69]).

E. Strengths and limitations of conceptual understanding research

The strengths and limitations of this body of research include the following.

1. Strengths

- Misconceptions research has raised consciousness among instructors about students’ learning difficulties, and about a misconception prevalent among instructors, namely, that teaching done in a clear, elegant manner, even charismatic instructors, quite often does not help students overcome misconceptions.
- Curricular interventions and assessments (both to be discussed in different sections) have emerged based on misconceptions research.
- Classroom instruction has changed as a result of misconceptions research, with many “active learning” strategies (e.g., the use of polling “clicker” technologies to teach large lecture courses) being practiced that have been shown more effective than traditional instruction at helping students overcome misconceptions.
- The pieces or resources and ontological categories views are attempting to map human cognitive architecture, which if successful can help in designing effective instructional strategies.

2. Limitations

- Designing “definitive” experiments that falsify one of the theoretical views remains elusive, hence the debate among proponents of the three theoretical views continues.
- Although many misconceptions have been cataloged across both introductory and advanced physics topics, it is daunting to ever achieve a complete list.

F. Areas for future study

Research on identifying and documenting misconceptions has been progressing for several decades and has covered an extensive range of physics topics [10], so future research in this area is limited to alternate populations (e.g., upper-division students, see [70]) and yet-to-be-investigated topics. Opportunities for continued research on the nature of student thinking and reasoning exist, including how students’ ideas progress over time. There is a pressing need for studies to help articulate general instructional strategies for guiding students to adopt scientific conceptions, especially when those conflict with students’ existing conceptions. Another promising area for future research is to design experiments to test the three competing viewpoints outlined in the theoretical framework in order to arrive at a more unified view.

III. PROBLEM SOLVING

In addition to research on conceptual understanding, another key focus of physics education research on student learning is *problem solving*, likely because problem solving is a key component of most physics courses. As with other studies of student learning, this research focuses first on documenting what students do while solving problems, and follows with the development and evaluation of instructional strategies to address student difficulties. Research also focused on efforts to develop students’ abilities to “think like a physicist” [71–73]. Since problem solving is a complex cognitive process, this area of research also has a strong overlap with the section on Cognitive Psychology

(Sec. VI). For additional reviews of research on physics problem solving, see Refs. [74,75]. For a historical overview of early pioneers in the field of physics education research (such as Karplus and Arons), refer to the account by Cummings [76].

A. Research questions

The research in problem solving include the following categories and questions.

1. Expert-novice research

What approaches do students use to solve physics problems? How are the problem-solving procedures used by inexperienced problem solvers similar to and different from those used by experienced solvers? How do experts and novices judge whether problems would be solved similarly? Early studies of physics problem solving investigated how beginning students solve physics problems and how their approaches compare to experienced solvers, such as professors [77–83]. This area of research also includes categorization studies to infer how physics knowledge is structured in memory [84–86].

2. Worked examples

How do students study worked-out examples? How do students use solutions from previously solved problems when solving new problems? How do students use instructor solutions to find mistakes in their problem solutions to homework and exams? What features of worked-out problem solutions facilitate student understanding of the example? This body of research explores how students use worked-out problem solutions or previously solved problems to solve new unfamiliar problems [87,88]. It also includes how students use instructor solutions to self-diagnose errors in their own problem solutions [89,90], and how to design effective examples [91].

3. Representations

What representations do students construct during problem solving? How are representations used by students? What is the relationship between facility with representations and problem-solving performance? What instructional strategies promote students' use of representations? How do students frame problem-solving tasks? This research explores the use of external representations for describing information during problem solving, such as pictures, physics-specific descriptions (e.g., free-body diagrams, field line diagrams, or energy bar charts), concept maps, graphs, and equations. Some studies focus on what representations are constructed during problem solving and the manner in which they are used [92–96], whereas other studies explore the facility with which students or experts can translate across multiple representations [97–99]. How students frame the problem-solving task impacts problem solving [100,101], as well as whether problems use numerical versus symbolic values [102].

4. Mathematics in physics

How are the mathematical skills used in physics courses different from the mathematical skills taught in math courses? How do students interpret and use symbols during quantitative problem solving? This area of research explores how quantitative tools from mathematics courses are applied during physics problem solving. Some examples include the use of symbols in equations to represent physical quantities [103–106], vector addition [107], arithmetic, algebra, geometry, calculus (e.g., integration) [108], and proportional reasoning [109].

5. Evaluating the effectiveness of instructional strategies for teaching problem solving

How does instructional strategy X [e.g., cooperative group problem solving] affect students' problem solving skills? To what extent do conceptual approaches to problem solving influence students' conceptual understanding of physics? How do students interact with online computer tutor systems? What features of Web-based homework systems successfully enhance students' problem solving skills? Several instructional strategies have been developed and tested, including the use of alternate types of problems [110–114], adopting an explicit problem-solving framework (i.e., a consistent sequence of problem solving steps) [115], conceptual approaches [116–118], cooperative group problem solving [119], and computer homework or tutor systems to help students become better problem solvers [71,120,121].

B. Theoretical frameworks

This section identifies a few prominent theories about learning and problem solving from cognitive science and educational psychology. Although these frameworks are viewed as useful and relevant for PER, problem-solving researchers in PER often do not clearly define or draw upon a theoretical basis for their research studies. The frameworks reviewed here are intended to provide a starting point for discussion. The following frameworks are included: information-processing models, problem solving by analogy, resources model, and situated cognition.

1. Information-processing models

According to one theory of human problem solving [122–124], problem solving is an iterative process of representation and search for a solution. This theory defines a mental state called the "problem space" consisting of a person's available knowledge and their internal representation or understanding of the task environment, including the initial state (given information), goal state or target, and appropriate operations that can be employed. After representing the problem, a solver engages in a search process during which they select a goal and a method to apply (such as a general heuristic), applies it, evaluates the result of this choice, modifies the goal or selects a subgoal, and proceeds in this fashion until a satisfactory solution is achieved or the problem is abandoned. Gick [125] expanded upon this model to include the role of schema activation. If constructing an internal representation of the problem activates an existing schema or memory framework for solving the problem, a solution is implemented immediately. If no schema is activated, then the solver engages in general search strategies [125].

Within this theory of problem solving, Newell and Simon [123] also describe the organizational structure of memory as an information-processing system consisting of two branches: short-term or "working" memory (STM) and long-term memory. Short-term memory is constrained and can only hold a small amount of information for a limited time, whereas long-term memory is essentially unlimited [124]. However, in order to access information stored in long-term memory, it must be activated and brought into working memory. If problem information and activated knowledge exceed the limits of STM, a solver may experience *cognitive overload* [126], which interferes with attempts to reach a solution. To alleviate this effect, problem information is often stored externally (e.g., written down on paper) or processed with the aid of a tool (e.g., computer) in order to free up space in working memory to devote to the task. As a problem solver becomes more proficient, knowledge and procedures may become *chunked* and some skills become automatic, which also optimizes the capacity of STM.

2. Problem solving by analogy

Instructors report that students often look for similar example problems in physics textbooks when attempting to solve homework problems. This process of using a similar, familiar problem to solve a new problem is referred to as *analogy*. There are three main views of analogical transfer, including structure mapping [127,128], pragmatic schema view [129,130], and exemplar view [131,132]. These views agree that there are three main criteria for making use of a known problem: (1) a person must be reminded of the previous problem and sufficiently remember its content, (2) they must compare and adapt the attributes from one problem to the other (what Gentner calls structure mapping), and (3) the solver must evaluate the effectiveness of this transfer for solving the new problem [133]. To be useful for future problem solving, a person must also abstract common features and generalize from the examples. The work of Holyoak and Koh [130] suggests that similarity can be both at a surface level and at a structural level, and while surface similarity can aid with reminding, it is structural similarity that is important for appropriately solving the new problem. Exemplar-based models emphasize the role of specific example problems for guiding the application of abstract principles [131,132]. Some studies have explored the value of using isomorphic problems in helping students draw analogies between problems [134]. For a more comprehensive review of these views, see Reeves and Weisberg [133].

3. Resources model and epistemic games

Research on how students use and understand mathematical symbols and equations in physics has built upon a theoretical framework called the resources model [44,135]. This model describes knowledge in terms of students' resources for learning, both "conceptual" and "epistemological" in nature [44]. This model helps to explain why students might have the requisite knowledge and skills for solving a problem but fail to activate it in a particular context, and suggests that instruction should be designed to make productive use of students' resources. Tuminaro [136] and Tuminaro and Redish [106] extended this resources model to give a detailed framework for analyzing students' application of mathematics to physics problems. They identified six *epistemic games* or problem-solving approaches that students participate in while using math in physics, including mapping meaning to mathematics, mapping mathematics to meaning, physical mechanism, pictorial analysis, recursive plug and chug, and transliteration to mathematics.

4. Situated cognition

Some research on instructional strategies for teaching problem solving (e.g., cooperative problem solving) is based on the situated cognition model of learning. In this view, knowledge is "situated" or influenced by the particular task and context in which learning is taking place [137]. They propose that school activities should incorporate authentic tasks, and they propose a teaching method called cognitive apprenticeship to facilitate an enculturation process. Cognitive apprenticeship includes three main phases: modeling (demonstrating how to do something), coaching (providing opportunities for guided practice and feedback), and fading (gradually removing scaffolding).

C. Methodology (data collection or sources and data analysis)

1. Contexts

Basic research to understand how students solve problems and identify common student difficulties is typically done outside the classroom, in an experimental or clinical setting with paid volunteers. Studies of pedagogy and curricular interventions are conducted in conjunction with a course where the data collected includes student responses to course materials such as homework and exams.

2. Participants

Studies of physics problem solving frequently involve undergraduate students in large-scale introductory courses (algebra based or calculus based) at a single institution over one or two semesters, sometimes longer. Studies conducted with high school students or disciplinary majors are less common. The "experts" in expert-novice studies are often physics faculty or graduate teaching assistants (TAs) who are experienced with teaching introductory courses.

3. Data sources and analysis

One common data source for problem-solving research is students' written solutions to physics problems that have been designed by the researcher(s) and/or adapted from existing problem sources. In most cases the problems are free response, but occasionally they are in a multiple-choice format. These data are typically analyzed by scoring the solutions according to particular criteria, as determined by comparing the written solutions to features of an "ideal" instructor solution. Sometimes rubrics are used to make scoring criteria objective and improve agreement among multiple scorers. Students' written solutions are used in curricular intervention studies to compare performance in reformed courses to traditional courses on a set of common assessment items.

Another widespread data source, especially for cognitive studies, is think-aloud problem-solving interviews. To collect such data participants are asked to solve physics problems and verbalize their thoughts during the process while being video- and/or audiotaped. Transcripts from these interviews are analyzed using standard qualitative methods such as case studies or grounded theory [57] to elucidate common themes or problem-solving approaches from the statements.

Studies comparing expert and novice problem solvers have traditionally used categorization tasks that require participants to group problems based on similarity of solution. Early studies used card-sorting tasks where subjects physically placed problem statement cards into piles and the resulting category was assigned a name [84,138]. More recent studies have used alternate categorization tasks, such as selecting which of two problems would be solved most like a model problem [85], multiple-choice categorization formats [117], or ranking the similarity of two problems on a numerical scale [139]. Some categorization tasks are analyzed using qualitative methods (to elicit commonalities across card groupings), whereas others compare quantitative ranking or the frequency of similarity judgment responses for different groups of subjects. New approaches are emerging for interpreting card-sorting categorization data [140].

Eye-tracking technology is a rare data source, but in some cases it is used to study gaze patterns or fixation time upon particular features of problems or problem solutions presented on a screen [141–143]. These problem features could include representations (such as pictures or physics diagrams), text, equations, and mathematical steps in example solutions.

D. Findings

1. Expert-novice studies

Early studies of problem solving identified differences between beginning problem solvers and experienced problem solvers in both the way they organize their knowledge of physics and how they approach problems.

- *Categorization*. Expert-novice studies found that when asked to group problems based on similarity of their solution, beginning problem solvers used literal objects from the surface attributes of the problems as category criteria (such as "spring" problems or "incline plane" problems), whereas experienced problem solvers considered the physics concept or principle used to solve the problem when deciding on problem categories, such as grouping conservation of energy problems together [84,144]. It was later shown that expert novice behavior is a continuum not a dichotomy; sometimes beginning students exhibit expertlike behaviors and experienced solvers behave like novices [85]. Further, studies have also demonstrated that novices also rely on terminology and variable names in the problems to make categorization decisions, not just objects in the problem [85,145]. More recently a study examined the type of problems needed to accurately distinguish between experts and novices [146].
- *Problem-solving approaches*. Think-aloud problem-solving interviews found that expert problem solvers typically begin by describing problem information qualitatively and using that information to decide on a solution strategy before writing down equations [77–83]. A successful solver's strategy usually includes the appropriate physics concept or principle and a plan for applying the principle to the particular conditions in the stated problem [147]. In contrast, beginning physics students typically started by writing down equations that match given or desired quantities in the problem statement and performing mathematical manipulations to get an answer [84,148]. A study showing how little good students retain from an introductory course they took in the freshman year when tested on the knowledge in their senior year is sobering [149].
- *Metacognition*. In addition, these studies concluded that experts monitor their progress while solving problems and evaluate the reasonableness of the answer, whereas beginning students frequently get stuck and lack strategies to progress further [81,87,150–152]. Explicit instruction on strategies for evaluating a solution, such as the use of limiting and extreme cases, has been shown to improve performance [151].

2. Worked examples

A common practice in problem-solving instruction is to provide students with worked-out example problems that demonstrate solution steps. This section briefly reviews research on how students study example problem solutions in textbooks, how they refer to previous examples during problem solving, and how they use instructor solutions to detect and correct errors in their own solutions. This research also addresses how to design example solutions to optimize their usefulness.

- *Using examples*. Research conducted by Chi *et al.* [87] and Ferguson-Hessler and de Jong [153] on how students study worked-out example solutions in textbooks concluded that good and poor students used worked examples differently when solving a new problem. In Chi *et al.* [87], the labels "good" and "poor" were determined by a student's success at solving problems after studying the worked examples (12 isomorphic or similar problems and 7 unrelated problems from a textbook). Good students (who scored an average of 82% on problems) referred to a specific line or line(s) in the example to check procedural aspects of their solution, whereas poor students (who only scored an average of 46% on problems) reread the example from the beginning to search for declarative information and a solution procedure they could copy [87]. Ferguson-Hessler and de Jong [153] confirmed these findings and described the actions of good students as "deep processing" of the examples whereas poor students engaged in "superficial processing."

A study by Smith, Mestre, and Ross [143] used eye tracking to investigate what aspects of a problem solution students look at while studying worked-out examples. They found that although students spent a large fraction of time reading conceptual, textual information in the solution, their ability to recall this information later was poor. The students' eye-gaze patterns also indicated they frequently jumped between the text and mathematics in an attempt to integrate these two sources of information.

- *Self-diagnosis*. When students receive feedback on their performance on homework and exams, instructors generally expect students to reflect on and learn from their mistakes. However, research indicates that such self-diagnosis of errors is difficult for students. In a study by Cohen *et al.* [89], only one-third to one-half of students were able to recognize when they used an inappropriate physics principle for a problem. These results were affected by the support provided to the students, indicating that having access to a correct solution and self-diagnosis rubric helped students explain the nature of their mistakes better than an instructor solution alone or only access to the textbook and class notes [90]. A classroom intervention by Henderson and Harper [154] found that requiring students to correct mistakes on assessments improved student learning of content and also improved skills relevant for problem solving, such as reflection.
- *Example structure and the worked-example effect*. Ward and Sweller [91] conducted classroom experiments using topics of geometric optics and kinematics to investigate how the design of example solutions influences learning and problem solving. They found that under certain conditions reviewing correct example solutions was more effective for learning than actually solving problems, what is referred to as the *worked-example effect*. To optimize their usefulness, they claim that examples must direct students' attention to important problem features and minimize cognitive load by keeping information sources separated, such as diagrams, text, and equations.

3. Representations

The term *representation* has multiple interpretations, but for the problem-solving research reviewed here it is used to refer only to concrete, external descriptions used by a solver. Some examples include pictures or sketches, physics-specific descriptions (e.g., free-body diagrams, field line diagrams, ray diagrams, or energy bar charts), concept maps, graphs, and equations or symbolic notation. Some researchers go on to make a distinction between general and physics-specific representations. Reif and Heller [151] suggest that a *basic* description includes initial steps taken to understand a problem, such as introducing symbolic notation and summarizing problem information verbally or pictorially. They separate this from a *theoretical* description that specifies systems and interactions for objects using physics concepts, such as describing motion with position, velocity, and acceleration or describing interactions by force vectors.

- *Representational format affects performance*. Multiple studies have determined that some students give inconsistent answers to the same problem-solving question when it is presented in different representational formats [98,99]. The four formats tested included verbal, diagrammatic (or pictorial), mathematical or symbolic, and graphical. The pattern of performance differed by the task and problem topic, but performance on mathematical quizzes was usually worse than other formats despite students' preference for calculation questions [98].
- *Physics-specific descriptions (might) facilitate problem solving*. A clinical study by Heller and Reif [92] found that subjects who received training materials on procedures for generating physics-specific, theoretical descriptions in mechanics performed better on a set of three problems than an unguided group. The group scored significantly higher on all four aspects of their solutions: describing the motion of objects, drawing a force diagram, correct equations, and the final answer.

They concluded that these "descriptions markedly facilitate the subsequent construction of correct problem solutions" (p. 206). In a study by Larkin [93], five students who were trained to use physical representations for direct-current circuits solved significantly more problems on average than a group of five students who received training on generating and combining equations to obtain an answer. The subjects were presented with three problems, and a problem was considered solved if the student produced a correct solution within a 20-minute time limit. Other studies have produced mixed results regarding the relationship between representations and problem-solving success. Studies by Rosengrant, Van Heuvelen, and Etkina [95,142] found that students who drew correct free-body diagrams were much more likely to solve a problem correctly than those who drew an incorrect diagram or none at all, but those who drew an incorrect diagram performed worse than students who drew no diagram. In a study by Van Heuvelen and Zou [96], students reported multiple work-energy representations were useful tools for solving problems, but the researchers observed little use of the representations on exams and could not determine their relationship to problem-solving performance. Similarly, Rosengrant and Mzoughi [155] found favorable student reactions to the use of impulse-momentum diagrams during class, but no differences in performance. De Leone and Gire [156] did not find a significant correlation between the number of nonmathematical representations used by a problem solver and the number of correct answers, but the nature of the errors made was different for representation users and nonusers.

- *Successful problem solvers use representations differently than unsuccessful solvers*. Some studies indicate that successful problem solvers use qualitative representations to guide the construction of quantitative equations [157] and they use physics diagrams as a tool to check the consistency of their work [158]. In contrast, unsuccessful solvers either do not generate representations or do not make productive use of them for problem solving. Kohl and Finkelstein [159] suggest that both novices and experts generate multiple representations during problem solving and sometimes do so in a different order, but experts can more flexibly move between those multiple representations.

4. Mathematics in physics

Mathematics and equations are often referred to as the "language of physics." This area of study identifies common student difficulties associated with math skills required for physics problems and describes how equations are used differently in the context of physics compared to math.

- *Vectors*. Studies have documented several student difficulties associated with using vectors, particularly for velocity, acceleration, forces, and electric fields (see, e.g., Refs. [107,160–162]). In particular, Nguyen and Meltzer [107] identified widespread difficulties with two-dimensional vector addition and interpreting the magnitude and direction of vectors in graphical problems.
- *Algebraic or proportional reasoning*. Some students have difficulty translating a sentence into a mathematical expression, and in doing so they often place quantities on the wrong side of an equals sign. Cohen and Kanim [109] studied occurrences of the algebra "reversal error" in both algebra and physics contexts, concluding that sentence structure was the most influential factor in guiding translation of sentences expressing relationships among variables into equations. More recently Christianson, Mestre, and Luke [163] found that the reversal error essentially disappears as students practice translating algebraic statements into equations, despite never being given feedback on whether or not their equations were correct.
- *Symbols and equations*. In contrast to mathematical expressions, equations used in physics have conceptual meaning associated with symbols and the relationships among physical quantities [106,136]. Students generally have poorer performance on symbolic questions than on numeric questions, and Torigoe [105] attributes this difference to students' confusion about the meaning of symbols in equations and difficulty keeping track of multiple quantities. Sherin [103,104] identified 21 different symbolic forms for equations in physics and how students interpret them, concluding that it is possible for students to have mathematically correct solutions but inappropriate conceptual understanding for the equations.
- *Use of mathematical skills is context dependent*. In many cases, students have the mathematical resources to solve problems but do not attempt to apply them in the context of a physics class or vice versa [164–166]. This suggests that the availability of resources depends on the context in which a skill was learned and the solver's perception of appropriate knowledge to activate in a particular situation, referred to as *epistemological framing* [136,167]. Cui, Rebello, and Bennett [108] found that students needed prompting and guidance to make the connection between their knowledge of calculus and a physics problem. Sayre and Wittman [168] found that students had a more solid understanding of Cartesian coordinate systems and would persist in applying this system even when a different (polar) coordinate system was more appropriate. They used Resource Theory and Process/Object Theory to explain the development of

unfamiliar ideas, suggesting a "plasticity continuum" measure in which resources can range from being plastic (uncertain about how and when to apply them) to being solid (familiar and confident in their understanding of how to apply them).

5. Instructional Strategies

- *Alternate problem types*. Standard problems presented in textbooks are often well defined and arguably do not reflect the nature of scientific thinking [74]. Several types of problems have been developed as alternatives to these standard problems that emphasize conceptual reasoning for realistic contexts. Some examples include context-rich problems [110,169], experiment problems [113,170], jeopardy problems [114], problem posing [111], ranking tasks [112,171], real-world problems [172], thinking problems [173], and synthesis problems [174].

The effects of implementing alternative problems are sometimes difficult to separate from the instructional strategies used in conjunction with the problems. For example, context-rich problems [110,119] result in improved problem-solving performance when used in conjunction with a problem-solving strategy and cooperative group problem solving. Specifically, students made fewer conceptual mistakes, generated more useful descriptions, and wrote equations that were consistent with their descriptions.

Experiment problems are typically utilized during laboratory sessions, where students can explore the question using objects similar to those in the problem [113]. The article only describes the problems, and does not provide research results on their effectiveness. The use of jeopardy problems resulted in high scores on the mechanics baseline test and the Force Concept Inventory (FCI) [114], but they caution that other instructional curricula were also used in conjunction with these problems.

In problem posing [111], students must generate a problem statement that meets certain criteria (for example, they are shown a diagram of two blocks connected by a string and they must pose a problem that can be solved by Newton's second law). This was found to be an effective tool for assessing a student's knowledge of physics concepts and their problem-solving skills, especially their ability to link appropriate problem contexts with physics principles.

- *Problem-solving frameworks*. Explicitly modeling an organized set of problem-solving steps and reinforcing this framework throughout a course results in higher course performance [92,115,175,176]. Often such frameworks are based on the work of the mathematician Polya [177]: understand the problem, plan the solution, execute the plan, and look back [73,178,179] but are expanded to include explicit guidelines for constructing an initial description of the problem.
- *Qualitative approaches*. Some curricula emphasize performing a qualitative analysis of a problem before writing down quantitative equations. Some examples include Overview, Case Study Physics [73,118,180], hierarchical problem analyses based on principles [181,182], and strategy writing, where a strategy consists of the principle, justification, and procedure needed for solving a problem [117].
- *Problem-solving instruction*. Some studies describe how courses can be restructured to include explicit instruction on problem-solving skills, such as through the use of cooperative group problem solving [119] or a curriculum called Active Learning Problem Sheets [183,184]. For additional instructional strategies, see a meta-analysis presented in Ref. [185].
- *Computer tutors*. Early uses of computers as problem-solving tutors focused on prompting students to use a systematic approach for solving Newton's laws and kinematics problems [186–188]. These tutoring systems have expanded substantially to include sophisticated hints, guidance, and feedback [71,189,190]. Other computer-assisted programs focus on Web-based homework such as the "computer-assisted personalized assignment" system (CAPA) [120,191–193], Mastering Physics, and CyberTutor [121].

Studies comparing paper-and-pencil homework to Web-based homework have found mixed results, with some studies citing no differences in performance [194] and others citing improved performance with Web-based tutors or homework systems [195–198]. Additional research results about homework are summarized in Sec. V.

E. Strengths and limitations of problem-solving research

1. Strengths

A strength of problem-solving research is that it is an established subfield of PER with a strong history and associations to other fields, such as mathematics and cognitive science. As a result, we have a great deal of information about how students solve problems and instructional strategies to address common difficulties. In addition, we can make use of established research methodologies from cognitive science and psychology when conducting research in physics education.

2. Limitations

Many of the conclusions from problem-solving research have been drawn from studies with a low number of subjects, a limited range of problem types and topics, and inconsistent measures of problem-solving performance. In addition, the complex nature of problem solving and variability in problem features make it difficult to isolate particular factors that may be responsible for improvements (or decrements) in problem-solving performance.

- *Low numbers of subjects*. Some early studies had very few subjects and/or the results have not been replicable on a larger scale (see, for example, Ref. [138] on revisiting categorization). Also, few studies have investigated problem-solving approaches of transition stages in the expert-novice continuum, such as undergraduate majors and graduate students [199,200].
- *Limited range of problem-solving tasks*. In many expert-novice studies, the experts are solving exercises that are simple for them, not novel problems they have not seen before. Very few studies have investigated the approaches that experts take on complex problems (see, for example, Ref. [152]). There is little research comparing problem solving across physics topics, for example, contrasting problem solving in mechanics with problem solving in electricity and magnetism or solving problems in upper-division physics courses.
- *Inconsistent problem-solving measures*. There is no widespread, consistent measure of problem-solving performance [201]. In most of the studies reviewed here, problem solving was measured using vague techniques such as mean number of correct answers, exam scores, or course grades without presenting detailed information about the way in which problem solutions were scored. There is some current research to develop a valid and reliable measure of problem solving, but for now this is still a limitation of many studies [202–204].
- *Failure to systematically consider problem features*. Problem solving is a complex topic of study, and most research studies do not systematically explore the effects of changing individual problem or solution features on problem-solving performance. For example, problems can differ along the following dimensions: physics concepts and principles required to solve it (or a combination of multiple principles), the format of the problem statement (text, picture, diagram, graph), the mathematics required for a solution, values provided for quantities (numeric) or absent (symbolic), presence of additional distracting information, context (e.g., real objects like cars or superficial objects like blocks), and the familiarity of the problem context (e.g., sports compared to nuclear particles).

F. Areas for future study

This review identified five key areas of existing research on problem solving: expert-novice research, worked examples, representations, the use of mathematics in physics, and evaluating the effectiveness of instructional strategies for teaching problem solving. These areas include many opportunities for continued research; however, some of the most prominent gaps in research on problem solving include research on worked examples, multiple representations, reducing memory load, and adoption of reformed instruction. In particular, the research on the worked-example effect in physics is sparse, and there are few guidelines for how to best design instructor solutions. Research on multiple representations is both sparse and contradictory, with little evidence regarding the relationship between use of representations and problem-solving performance. Another area that would benefit from future study is developing strategies for effectively reducing memory load while still highlighting important aspects of problem solving. Although there are several instructional strategies and curricula for teaching problem solving, adoption of these practices is not particularly widespread and this warrants additional study.

IV. CURRICULUM AND INSTRUCTION IN PHYSICS

An abundance of instructional methods and curricular materials have been developed to span all aspects of the standard physics course: lecture, recitation, and laboratories. Some of these instructional reforms aim to incorporate active engagement of students in traditional courses, whereas other reforms involve comprehensive structural changes, such as combining lecture, recitation, and labs into a single class environment. Advances in technology have also introduced classroom polling technologies into lectures, computers and sensors for collecting and analyzing laboratory data, Web-based homework and tutoring systems, and computer animations or simulations of physical phenomena. As a result, the development and evaluation of instructional innovations continues to be an active area of PER.

Several PER-based curricular materials have been developed and tested for their effectiveness. There are also several text resources available for physics instructors to learn more about PER-based teaching strategies. Redish's [205] book *Teaching Physics with the Physics Suite* gives an overview of several research-based instructional strategies, curricula, and assessments, and also includes a chapter on cognitive science. Books that provide hints for teaching specific physics topics include Arons's book *A Guide to Introductory Physics Teaching* [206] and Knight's *Five Easy Lessons: Strategies for Successful Physics Teaching* [207]. For a comprehensive review of research on active-learning instruction in physics, see Ref. [208].

A. Research questions

All examples of instructional reform presented here are (either implicitly or explicitly) related to the overarching research question of evaluating the effectiveness of instructional strategies and materials for teaching physics, and determining the conditions under which the instruction does or does not work. We will, therefore, forego listing specific research questions below since nearly all would be of the "Did it work, and under what conditions?" type. Strategies include lecture-based instruction, recitations, and laboratory-based strategies; changes to the overall structure of a class; and general curricular materials such as textbooks and simulations. Many of these examples have been termed "interactive engagement" methods in comparison to traditional, passive lecture methods [209].

1. Lecture-based methods

These instructional methods seek to enhance the lecture format of a physics course by making in-class experiences interactive through student-student interactions and student-instructor interactions [210]. Reforms to lecture often include opportunities for students to discuss topics with their classmates, like in active learning classes supported by classroom polling technologies [211–213], and Interactive Lecture Demonstrations [214]. Another instructional method is to require students to complete prelecture assignments to better prepare them for lecture, such as answering conceptual questions in Just-in-Time Teaching [215] or viewing a presentation such as multimedia learning modules [216–218]. Other lecture-based approaches promote self-reflection about learning, such as requiring students to keep a journal and/or submit weekly reports describing what they have learned and what they are still confused about [219–221].

2. Recitation or discussion methods

These approaches seek to make traditional, passive recitation sessions more interactive and collaborative. Curricular materials include the *Tutorials in Introductory Physics* developed at the University of Washington [27] and adaptations thereof [222,223], such as *Activity-Based Tutorials* and *Open-Source Tutorials* at the University of Maryland [224–226]. Another recitation method utilizes collaborative learning in discussion sessions, such as *Cooperative Group Problem Solving* at the University of Minnesota [110,119].

3. Laboratory methods

The introduction of technological tools for collecting and analyzing data has transformed the physics laboratory and spawned several curricula to accompany these changes, including *RealTime Physics* [227–230] and video analysis software [231]. Other laboratory materials focus on developing scientific abilities (e.g., designing experiments, testing hypotheses) such as *Investigative Science Learning Environments* [232,233], *Computer-based Problem-Solving Labs* [234], and *Scientific Community Labs* [235].

4. Structural changes to classroom environment

Some methods of instructional reform involve alterations to the standard classroom structure. One example is *problem-based learning*, which modifies the lecture setting to be appropriate for cooperative group work [236,237]. Several other examples revise the course to integrate some aspects of lecture, labs, and recitations into a single setting as in *Physics by Inquiry* [238,239], *Workshop Physics* [240,241], *Studio Physics* [242–246], *Student-Centered Active Learning Environment for Undergraduate Programs* (SCALE-UP) [247], and *Technology-Enabled Active Learning* (TEAL) [248–250].

5. General instructional strategies and materials

This class of reform includes physics textbooks that have been written to incorporate results from physics education research such as *Understanding Physics* [251] and *Matter and Interactions* [252] and books about teaching physics such as Redish's [205] *Teaching Physics with the Physics Suite* and Arons's [206] *A Guide to Introductory Physics Teaching*. It also includes computer-based curricula such as animations and interactive simulations of physical phenomena [253–255] and instruction that incorporates computer programming experiences into the course [256,257]. Other curricula cited include high school materials such as *Modeling Instruction* [258] and *Minds on Physics* [259], since these are sometimes adopted for undergraduate courses for nonscience majors [260].

B. Theoretical frameworks

Most modern theories of instructional design for science courses are based on constructivism and its associated theories, such as situated, sociocultural, ecological, everyday, and distributed cognition [261]. The constructivist view of teaching and learning emphasizes the active role that learners take by interacting with the environment and interpreting information in relation to their prior understanding and experiences [262,263]. Constructivism has roots in early theories of education; for example, the notion that experience and prior knowledge influence learning was discussed by the education theorist Dewey [264] and apparent in Piaget's theories of cognitive equilibration via assimilation and accommodation. For a historical review of constructivist theories including a discussion of Piaget, see Ref. [263].

Instructional strategies and materials for teaching undergraduate physics have incorporated constructivist theories by reforming traditional styles of teaching (passive, lecture-based methods) to be more student centered and active, approaches collectively referred to as interactive engagement methods [209]. Examples of instructional strategies and materials to make lectures more interactive are described in the Sec. IV D 1.

Design principles for creating constructivist learning environments include assigning open-ended tasks (e.g., problems) that are authentic and challenging, giving students opportunities to work collaboratively with their peers, and providing appropriate scaffolding for activities [265]. Physics curricula that incorporate alternate problems and cooperative group work include *problem-based learning* [236] and *cooperative group problem solving* [110,119]. Other instructional strategies that also utilize teamwork and cooperative learning in the classroom include *Physics by Inquiry* [238,239], *Tutorials in Introductory Physics* [27], *Workshop Physics* [240,241], *Studio Physics* [242,243,245,246], *SCALE-UP* [247], and *TEAL* [248,249].

Another constructivist-based approach to instruction is the use of *learning cycles* [266,267]. A three-phase learning cycle developed for the Science Curriculum Improvement Study (SCIS) included exploration, concept introduction, and concept application [266]. An extension of this work is the 5E cycle of engagement, exploration, explanation, elaboration, and evaluation [262]. Both approaches include an initial "exploratory" period in which students participate in hands-on activities before the formal introduction of concepts. Several physics laboratory curricula incorporate this aspect of early exploration, such as *Physics by Inquiry* [238,239] and *Investigative Science Learning Environments* [232]. The upper-division curricula *Paradigms in Physics* also cites the use of learning cycles [268].

C. Methodology (data collection or sources and data analysis)

1. Contexts

Studies of the effectiveness of particular instructional strategies and/or materials are typically conducted in the context of physics courses. Oftentimes students' performance in a reformed course is compared to other sections of a similar course taught traditionally, or to past years in which the course was taught differently. Comparisons are typically made between courses within a single institution, but occasionally researchers make cross-institutional comparisons [209].

2. Participants

Studies of reformed instruction frequently involve undergraduate students in large-scale introductory courses (algebra based or calculus based), and in very few instances there are reforms made to upper-division undergraduate physics courses or graduate student courses. Refinements to instructional methods can take place over a period of several years with multiple cohorts of students.

3. Data sources and analysis

The development and evaluation of reformed curricula typically involves a mixture of quantitative and qualitative research methodologies. The development of curricula is more descriptive and often begins with identifying common student difficulties or misconceptions, so it lends itself to qualitative methods such as interviews with students or faculty. Examples include the process undergone to develop the *Physics by Inquiry* and *Tutorials in Introductory Physics* at the University of Washington [5,162,269–271] or the procedures to validate clicker questions at The Ohio State University [211].

Evaluating the effectiveness of reforms made to large introductory physics courses lends itself to quantitative, statistical methods and the use of quantitative measures. Common data sources for evaluating the effectiveness of instructional strategies or materials include student responses to standard course assignments, such as performance on course exams, laboratories, and homework. Sometimes rubrics are written to facilitate scoring for open-ended items [272]. For instructional reforms that have taken place within the past two decades, a common data source is students' performance on pre-post concept inventories such as the Force Concept Inventory and Force and Motion Concept Evaluation (FMCE) or attitudinal surveys. (For a comprehensive discussion of inventories, see Sec. V.) Quantitative data analysis methods are utilized to compare scores, and a commonly cited pretest–post-test measure is the normalized gain [209]. Occasionally qualitative research methods are utilized to examine student discourse and student interactions during a reformed class with the aid of video and audio recording [249].

D. Findings

1. Lecture-based methods

- *Peer discussions and classroom communication systems*. Several methods have been developed to encourage active student involvement during a lecture class, including active learning via classroom polling-facilitated instruction [212] and *Peer Instruction* [213]. In these approaches, students engage in a questioning cycle that typically includes viewing a question presented by the instructor, discussing ideas with classmates, responding to the question, and engaging in a classwide discussion. In questionnaires and interviews, students reported having a positive attitude toward active learning using polling technologies and felt that they learned more than they would have in a passive lecture class [212]. Studies of the effectiveness of *Peer Instruction* cite increased conceptual understanding and problem solving performance [273] and decreased course attrition [274].

The nature and quality of the questions asked is also important [211,275]. Qualitative questions and questions that require reasoning promote discussion more than questions requiring only a calculation or recall of information [275], and students and instructors both report a preference for such conceptual questions over numerical problems [212,276]. Beatty *et al.* [275] suggest several ways in which questions can be crafted to include subtle or common student conceptions, such as sequencing two questions to look similar but require different reasoning in an "oops-go-back" format [275]. Researchers at The Ohio State University went through a process of validating several clicker questions with a sequence of student interviews, faculty interviews, and classroom response data [211]. They identified some situations in which students' perspectives were not anticipated by the question designers, highlighting the importance of validating course materials. The manner in which credit is assigned for responses has been shown to affect the nature of students' conversations [277]. In high-stakes classrooms, group conversations are dominated by high-ability students, whereas low-stakes classrooms promote discussion among students of varying ability levels.

These lecture-based methods frequently employ student response systems such as clickers or other polling or communication devices (see, for example, Ref. [278]). Studies of electronic response systems for gathering student answers indicate that it is not the technology that is important, but the way in which it is used [276,279,280]. Clickers offer advantages for instructors, such as acquiring an immediate gauge of student understanding and keeping an electronic record of response patterns [280]. For a review of research on clickers and best practices for their use, see Ref. [279].

- *Interactive lecture demonstrations* (ILDs). This curriculum follows a sequence in which students make a prediction about what they expect to see in a physical experiment or demonstration, discuss their prediction with peers, observe the event, and compare the observation to their prediction [214]. Each stage of the ILD is guided by questions printed on worksheets. Studies evaluating the effectiveness of ILDs indicate that they significantly improve learning of basic physics concepts measured by the FMCE [29] and that the prediction phase of classroom demonstrations is particularly important for forming scientific concepts [281].
- *Prelecture assignments*. One prelecture method, called Just-in-Time teaching (JiTT), requires students to submit answers to conceptual questions electronically before class [215]. The instructor uses these responses to gauge student understanding and incorporates them into the lecture. JiTT has been shown to improve students' understanding of concepts such as Newton's third law [282]. In another prelecture method, students view multimedia learning module (MLM) presentations with interspersed questions [283]. Students who viewed MLMs performed better on assessments of basic content administered both immediately after viewing the MLMs as well as two weeks later compared to students who read the same material in a static text format [217,218], and MLM viewers also performed better on before-lecture assessment questions indicating the presentations improved students' preparation for class [216]. In addition, when MLMs were implemented, students perceived the course as less difficult, they had a more positive attitude toward physics, and they found lectures more valuable since the MLMs allowed class time to be spent on activities to refine their understanding rather than covering basic content [217]. Similar results were found in a different implementation of the MLM approach [284,285].
- *Reflection on learning*. One practice to facilitate students' reflection on their learning is requiring them to keep a journal and/or submit weekly reports [219–221]. In those studies, students were asked to write down every week what they learned in physics class, how they learned it, and what they still had questions about. Students received credit for submitting responses that were similar in weight to homework. When students' weekly journal statements were coded across several aspects evaluated for their appropriateness (see Ref. [221] for the code criteria), the researchers found a relationship between performance and reflection quality, where students with high conceptual gains on concept inventories tended to show "reflection on learning that is more articulate and epistemologically sophisticated than students with lower conceptual gains" [221].

2. Recitation or discussion methods

In physics courses, a recitation or discussion session refers to a smaller classroom environment that meets once or twice a week, typically consisting of 15–30 students and taught by a graduate or undergraduate teaching assistant. It is intended to provide students with individualized help and feedback, such as with conceptual activities or problem-solving practice. In "traditional" recitations the TA would solve homework problems on the board, whereas reformed sessions are revised to be more interactive.

- *Tutorials. Tutorials in Introductory Physics* (TIP) is a supplementary curriculum developed at the University of Washington (UW) [27]. It consists of a series of pretests, worksheets, and homework assignments intended to develop conceptual understanding and qualitative reasoning skills. The tutorials are designed for use in a small classroom environment in which students work in groups of 3 or 4, but it can also be adapted for larger class settings. The *UW Tutorials* were developed from several interview-based studies on student understanding of topics in velocity and acceleration [269,270,286], kinematics graphs [287], Newton's laws [288], energy and momentum [289,290], geometric optics [31,291], and electricity and magnetism [12,270]. More recent instructional strategies and materials have been developed for topics in light and electromagnetic waves [11,292–294], gas laws [295,296], thermodynamics [297,298], hydrostatic pressure [299], static equilibrium [300], and special cases in momentum [301].

The University of Maryland expanded upon the UW tutorial framework to develop *Activity-Based Tutorials* (ABT), which also relate concepts to mathematics and include technological tools such as computers for data acquisition and displaying videos or simulations [225,226,302]. Another adaptation from University of Maryland, called *Open-Source Tutorials* (OST), permits instructors to modify worksheet materials to meet their needs and places emphasis on developing both students' concepts and epistemologies or beliefs about learning physics [224]. Studies of the effectiveness of TIP and ABT tutorials have produced high gains on multiple-choice questions and pre-post diagnostic measures such as the Force and Motion Concept Evaluation, but sometimes with mixed attitudinal reactions from students [303–305]. Although students reported they viewed tutorials as useful and they liked working in groups, very few (20%) said they enjoy tutorials [303]. A comparison of the three types of tutorials indicates that students using the OST version perform better on course exams and diagnostic tests than TIP or ABT [306].

- *Cooperative learning*. Cooperative learning refers to a strategy for collaborative group activities developed by Johnson *et al.* [307,308]. The principles of cooperative learning have been applied to physics classes in an approach called cooperative group problem solving [110,119] and also integrated into other approaches, such as Just-In-Time Teaching, problem-based learning [236], and Workshop-Studio-SCALE-UP physics. Cooperative group problem solving necessitates the use of problems appropriate for group work, such as context-rich problems [178] or ill-structured problems [236] and has been shown to improve problem-solving performance (for a comprehensive list of alternate problem types, see the Sec. III). However, there are several structural aspects that influence group functioning, such as the size of the group (the optimal size is three), composition (heterogeneous by ability), seating arrangement, use of group role assignments, and testing procedures that include both group and individual components [110].

3. Laboratory methods

These approaches seek to revise traditional labs, which are often described as confirmatory or "cookbook" labs in which students follow step-by-step instructions to verify a result. General approaches to making labs more interactive include engaging students in class discussions before and at the end of class; revising materials such that students must plan and make decisions; and probing for student understanding during class (see the discussion of laboratory reform in Ref. [205]). Also reviewed here are technological tools used in labs.

- *Microcomputer-based labs*. With the introduction of microcomputer-based laboratory (MBL) tools such as sonic rangers, it became possible for students to observe a graph produced simultaneously with the motion of an object. Research on these tools and curricula developed to accompany them determined that they improved students' understanding of kinematics concepts and graphs (e.g., position, velocity, and acceleration) in comparison to paper-and-pencil activities [309,310]. Some researchers attribute the success of MBLs to their real-time data collection, by which students can observe both the motion of an object and a graph of the motion generated simultaneously [309], whereas others do not see a difference between real-time and delayed-time analysis of motion [311,312]. A curriculum that utilizes MBL tools to address common student preconceptions is *RealTime Physics Active Learning Laboratories* [227–230] and computer-based problem-solving labs at the University of Minnesota (UMN), written in LabVIEW [313]. Studies of the *RealTime Physics* laboratories indicate improved student learning of concepts in dynamics, as measured by performance on the FMCE exam [314]. Studies of the early implementation of UMN computer-based problem-solving labs indicated no significant differences between groups using computers and groups using traditional tools with respect to course grades or the Test of Understanding Graphs in Kinematics (TUG-K) inventory; however, the experimental (computer) treatment group scored slightly better on the Force Concept Inventory post-test [234].
- *Video analysis software*. Another computer-based tool that has been integrated into instruction is video analysis software [231,315]. Students who had opportunities to collect and analyze data resulting from the motion of objects in videos performed better on interpreting kinematics graphs; a teacher demonstration alone was not sufficient for producing significant results [231]. The software was also perceived as useful by students and improved students' comfort with using computers [316].
- *Engaging students in the process of science*. Some instructional strategies for the laboratory seek to engage students in the authentic practice of science in open-ended experiments, such as observing new phenomena, developing hypotheses, designing their own experiments to test them, measurement and uncertainty, and reporting results in written papers [72,317,318]. One curriculum that emphasizes the development of scientific abilities is the *Investigative Science Learning Environment* (ISLE) labs

[233,317,319,320]. ISLE has been shown to develop students' scientific abilities, such as multiple representations, designing an experiment, collecting and analyzing data, evaluating an experimental outcome, and scientistlike communication [233,272,321]. Students who participated in physics laboratory design activities were also able to transfer skills such as evaluating assumptions and experimental uncertainties to a novel task in biology [233]. In ISLE students go through a learning cycle of observation and exploration, developing explanations, designing experiments to test out and refine their hypotheses, and applying knowledge to solve problems [233,266]. Although labs are a key component of the ISLE curriculum, it also includes worksheets and activities appropriate for recitations and large class meetings in the *Physics Active Learning Guide* [319].

Another laboratory curriculum is the *Problem Solving Labs* developed at the University of Minnesota [322]. Each laboratory experiment is motivated by a context-rich problem, and students are required to complete guiding warm-up questions to reach a prediction before attending the lab session. During lab they discuss predictions in their cooperative groups, and are prompted to make decisions such as planning the data collection and analysis. Although there are research results with respect to UMN's context-rich problems and cooperative group problem solving [110,119], in addition to a comparison of computer and noncomputer labs [234], research on the effectiveness of the problem-solving laboratory curricula itself is not currently available and warrants future study.

A laboratory curriculum called *Scientific Community Labs* was developed at the University of Maryland [323]. The labs take a conceptual approach to learning measurement techniques, including the concept of uncertainty [235]. During lab, students are given a question and must devise an experiment, collect and analyze data, and present their results to the class. Confidence in the results (or "error analysis") is an explicit part of the discussion. Students using the reformed laboratories exhibited higher gains on a physics measurement questionnaire and more sophisticated reasoning when comparing data sets, such as using "range overlap" rather than percent difference [235].

4. Structural changes to the classroom environment

This section reviews instructional interventions that alter the standard classroom structure. In some instances this includes changing the lecture seating and incorporating technology to accommodate group work on meaningful problems, and in other situations the entire course is restructured to combine lecture, lab, and recitation into a single interactive environment (e.g., workshop and studio class settings).

- *Problem-based learning* (PBL). Developed at the University of Delaware in the early 1990s, problem-based learning [324] has been implemented in a variety of undergraduate science courses, including some in physics [236]. Students work in small cooperative learning teams to solve a complex, real-world problem that may take several weeks to investigate. Lectures are kept to a minimum, and the instructor and/or peer tutors help facilitate the functioning of groups and ask probing questions during class. Research on PBL in physics classes has determined that it helps students develop critical thinking and communication skills [325,326].
- *Physics by Inquiry* (PbI). Developed at the University of Washington, the *Physics by Inquiry* curriculum includes a workbook of narratives, laboratory-based experiments and exercises, and supplementary problems [238,239,327]. Students keep a notebook of their observations and exercises, to help them reflect on how their understanding of science concepts changes throughout a course. The targeted audiences for PbI include preservice and in-service K-12 science teachers, underprepared students, and nonscience students. Many of the research studies to develop and evaluate the PbI materials were conducted concurrently with the studies of *Tutorials in Introductory Physics* described above. Examples include curricula on light and shadows [31] or on density (e.g., sinking and floating) [328]. A study comparing Physics by Inquiry to traditional instruction found that elementary education students in a PbI course performed significantly better on a quantitative and a qualitative exam problem than physics and engineering students in a traditional introductory physics course [329].
- *Workshop Physics* (WP) *and Explorations in Physics* (EiP). *Workshop Physics* was developed at Dickinson College, where the calculus-based physics course was restructured to eliminate formal lectures [240,241,330]. Instead, students meet for three 2-hour sessions per week to engage in experiments and analyze data using computer tools, such as curve fitting with spreadsheet software. Research on WP indicates that students prefer the workshop method, have improved performance on conceptual questions (at some but not all institutions), and equal problem-solving performance in comparison to traditional sections of the same course [240]. The WP activities are published in multiple workbook volumes, called the *Workshop Physics Activity Guide* [241]. EiP is the equivalent of Workshop Physics, but designed for nonscience students [331].
- *Studio Physics*. The term "studio" comes from Wilson's description of a "comprehensive unified physics learning environment" or CUPLE physics studio implemented at Rensselaer [246]. In *Studio*

Physics, the lecture and laboratory sessions are combined into a single course session of 30–45 students meeting for two hours, 2–4 times per week [242] with computer-based activities and collaborative group work. Early implementations of Studio Physics showed low gains on the Force Concept Inventory exam (similar to scores from traditional courses), but the introduction of research-based methods such as interactive lecture demonstrations and cooperative group problem solving significantly improved learning gains [242]. Hoellwarth, Moelter, and Knight [243] found that students in a studio section had significantly higher normalized gain on the FCI and the FMCE than students in a traditional classroom; however, students' scores on quantitative final exam problems were the same or slightly worse in the studio sections.

In another instantiation of studio, called New Studio Physics, the lecture remains separate but the recitation and laboratory sessions are combined into a 2-hour session, two times per week [245]. Implementation of New Studio Physics found gains on the Force Concept Inventory exam similar to those obtained by other interactive engagement instructional methods.

- *SCALE-UP*. The acronym SCALE-UP stands for student-centered active learning environment for undergraduate programs, which was first developed and studied at North Carolina State University in physics classes and has been adopted by several other programs and institutions [247,332]. SCALE-UP uses a very carefully designed classroom in which students (typically 50–100) work at round tables in teams (typically 3 teams of 3 students at each table) with networked laptops and access to whiteboards. There are multiple display screens around the room. Students engage in hands-on activities including viewing interactive computer simulations, responding to questions or problems, and conducting hypothesis-driven laboratory experiments. Research on SCALE-UP cites improved problem-solving scores on course exams, conceptual understanding gains as measured by pre-post concept inventories, slightly more positive attitudes, and reduced attrition (see "How do you know it works?" in Ref. [333], and see also [247,334]).
- *TEAL*. One extension of Studio-SCALE-UP is the *Technology-Enabled Active Learning* (TEAL) project at the Massachusetts Institute of Technology [248,250]. TEAL uses the classroom design of SCALE-UP, which combines lecture, recitation, and laboratories into a single media-enhanced classroom. Like SCALE-UP, it incorporates Web-based homework assignments, conceptual questions with clickers, and two- and three-dimensional visualization tools for learning electromagnetism concepts [250]. Studies of TEAL indicate higher learning gains on pre- and post-tests of multiple-choice conceptual questions (which were adapted from a variety of research-based sources) and a lower failure rate than the traditional course structure [249,335]. Open-ended surveys and focus groups indicated that students' attitudes toward the TEAL classrooms varied, with several students expressing positive comments regarding the usefulness of the collaborative activities and others expressing they preferred a lecture format, or thought that group work was too unguided [249].

5. General instructional strategies and materials

- *Textbooks*. *Understanding Physics* is a text written by Cummings *et al.* [251] based on results from physics education research. The text includes touchstone example problems and alternate types of problems, microcomputer-based laboratory tools, and topic organization to facilitate student understanding. The text *Physics for Scientists and Engineers: A Strategic Approach* by Knight [336] also incorporates PER-based ideas such as multiple representations, alternate problem types, and a four-step problem-solving strategy for all worked examples: model, visualize, solve, and assess. The curriculum *Six Ideas that Shaped Physics* was developed and tested over a period of more than eight years as part of the Introductory University Physics Project (IUPP) at Pomona College [337]. The volumes are structured to emphasize the hierarchical structure of physics (e.g., volume C on conservation laws) and provide support for active learning. Courses using the materials found very high gains on conceptual inventories including the FCI and Brief Electricity and Magnetism Assessment (BEMA), with normalized gains in the range of 0.50–0.70 (see "Evidence for Success," in Ref. [338]).

Another research-based curricula is *Matter & Interactions* (M&I) developed at North Carolina State University [252]. M&I approaches calculus-based introductory physics from a "modern" perspective, highlighting the atomic nature of matter, relativity, and computational modeling of complex systems [339–341]. *Matter & Interactions* has resulted in higher gains on the BEMA pre-post diagnostic test [342], but lower gains on the FCI compared to traditional instruction—a finding that perhaps could be explained by differentially more coverage of FCI topics in the traditional curriculum [343].

- *Problem solving*. For a review of instructional strategies and materials for teaching problem solving, see Sec. III. The methods reviewed there include alternate types of problems, qualitative approaches, teaching an explicit framework, and Web-based homework and computer tutors for problem solving.
- *Animations, simulations, and computer programming*. A recent trend in instructional strategies and materials is the use of technological tools in the classroom, such

as animations, simulations, visualization tools, and computer programming experiences. With respect to research on animations, Dancy and Beichner [344] compared students' responses to Force Concept Inventory questions with static pictures and computer animations for the same questions. They found that the students were better able to interpret the question's intention when viewed in animated form, and they gave an answer more diagnostic of their understanding. Another study by Yeo *et al.* [345] found that students using interactive media often retained their intuitive (incorrect) conceptions about motion unless they had explicit guidance from researcher interventions. Steinberg [346] did not find performance differences between students who interacted with a simulation for learning air resistance compared to those who completed paper-and-pencil versions; however, students who interacted with computer simulations were more reliant upon the computer to provide them with the correct answer. Computer visualization tools have been shown to be particularly effective for abstract concepts in electricity and magnetism, such as the concept of a field [217,218,249,335,347]. Although textbooks often include access to online visualization or animation tools, research support for their development and effectiveness is not always available and/or cited [348].

One of the most widely cited projects to develop and test interactive science simulations is the University of Colorado's PhET project [349]. The Web site includes a range of free simulations of phenomena for physics, biology, chemistry, Earth science, and math; it also includes a bank of activities created by teachers to accompany the simulations. The physics section includes (to date) more than 70 simulations. There have been several research studies published regarding the effectiveness of the PhET simulations [253,254,350]. One study in particular [255] found that students who used a computer simulation to explore direct-current circuits rather than real light bulbs, meters, and wires had higher scores on their lab report, scored higher on final exam questions pertaining to circuits, and were more successful at performing tasks including assembling a real circuit. They cite that computer simulations are particularly useful for instruction in that they allow students to view "hidden" phenomena, such as current flow, and focus their attention on relevant details of the representation.

Finally, a few studies have explored the use of computer programming in physics courses. An early study with programming in a course for physics majors was the project M.U.P.P.E.T.: *Maryland University Project in Physics and Educational Technology* [257]. The M.U.P.P.E.T. project determined that programming in introductory physics facilitated the use of more realistic problems and modern topics, and the use of carefully designed projects could encourage students' use of qualitative and analytic reasoning skills. Another instantiation of computer programming in introductory physics is the use of VPython in the *Matter & Interactions* curriculum [256]. In addition to teaching students some basic programming and debugging skills, the computational tools also allow students to visualize abstract phenomena and tackle problems that cannot be solved analytically. For a more comprehensive review of research on integrating computational physics into the undergraduate curriculum, see Ref. [351].

Homework. There are few studies related to effective procedures for the assignment, collection, and grading of homework in large enrollment courses. One study found that allowing students some choice in which problems to solve and making a fraction of problem solutions available before homework is due provided more timely feedback to students, and resulted in an increase in self-direction [352].

- *Preservice and in-service teachers*. As mentioned previously, the *Physics by Inquiry* laboratory materials are appropriate for educating preservice and in-service K-12 teachers about physical science [238,239]. A different curriculum targeted toward elementary teachers is *Physics and Everyday Thinking* (PET) [353]. PET has shown a shift toward expertlike attitudes as measured by the Colorado Learning Attitudes about Science Survey [354].
- *High school curricula*. Although not reviewed here, it is worth mentioning that there have been several texts and workbooks developed for use in middle school physical science and high school physics, which are sometimes adapted for undergraduate courses with nonscience majors. Examples include *Modeling Instruction* [258,260], *Minds on Physics* [259], *It's About Time: Active Physics* [355], *Tools for Scientific Thinking* [239], and a curriculum for 7th-8th graders called *InterActions in Physical Science* [356].
- *Upper-division curricula*. One curriculum that reforms the junior-level and senior-level courses for physics majors is the *Paradigms in Physics* project at Oregon State University [268,357]. It includes revisions to the structure or order of topics, content of courses, and instructional methods, in an effort to better reflect physicists' views and experiences in the field. Examples include adding laboratory experiments as a component in several courses, making a stronger link between mathematics and physics, and introducing computational examples and problems [357]. Evaluation of the program indicates improved retention of physics majors and improved physics problem-solving abilities [268]. The development and evaluation of *UW Tutorials* curricula for upper-division mechanics courses also shows promising effects on learning advanced topics [358]. The University of Colorado-Boulder has transformed their

upper-division electricity and magnetism course [359–362] and quantum mechanics course [363] by integrating active engagement methods such as using conceptual questions during lecture, tutorials during recitations, and homework that utilizes multiple representations. They have identified several areas in which upper-division students have difficulty, such as communicating understanding, visualizing fields, problem-solving techniques [359], and the use of conceptual clicker questions can give faculty insight into students' understanding [362].

Reform to graduate education is less common. Carr and McKagan [364] researched the implementation of PER methods into a graduate quantum mechanics course, including changes to the content, textbook, instructional methods, and assessment. Materials were adapted during the course in response to student surveys and interviews. They found that students became more interested in course material when the instructor made explicit connections to modern research, and students were responsive and positive toward the lecture modifications, such as the increase in peer discussions during class.

- *The "learning assistant"* (LA) *model*. A curricular reform developed at the University of Colorado [365] uses undergraduate students who performed well in an introductory course to serve as learning assistants in following years. LAs facilitate small group interactions in lectures and recitation sections. Use of LAs in teaching has resulted in higher learning gains for students in the BEMA and FMCE than achieved with traditional instruction; serving as a learning assistant also increases dramatically the likelihood that they will enter the teaching profession. Another study indicates that LAs help improve student performance when used in lab settings [366].

Although many of the instructional materials described here are stand-alone curricula, they are sometimes implemented in combination. Redish's [205] text outlines how the *Physics Suite* integrates several curricula into an activity-based physics program, including *Understanding Physics* [251], alternate types of problems, interactive lecture demonstrations [214], *Workshop Physics* [241], and computer tools in workshops and labs with *RealTime Physics* [227–230].

E. Strengths and Limitations

1. Strengths

A strength of research on curriculum and instruction in physics is the sheer abundance of methods that have been developed and evaluated with research. In particular, many early approaches have been tested and refined over several decades to the point of becoming mainstream in introductory undergraduate courses (such as the use of conceptual questions and polling technologies, like clickers). Adoption of research-based instruction is also spreading as a result of new faculty workshops sponsored by the American Association of Physics Teachers, American Physical Society, and American Astronomical Society [367].

2. Limitations

This review identified some instructional strategies and curricula that have not been extensively tested for their effectiveness in physics, such as the *Just-in-Time* teaching method and several laboratory-based curricula. In addition, although some textbooks are PER based, many of the most commonly used texts for undergraduate courses are not, and this can create a mismatch between the goals of reformed instruction and the materials actually used in the course; to date, a large percentage of college physics courses are inconsistent with evidence-based reforms [368,369].

F. Areas for future study

- *Upper-division undergraduate courses and graduate education*. An area of future study includes expanding evidence-based instructional strategies and materials that have been developed for introductory undergraduate physics courses to other populations, such as upper-division (junior- and senior-level) courses for physics majors and courses for graduate students. Although there have been some studies of curriculum and instruction in upper-division courses [268,362,364,370], this area would benefit from additional research.
- *High school–university cross-curricular studies*. Although several of the physics curricula are identified as appropriate for a particular student population, it is possible that some curricula for high school might be appropriate for university nonscience majors (see Ref. [260]) and that university curricula might be appropriate for advanced high school courses. The application of instructional strategies and materials to alternate groups of students is a possible area for future research.
- *Courses for biology and premedicine students*. There is some current interest in reforming the introductory course for biology and premedicine students [371]. Additional discussion is needed to determine how current courses are (or are not) meeting the needs of this student population, and research is needed to develop and test reformed instruction and curricula. At least one research group in Germany has developed and tested the effectiveness of a reformed physics laboratory curriculum for medical students [372]. They found that including medical applications (e.g., connecting electricity and neural functions) improved students' attitudes towards physics and improved their ability to relate concepts of physics and medicine as measured by the construction of concept maps.

- *The influence of instructor modifications on the success of reformed curricula*. Instructors do not always implement PER-based instructional methods in the way that the developers intended them, and it is unclear what effect this has on the effectiveness of the method. An area of current research is to explore ways in which instructors modify methods when implementing them in their classes, their reasoning for doing so, and how this influences student learning. For examples of some studies, see Sec. VII.
- *Lab instruction*. Much of the early research on labs focused on the integration of computer-based sensors and technological tools into the laboratory, but it is possible that many institutions still use confirmatory-style, "cookbook" labs, despite the use of reformed instruction in other aspects of the course (such as lecture and recitation). With the exceptions of ISLE, *RealTime Physics*, and integrated course structures (workshop, studio, SCALE-UP, and TEAL), there has not been much published research on the effectiveness of laboratory curricula.
- *Textbooks*. Although there is unpublished research about the "readability" of introductory physics textbooks [373], there is a need for more research on how students read and interact with physics textbooks and components in differing formats, such as equations, diagrams, text, colored photos, etc. Cognitive load theory suggests that many "enhancements" to textbooks may create extraneous memory load that detract from, rather than add to, learning, although little research exists on the design of effective textbooks (see Refs. [374,375] for reviews).
- *Instructional technology*. Although there were several aspects of technology identified here (e.g., classroom polling technologies, multimedia presentations, computer data collection and analysis tools, and computer programming), it is expected that technology and computer-based instruction will continue to evolve, and this will require future research on the effective use of these technologies in classrooms.

V. ASSESSMENT

The development and validation of multiple-choice concept inventories was (and continues to be) an influential area of research in physics education, particularly the development of mechanics and force concept tests, such as the Mechanics Diagnostic Test [376], Mechanics Baseline Test [14], the FCI [15], and the FMCE [17]. In particular, the FCI has experienced widespread use to evaluate the effectiveness of instruction within and across several institutions [209]. The bank of available concept inventories has expanded to more than 30 physics inventories, including surveys of attitudes and beliefs [377].

The process for developing and validating concept tests typically begins with qualitative studies of student difficulties and incorrect responses (or a review of existing research on common misconceptions), drafting and piloting questions, and a series of statistical procedures to evaluate items and scores (see the handbook [378]; see also [379]). For a comparison across several different inventory development methodologies, see Lindell, Peak, and Foster [380]. Research on assessment has broadened to include exploring correlations between inventory scores and other measures of performance, comparing scores across multiple populations (culture and gender), and exploring the value of complex models of student learning beyond pre-post scores. Assessment also refers to alternate formats for measuring student understanding such as rubrics and exams, which will also be discussed in this section.

A. Research questions

1. Development and validation of concept inventories

To what extent do scores on a multiple-choice concept test agree with students' written responses to open-ended questions (a measure of test score validity)? How does the context and format of concept test questions impact student responses? How do the conditions under which a concept test is administered affect students' performance on the test? This area of research includes developing and validating concept inventories, methods for analyzing results from concept inventories [381], and issues related to context sensitivity and studies of testing conditions [382].

2. Comparing scores across multiple measures

What is the relationship between students' scores on the Force Concept Inventory and performance in a physics course as measured by grades? What is the relationship between students' performance on concept tests and measures of their mathematical skills? Several studies have compared scores on concept inventories to other measures such as SAT scores [383], scientific reasoning ability [384], scores on a math diagnostic test [385], as well as how students blend mathematical and conceptual knowledge [386]. Other possible measures of comparison include grade point average, course grades, or final exam scores. Some studies also compare tests on one inventory to those on another, such as comparing the FCI to the FMCE [387].

3. Comparing scores across multiple populations (culture and gender)

How do females' average scores on the Force Concept Inventory test compare to males' average scores? How do FCI scores compare across cultures (e.g., American and Chinese students)? Are there different gender patterns in clicker responses? There have also been some recent studies comparing inventory scores across multiple social or cultural groups, such as comparing American and Chinese students [388,389] and comparing males and females [390–392]. There is also some research on gender differences in the use of online homework [393] and in the response patterns of males and females using clickers [394].

4. Course exams and homework

To what extent does performance on multiple-choice questions agree with students' performance on equivalent open-ended problems? How frequently should homework be assigned and graded, if at all? What format of grader feedback on homework is most effective for promoting student learning? This area includes research on multiple-choice course exams, particularly their equivalence to performance on open-ended problems [395,396] and homework grading [352].

5. Rubrics for process assessment

To what extent do scores on rubric-based grading agree with other measures of student performance (e.g., traditional grading of a problem solution)? What is the level of agreement (consistency) among multiple people using a rubric to score the same student papers? Research questions that pertain to the development of rubrics focus on the validity and reliability of scores on the assessment. Examples of physics rubrics include the scientific abilities rubrics developed for the ISLE curricula [317] and problem-solving rubrics [205,397]. There are also assessments under development for problem solving [398].

6. Complex models of student learning

To what extent do predictions from mathematical models agree with measures of student learning? How are student responses to conceptual questions affected by instruction? What is the pattern of student learning throughout a course? Some researchers are attempting to create mathematical models to describe student learning [399,400] and others are investigating the peaks and decays of learning that occur "between" pre- and post-testing [401,402].

B. Theoretical frameworks

Research on the development and validation of assessments is typically guided by theories of test development from fields such as quantitative measurement in education (see Ref. [378] for an introduction to classical test theory; see also Ref. [403]. It is also guided by evolving definitions of reliability and validity (see Chap. 3 of Ref. [205] for a review). Historically, validity was defined as determining whether an assessment measures what it claims to measure and relied on statistical measures for each "type" of validity, such as content, construct, predictive, and concurrent validity. The current view of validity has shifted away from the "types" paradigm to a unified concept of validity that includes multiple sources of evidence [404]. The most recent edition of the *Standards for Educational and Psychological Testing* describes validity in the following way:

> *Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests… The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself. (p. 9).*

The 1999 edition of the *Standards for Educational and Psychological Testing* cites five primary sources of evidence for validity: evidence based on test content, response processes, internal structure, relations to other variables (external structure), and consequences of testing. For example, the aspect of *content* considers whether the tasks or items on an assessment are representative and relevant to the domain or topic being tested. The aspect of *response processes* considers whether the processes engaged in by a respondent while taking a test are consistent with the assessment developers' intentions. Internal and external structure aspects consider interrelationships among parts of a test and comparisons with other measures of the same construct.

In addition to issues of validity, the *Standards for Educational and Psychological Testing* also address the importance of reliability and utility for designing tests. Reliability refers to the consistency or stability of scores and score interpretations on an assessment, such as agreement among multiple graders for a free-response question. Utility refers to the potential usefulness of the assessment as perceived by its users, such as researchers and instructors.

The development of assessments is also guided by research literature on test design, such as the procedures outlined by Engelhardt [378] and Treagust [379]. According to these sources, a general procedure for developing concept tests in science includes the following steps: (1) identify knowledge or define the content area in science you want to test; (2) identify student misconceptions associated with that topic by examining the existing literature, or by student interviews and responses to open-ended questions; (3) develop items and pilot test them in two-tier form—asking for response and written reasoning for the response.

C. Methodology (data collection or sources and data analysis)

1. Contexts

Basic research to identify common student difficulties is typically done outside the classroom, in an experimental or clinical setting with paid volunteers (refer Sec. II). During the development of concept assessments, sample items may also be pilot tested with student volunteers, or could be tested in the context of a course. Statistical analyses of quantitative assessment scores are typically conducted in conjunction with a course where the data collected includes student responses to concept inventory questions, homework, or exams.

2. Participants

Studies of physics assessments frequently involve undergraduate students in large-scale introductory courses (algebra based or calculus based) at a single institution over one or two semesters, sometimes longer. Studies conducted with high school students or disciplinary majors are less common. Very few meta-analyses have been conducted to compare concept inventory scores across multiple institutions. One such analysis is Hake [209], who examined Force Concept Inventory scores for 62 mechanics courses, comparing interactive-engagement instructional strategies to “traditional” instructional methods.

3. Data sources and analysis

The early development stages of a concept inventory typically include interviewing students or reviewing existing research literature to identify misconceptions and/or propose assessment items, and pilot testing these items in both a free-response format and a multiple-choice format. After an assessment has been drafted, the statistical analysis of scores (often to measure the validity and reliability of scores on items) typically involves numerical data and quantitative data analysis. Studies comparing performance on multiple measures or across student populations typically utilize statistical measures, such as correlation coefficients.

D. Findings

1. Development and validation of concept inventories

Concept inventories used in physics. As stated previously, more than 30 physics concept inventories or attitude surveys have been written [377]. Many are still under development and undergoing validation procedures, and about half of them have been published. The inventories with accessible references are listed below, in chronological order of their publication date.

Published concept inventories and validation studies:

- MDT: Mechanics Diagnostic Test [376]
- MBT: Mechanics Baseline Test [14]
- FCI: Force Concept Inventory [15]. In the initial article published about the FCI, they identified six conceptual dimensions included on the test (kinematics, Newton’s three laws, the superposition principle, and kinds of force). These came into question when factor analysis could not statistically confirm the validity of those categories, prompting a discussion regarding what the FCI actually measures [405–407].
- FMCE: Force and Motion Conceptual Evaluation [17]. An analysis by Ramlo [408] established validity and reliability estimates for the FMCE.
- TUG-K: Test of Understanding Graphs in Kinematics [13]
- CSEM: Conceptual Survey in Electricity and Magnetism [16]
- ECS: Energy Concepts Survey [409]
- DIRECT: Determining and Interpreting Resistive Electric Circuits concept Test [410]
- BEMA: Brief Electricity and Magnetism Assessment [411]
- R-FCI: Representational variant of the Force Concept Inventory [412]
- QMCS: Quantum Mechanics Conceptual Survey [413]
- CUE: Colorado Upper-Division Electrostatic assessment [414]
- FVA: Force, Velocity and Acceleration assessment [415]

Other assessments:

- LCTSR: Lawson classroom test of scientific reasoning [416] (reproduced in Coletta and Phillips [384]). This assessment includes questions on conservation, proportional thinking, identification of variables, probabilistic thinking, and hypothetico-deductive reasoning.
- *Statistical analysis of scores*. In a meta-analysis of FCI scores, Hake [209] uses a measure he refers to as “normalized gain” to compare pretest and post-test scores, which is examined in more detail in Ref. [417]. The average normalized gain for a group of students is computed by taking the average percent gain (average post-test score %—average pretest score %) divided by the possible gain (100%—average pretest score %). Hake’s analysis concluded that courses utilizing interactive engagement methods produced an average normalized gain of 0.48 ± 0.14 standard deviations on the FCI, whereas courses taught using “traditional” methods produced an average normalized gain of 0.23 ± 0.04 standard deviations. It has become a common practice in PER to compare the normalized gain values obtained in a study to these values reported by Hake [209].

Although the normalized gain measure is popular, it has received some criticism. Marx and Cummings [418] suggest revising the gain measure to a proposed measure of “normalized change.” The main distinction between the two measures is that normalized change revises the formula for situations in which the post-test score is lower than the pretest score, and drops students who score 100% on the pretest (a normalized gain of infinity).

Ding and Beichner [381] identify five statistical means for analyzing test scores to assess the validity and reliability of those scores: classical test theory, factor analysis, cluster analysis, item response theory, and model analysis. Other publications explore these and other analysis methods in greater detail, including item response theory [121,419,420], cluster analysis [421,422], Rasch model based analysis [423], concentration analysis [424], and model analysis [399].

- *Influence of testing conditions*. A study by Ding *et al.* [382] outlined several testing conditions that affected pre- and postinstruction scores on the Conceptual Survey in Electricity & Magnetism (CSEM) exam. For example, they found that moving the test administration time by a few days or lectures could have a substantial effect on the CSEM pretest scores, since topics of charge and Coulomb force are often discussed in the first week of the course. This is an important consideration for situations in which concept tests are administered outside of a lecture class, such as during a laboratory or recitation section. They also investigated the effect of several different incentives on post-test scores: giving a few points to take the post-test regardless of performance did not seem to influence overall class average; giving students the option to replace an exam score with their post-test CSEM score resulted in a lower response rate but increased average post-test score (perhaps due to sampling from a higher-achieving group of students), and incorporating CSEM questions into the final exam produced a significant increase in post-test scores [383]. Henderson [425] observed only a small increase in scores when the FCI post-test was graded versus when it was ungraded, a difference of about half an item or 2%.
- *Web-based assessments*. Dancy and Beichner [344] compared performance on a paper-and-pencil version of the FCI to an animated version on the computer. They found that students were better able to understand the intent of the question when viewing it in animated format and this version was more indicative of their conceptual understanding.

Bonham [426] studied the reliability, compliance, and security of using Web-based assessments in an introductory astronomy course. The study found no statistical differences between administering an astronomy concept inventory and an attitude survey on paper or on the Web. The compliance rate increased with the amount of course credit given for completing Web-based assessments and the frequency of Email reminders. Most students abided by the request not to print, save, or copy portions of the assessments; a small number of students navigated away from the assessment to use other Web-based applications, suggesting they might have used Web-based resources. The study was unable to detect the use of non-Web resources for completing assessments (such as textbook, notes, or classmate).

- *Sensitivity to question context*. Stewart, Griffin, and Stewart [427] studied the effect of changing question context on the FCI on performance. They investigated eight types of question alterations: changing a concrete system to abstract (like changing a person pushing on a box to a force exerted on a block), changing the objects in the question to other realistic objects, removing redundant wrong answers, adding a picture, removing a picture, reordering multiple-choice answers, restructuring the order of a group of questions, and making responses symbolic instead of textual. The strongest positive effects were seen for removing a figure, changing the objects in the question, and removing questions from a group. Although altering a question sometimes vastly affected responses on that particular question, these effects tended to balance each other out and resulted in no overall significant change to the total score.

Steinberg and Sabella [50] report differences in students' answers when a FCI question is administered in multiple-choice format to when it is an open-ended question on an exam. Students typically scored higher on the exam questions, which were less "colloquial" than the FCI and often prompted them to draw the forces acting on objects. Although they observed a correlation between student performance on the two formats, there were several instances in which students gave inconsistent answers. The researchers caution educators to be wary that student understanding is more complex than simple correct or incorrect answers to multiple-choice questions, since changing the context of the test can alter responses. Dufresne, Leonard, and Gerace [49] also reported specific examples of how changing FCI questions alters students' answers. They describe one example in detail: the situation of two balls of unequal weight dropped from the same height. Although the majority of students surveyed responded correctly to the FCI question that the balls would reach the ground at about the same time, several of these students incorrectly believed that the forces acting on the balls were also equal. The possible influence of context familiarity was briefly addressed by Huffman and Heller [407], who suggested that the degree of familiarity with particular objects, such as hockey pucks or rockets, might affect a student's scores on those items. Context sensitivity has been found in other studies of physics concepts as well [428].

2. Comparing scores across multiple measures

- *Comparing inventory scores to other measures*. The widespread availability of physics concept inventories has prompted some researchers to question whether these tests can be used as placement exams. In other words, is a student's performance on a test like the FCI a predictor of their performance in an introductory physics course? Are strong math skills essential for learning physics? Several studies have compared scores on physics concept inventories to other performance measures such as SAT scores, grade point average, course grades, exam scores, math tests, problem solving, and Lawson's classroom test of scientific reasoning. Meltzer [385] reported significant correlations between normalized gain on a Conceptual Survey of Electricity (CSE) with pretest scores on an algebra or trigonometry math skills test for three out of four groups ($r = 0.38$, 0.46, and 0.30). Coletta, Phillips, and Steinert [383] reported strong correlations between normalized gain on the FCI and SAT scores ($r = 0.57$ at a high school and $r = 0.46$ at a university), and correlations between normalized gain on the FCI and scores on Lawson's classroom test of scientific reasoning ($r = 0.51$). Another study claims that math ability is a significant predictor of performance in a physics course [429,430]. Henderson [425] emphasizes that the FCI cannot be used as a placement test, because although high-achieving students (scoring higher than 60% on the pretest) usually get an A or B in the course, the test does not do a good job of predicting failure.

- *Comparing scores on multiple physics inventories*. One study has compared scores on the FCI to scores on the FMCE, which both cover one-dimensional kinematics and Newton's laws, but measure slightly different content and differ in format [387]. For example, the FMCE includes questions in pictorial and graphical format whereas the FCI is primarily in verbal and pictorial format, and the FCI is broader in that it includes some questions on two-dimensional motion with constant acceleration. The researchers reported that scores on the FCI were significantly higher than on the FMCE, and there was a strong correlation between scores on the two exams, even though some students showed inconsistent performance on the two tests. They also observed differences in gains on the FCI and FCME for different instructional strategies. For example, cooperative group problem solving showed equal gains on both tests, whereas using interactive lecture demonstrations showed higher gains on the FMCE than on the FCI.

There is also some inconsistency in observing correlations between diagnostic pretest scores and normalized gain. Hake [209] reported no correlation between FCI pretest scores and normalized gain, and Meltzer [383] reported no correlation between CSE pretest scores and normalized gain, whereas Coletta *et al.* [383] did observe a correlation between FCI pretest scores and normalized gain.

3. Comparing scores across multiple populations

- *Cross-cultural comparisons of concept inventory scores*. A comparison of concept inventory scores for Chinese and American students found that Chinese students scored higher on content knowledge tests like the FCI and BEMA, but scored at the same level as Americans on the Lawson Classroom Test of Scientific Reasoning [388,389].
- *Gender comparisons of concept inventory scores*. In introductory calculus-based physics courses for scientists and engineers at large-enrollment institutions, females are underrepresented and typically make up 20%–25% of the class [390,431]. Recent analyses of test scores in these courses indicate men significantly outperform women on some physics concept inventories, such as the FCI and FMCE, by an average of 10%–15%. There are conflicting views from different institutions regarding the extent to which interactive engagement instructional strategies can reduce and/or eliminate the gender gap in scores. Lorenzo, Crouch, and Mazur [391] claim that interactive engagement instructional strategies such as *Peer Instruction* at Harvard reduce or eliminate the preinstruction gender gap in FCI scores. In partially interactive classes the gender gap decreased from 13% before instruction to 7.8% after instruction, and in fully interactive teaching approaches it was essentially eliminated, decreasing from 9.2% to 2.4% [391]. In contrast, the University of Minnesota observed a persistence of the FCI gender gap on both pretest and post-test for cooperative group problem solving instruction [431]. The data examined included 40 fall term classes with more than 5500 students over ten years, showing a gap of $15.3\% \pm 0.5\%$ that decreased slightly to $13.4\% \pm 0.6\%$ although there was essentially no gender difference in course performance as determined by course grade ($1.5\% \pm 0.2\%$). The University of Colorado at Boulder observed persistence in the gender gap on FMCE scores from pretest to post-test but there were variations by instructor to the extent curricula were implemented, such as ConcepTests, online homework, and *Tutorials in Introductory Physics* [359]. In partially interactive classes, the gender gap averaged 10.9% preinstruction and 13.6% after instruction; in fully interactive, it decreased slightly from an estimated 11.2% to 10.0%. In a second-semester course on electricity and magnetism, there was a small gender gap on the BEMA pretest ($1.8\% \pm 0.5\%$), which increased to $6.1\% \pm 1.0\%$ on the post-test. There were some differences in course performance, where females outscored males on homework and participation, whereas males scored higher on exams, resulting in overall similar course grades [390,392]. Kost *et al.* [390] used logistic regression analysis to determine that the odds of scoring above 60% on the post-test were not different for males and females when controlling for pretest scores (FMCE, math, and attitude), suggesting that preparation and background are responsible for the persistence of the gap, not gender explicitly.

Some researchers have attempted to account for pretest gender differences by examining the preparation of males and females entering introductory college physics courses after high school. Kost, Pollock, and Finkelstein [390] found that although there is a correlation between years of high school physics and concept inventory post-test scores, gender differences could not be explained by whether students took high school physics because the gender gap was present for matched groups. Hazari, Tai, and Sadler [432] also found that background factors such as math SAT scores, calculus, and high school grades were equally predictive of success in university physics for both males and females. Factors that differentially affected male and female performance included high school pedagogy, assessment characteristics, and affective factors such as family attitude toward science and the father's encouragement [432–434]. Kost-Smith, Pollock, and Finkelstein [435] suggest that the underrepresentation of women in physics careers may be due to "an accumulation of small gender differences over time" in aspects such as retention, performance, and attitudes and beliefs such as personal interest (p. 1). A recent intervention involving a brief written exercise on the topic of "values affirmation" successfully reduced the gender gap on FMCE scores at the University of Colorado, suggesting that psychological interventions hold promise for addressing performance differences in science [436].

- *Gender comparisons on homework*. Kortemeyer *et al.* [120] determined that female students benefited more from the online homework system CAPA than male students as measured by final course grade. Further study of gender differences in use of the CAPA homework system found that more females than males reported that they consulted peers online or discussed problems with teaching assistants prior to attempting a problem (21% females compared to 9% males), whereas a higher fraction of men reported immediately attempting a problem after reading it (58% males compared to 39% females) [393]. There were no significant differences with respect to the time students reported spending on homework (females averaged 5.6 hours per week, males 5.1 hours per week). Males and females differed in the nature of their comments regarding multiple attempts to enter a solution. Male students stated that when they got a problem incorrect, they would use subsequent attempts to enter "random stuff" or manipulations of an equation, whereas female students stated that additional attempts gave them an opportunity to explore multiple approaches to the problem without stressing out about grades.

4. Course exams and homework

- *Multiple-choice course exams*. Scott, Stelzer, and Gladding [396] examined the reliability and validity of using multiple-choice midterm exams in an introductory calculus-based electricity and magnetism course. They compared students' multiple-choice exam scores to written explanations graded by instructors. The multiple-choice scores were found to be highly consistent with instructors' ratings of the students' explanations, indicating that multiple-choice exams "gave a statistically equivalent assessment of their understanding" (p. 5). Although multiple-choice exams are easier to grade than free-response questions or problems, they caution that high-quality multiple-choice questions are more difficult to create. Hudson and Hudson [395] also found a high correlation between performance on multiple-choice test questions and free-response questions, suggesting that multiple-choice tests can "provide essentially the same information as would hand-graded problems" (p. 838). More recent work has begun to explore whether computer-administered practice exams can help students better prepare for midterm exams [437].
- *Homework*. A common practice in introductory physics courses is to assign open-ended problems as homework; however, there are many different approaches for assigning and grading homework (if it is graded at all). For some courses, especially large enrollment courses, there is a concern that grading homework is time consuming and does not provide timely feedback to students. Bao, Stonebraker, and Sadaghiani [352] conducted research on a more flexible homework method in which students have some freedom to choose which problems to submit, and half of the problem solutions are available before the homework is due. Bao *et al.* [352] found that the majority of students preferred the new approach and believed that early access to the solutions improved their learning. In addition, students who used the new homework method scored significantly higher (5%–7%) on the Mechanics Baseline Test than students who used traditional homework methods.

One way to provide students with immediate feedback on their performance might be through the use of electronic scoring via web-based homework systems. As stated in the section on Problem Solving, there are several Web-based homework systems in physics including the "computer-assisted personalized assignment" system CAPA [120,191–193], Mastering Physics, WebAssign [560], and CyberTutor [121]. Studies comparing paper-and-pencil homework to Web-based homework have found mixed results, with some studies citing no differences in performance [194] and others citing improved performance with Web-based tutors or homework systems [195–198].

Lee *et al.* [121] used item response theory to analyze MIT students' performance on multiple attempts to answer a question on the Mastering Physics homework system. Among students who answered an item incorrectly and received tutoring help, 59% were able to get the solution on the second attempt, and another 2/3 of students solved it on their third attempt. Weaker students benefited more from the tutoring—students who consulted hints before submitting an answer (even though there was a penalty for this) showed a large increase of skill, 1.9 standard deviations better. This study also indicated that the hints provided by Mastering Physics improved performance on subsequent attempts of a problem, whereas simply indicating that an attempt was incorrect was insufficient to improve subsequent performance. In a different study, the same researchers found a negative correlation between incidence of homework copying and exam performance [438]. Students who repetitively copied homework performed worse on exam problems requiring analytic responses. However, there was no significant difference on normalized gain for the Mechanics Baseline Test concept inventory.

5. Rubrics for process assessment

A *rubric* is an assessment tool that indicates a score and the criteria that must be met to achieve that score. It can be composed of several subcategories (an analytic rubric) or can be holistic, and is often formatted as a table or grid. Rubrics are used in education research to provide criteria for objective scoring of student responses, but they are also used in classrooms to communicate expectations to students and help them self-assess their own work.

- *Assessing scientific abilities*. Etkina *et al.* [272] developed a set of tasks and rubrics to provide students with formative feedback on science-process skills, such as the ability to represent physical processes in multiple ways, design and conduct experimental investigations, and communicate their results. In classes that used the rubrics, students improved their abilities to design an experiment, devise a mathematical procedure to solve an experimental problem, and communicate procedures. In addition, these students performed better than students taught by traditional methods on multiple measures including use of diagrams in problem solving (60% compared to 20%) and performance on multiple-choice problems on final tests.
- *Assessing problem solving*. Most measures of problem-solving performance given in the classroom focus on the correctness of the end result or partial results, often comparing a student's written solution to features of an instructor solution. Scoring criteria and point values are often determined on a problem-by-problem basis. Docktor [205] developed a more descriptive problem-solving measure in the form of a general rubric to evaluate students' written solutions to physics problems along five processes: summarizing problem information and generating a useful description, selecting physics principles (physics approach), applying physics principles to the specific conditions of the problem, executing mathematical procedures, and the overall communication or logical progression of the solution. These skill categories were found to be consistent with research literature on problem solving, agree with instructors' beliefs, and are representative of the processes students engage in while solving physics problems (as measured by problem-solving interviews). The rubric is applicable to a range of problem types, physics topics, and solution formats.

Another effort to assess problem solving includes recent research by Marx and Cummings [203] and Cummings and Marx [439] to develop an assessment of students' ability to solve standard textbook-style problems. The draft instrument includes topics of Newton's laws, energy, and momentum, with 15 items (five from each topic).

Stewart and Ballard [440] assessed students' written presentation on open-response questions on four hourly exams. The problems were scored along three categories: math (symbols, relations, and numbers), language (sentences and words), and graphics (i.e., drawings). These presentation elements could explain 32%–26% of variance in test performance, but could not explain scores on a conceptual post-test. The incidence of language (sentences and words) and fewer numbers on a student's test was correlated with better performance.

6. Complex models of student learning

This section includes efforts to extend score-based analyses beyond pretest–post-test score comparisons.

- *Mathematical learning models that consider prior knowledge*. Pritchard, Lee, and Bao [400] present four mathematical models to describe student learning that differ in their assumptions about the influence of students' prior knowledge (as measured by pretest scores). The *tabula rasa* model assumes that change in performance (or test gain) is unaffected by prior knowledge, whereas the *constructivist* and *connectedness* models assume prior knowledge affects learning, and the *tutoring* model assumes efficient knowledge acquisition and therefore a high rate of learning. The researchers found that data from MIT had a lower value on the "connectedness" parameter than data from the University of Minnesota, indicating that MIT students were using memorization to answer questions on the FCI.
- *Model analysis*. Bao and Redish [399] propose a method for analyzing concept inventory scores to assess the level of confusion in a class. An individual student's responses are characterized mathematically as a vector, representing the probability they will apply a particular conceptual model. In addition to indicating how many students answered a question correctly, this model analysis approach indicates a "level of confusion" or the state of a class's knowledge.

- *Learning progressions*. A few researchers have investigated the progression of learning throughout an introductory electricity and magnetism course, looking at the "peaks" of high scores and "decays" or decrease in performance on conceptual questions that occur between the administrations of pre- and post-tests [401,402]. They observed a rapid increase in performance that occurred after homework assignments (not after lectures or laboratories) but decreased a few days later. The researchers also observed interference effects, where learning about electrical potential impeded students' performance on questions related to the vector nature of electric fields, but performance recovered later when learning about magnetic fields.

E. Strengths and limitations of assessment research

1. Strengths

- The topic of assessment using multiple-choice inventories lends itself to large-scale, quantitative studies. Some of the research studies or meta-analyses cited in this section include test scores for upwards of hundreds or even thousands of students, which provide confidence in the interpretation of results.
- The widespread availability of concept inventories in physics has facilitated their use in classrooms both nationally and internationally. Instructors have several tools they can use to evaluate the effectiveness of their instruction, and can compare published results from other institutions to the scores of their students.
- From the perspective of making physics instruction more effective, concept inventories historically have served a pivotal role in catalyzing dialogs among physics professors who were initially very skeptical that students (especially the ones they taught) could hold erroneous conceptual notions on fundamental physics ideas following instruction.

2. Limitations

- Many studies who report data on inventories do not give sufficient information about when they were administered and incentive conditions, which can have a significant impact on results (see Ref. [382]).
- It is typical for studies to report average scores for a particular class or group of students, but many do not indicate the distribution of scores (e.g., histograms) or separate the results for underrepresented minorities.
- Some research on conceptual tests, especially the Force Concept Inventory, have produced contradictory results. Although the test authors claim the questions and choices were developed based on students' open-ended responses to questions, subsequent studies have found that changing the format of test questions or the context of the questions can dramatically alter students' responses [49,50,427].
- Although several sophisticated statistical methods exist for analyzing test scores and have even been described in the PER literature (see Ref. [381] for a review), these procedures have not become widespread, possibly due to a lack of familiarity with those statistical methods among PER researchers.

F. Areas for future study

- *Homework and exams*. There is very little research on homework, such as how much to assign, when to collect it, and how to grade it (if at all). In addition, studies comparing paper and Web-based homework have conflicting results, with some stating that Web-based homework is better than paper, and some stating that they result in equal learning. There is also a need for research on the type and quality of feedback provided to students, such as what kind of tutoring is most helpful in Web-based homework systems. Similarly, there is a need for research on exams, including the optimal frequency of administering quizzes or tests, what types of questions are best (short answer, essay, multiple choice, problem solving, etc.), and whether comprehensive final exams have merit. Educational psychology research on the "testing effect" suggests that more frequent testing improves learning and retention of knowledge, but this has not been studied in science classes [441]. In addition, there exists little or no research on how students prepare for physics exams, or the administration and use of practice exams.
- *Development of additional assessments*. As indicated by the list in Ref. [377], there are several concept inventories under development and only about half of the inventories in physics are published. However, there is some need for alternate forms of assessment (other than concept inventories), such as tests for problem solving and mathematics, or measures of expertise such as categorization tasks. A multiple-choice problem-solving assessment for textbook-style problems is in development [398,439]. The PER community also lacks a consistent math skills exam; several institutions use their own, self-made exam to assess students' skills. Once such measures exist, additional research could investigate the relationship between scores on concept tests and problem-solving performance or the relationship between conceptual knowledge, mathematical skills, and problem solving.
- *Gender and other population differences*. In general, there is a need for studies in PER to report data that are separated for particular populations, such as underrepresented minorities. Research reported in this section indicates that females score lower on average than males on some concept inventories like the FCI

and FMCE by as much as 10%–15%. The reasons for this gender gap in performance is not well understood and needs further investigation. For example, does the gender gap also exist for electricity and magnetism tests (CSEM or BEMA)? Does it exist for all introductory course populations (nonscience, premedicine, engineers, majors, etc.)? How are these results related to the concept of *stereotype threat* [436]?

- *Testing conditions*. There is a need for additional research on testing conditions, such as the influence of when a concept inventory is administered, how it is administered, and incentives provided to students. Ding *et al.* [382] showed for the CSEM test that even a few days or lectures can influence pretest scores, and it would be useful to have similar information for other tests. There have been very few studies mapping the progression of learning, decay, and interference that occur during a course and "between" pre- and post-tests [401,402]. Also, the issues involved in administering concept inventories electronically has only been minimally explored [426].
- *Consistent procedures for analyzing and reporting scores*. The "normalized gain" is a commonly reported measure for comparing pretest and post-test scores across populations [209], but the statistical origin for this measure is unclear and alternatives have been suggested (such as normalized change). It is unclear why normalized gain is still favored, and the PER community should reach an agreement about how to analyze and report scores from concept inventories.

In addition, the PER community would benefit from handbooks that outline procedures for the development and testing of concept inventories, similar to the guidance in Ref. [378]. These handbooks should also incorporate the recent interpretations of validity and reliability from the field of quantitative measurement set forth by [404] standards for educational and psychological testing [403]. Many test developers still follow the validity "types" paradigm, which is inconsistent with the *Standards for Educational and Psychological Testing*.

VI. COGNITIVE PSYCHOLOGY

Cognitive psychology is a field of psychology within cognitive science [442] focused on studying mental processes such as problem solving, memory, reasoning, learning, attention, perception, and language comprehension. It emerged about 60 years ago, and in these ensuing decades the field has matured in terms of theories, methodologies, and traditions. PER, on the other hand, is not only a younger field of study, but it is also much more applied than cognitive psychology in two regards: (1) PER focuses on physics, whereas the content topic of research in cognitive psychology is usually not important since the focus is on exploring cognitive processes, and (2) quite often, but not always, a primary goal in PER studies is the improvement of teaching and learning in physics—the improvement of psychology instruction is almost never the goal in cognitive psychology research.

Until recently, PER researchers were trained as research physicists in traditional physics areas and then made the transition to education research and learned to do PER by on-the-job experience. Early PER practitioners had little or no formal training in cognition or education, although sometimes they were mentored by other physicists interested in or practicing PER. What this means is that "cognitive studies" in PER, namely, studies exploring cognitive processes related to physics learning or problem solving, have not been common. Although there is some overlap between cognitive studies, conceptual understanding, and problem solving, articles included in this section may not have findings that are directly applicable to improving classroom instruction, yet they carry important implications for physics learning. Early research on conceptual understanding and problem solving was generally focused on investigating students' difficulties learning physics and then designing instructional strategies and curricula to effectively address those difficulties, whereas cognitive research does not have an explicit focus on instruction.

A. Research questions

The research questions investigated under the cognitive psychology category generally fall into the following categories.

1. Knowledge and memory

How is knowledge organized and accessed or activated? In what ways is knowledge activation influenced by framing and context? What is the role of specific examples in knowledge storage and retrieval? Do experts and novices differ in their organization of knowledge and/or memory [84,144,151,443,444]?

2. Attention

Do experts and novices attend to different aspects of problems [78,141,445]? *What do novices attend to when learning from worked-out problems* [143]?

3. Reasoning and problem solving

Do experts and novices approach problem-solving tasks differently [78]? *How do experts and novices categorize physics problems* [84,85]? *What is better for developing problem-solving skills, working problems on one's own or studying worked-out examples* [91]?

4. Learning and transfer

What type of knowledge do students transfer and how does the knowledge transferred depend on context [446–448]? *What facilitates and impedes transfer of learning across physics-similar but surface-different contexts* [135,233]? *What are possible mechanisms for knowledge transfer, how can they be used to analyze or interpret interview data, and how does a learner decide that transfer has been achieved* [449,450]? It has been known for some time that transfer of learning across contexts is difficult to achieve (see Refs. [451,452] for reviews). Recently a number of studies have explored various aspects of transfer in physics contexts.

B. Theoretical frameworks

Given that PER is a relatively young field of study, the theoretical frameworks used in cognitive studies are borrowed from cognitive science and made to fit the occasion. Generally, theoretical issues in PER studies analyzing cognitive processes can be characterized in the following ways:

- *Expertise and expert performance*. The goal of studies of expertise is to gain insights on effective reasoning and performance. Although it is recognized that expertise research can help us think about how to make the transition from novice to expert more efficient through the design of instructional interventions, turning novices into experts is not necessarily the primary goal of this body of research. The study of expert performance (for reviews, see Refs. [35,453]) has revealed various insights about the way that experts store and apply knowledge, some of which will be summarized in Sec. VI D. Research on expert performance has allowed the characterization of knowledge in memory for both experts and novices and how experts and novices use their knowledge to solve or reason about problems. Studies of expertise typically compare the performance of experts and novices on physics tasks such as problem solving or problem categorization.
- *Transfer of learning*. Transfer studies focus on exploring what knowledge is deployed to reason about a situation and how the particular knowledge deployed depends on context—that is, how and whether or not changes in context of surface attributes of situations that do not change the underlying structure affect what knowledge is brought to bear and/or how it is used in reasoning. (See Ref. [454] for various studies of this type.)
- *Metacognition*. Metacognition, which refers broadly to thinking about one's own thinking [455], is thought to impact learning in that being reflective about one's own thinking processes can improve learning and retention due to deeper processing of the information [87] (for a broad overview, see Ref. [35]; for a recently developed assessment of physics metacognition, see Ref. [456]).
- *Construction and use of categories*. Although it may seem obvious, proficient problem solving is largely the ability to categorize new problems into types that the solver knows how to solve [85,87,153]. There are several theories regarding the construction and use of categories in learning, such as the prototype view, exemplar view, and rule-based views (for a review of concepts and categories, see Refs. [457,458]).
- *Attention during processing of information or problem solving*. When we look at a situation (e.g., a picture of diagram) or read a problem, we need to be selective in what we attend to since short-term memory is limited (see Ref. [445] for a brief review and additional references).
- *Knowledge organization and memory.* As we gain expertise in a subject, that knowledge is organized in memory in ways that promote quick access or retrieval. The knowledge organization of experts is believed to be hierarchical (see Refs. [35,84,151]). Experts' knowledge is bundled with relevant contexts in which the knowledge can be applied and with procedures for applying it. With their extensive experience in problem solving, experts develop *problem-type schemas*, which consist of representations of problem categories together with appropriate solution procedures [144,443,444].
- *Language*. The language people use shapes the way they think [459]. The language physicists use to describe and explain physical phenomena is largely metaphorical. The key here is that metaphors used in physics discourse were originally analogies with inherent limitations [460]. As those analogies were used by physicists, they turned into metaphors that students perceive as literal [460,461]. For example, current flows, energy barrier, force acts. In addition, the grammatical structure of physics statements often leads to ontological confusion [461,462]. For example, a particle *in* a potential well, weight *of* an object, heat transferred *into* a system.
- *Learning by analogy and analogical reasoning*. Studies of use of analogy in making sense of phenomena and solving problems indicate analogy is commonly used by people—that is, they tend to look for a similar problem that they know how to solve or that is already worked out and map how the solution to the analogous problem can be applied to solve the new problem. For theories of analogy and analogical transfer, see Refs. [128,130]. Theories of analogy use for understanding concepts suggest ways in which analogical scaffolding can be a useful tool for guiding students' use of representations to understand phenomena [463–465].

C. Methodology, data collection or sources and data analysis

Various methodologies have been used in physics cognitive studies that are commonly used in cognitive psychology.

1. Contexts

Cognitive studies in physics education are difficult to conduct on a large scale, so they typically take place in a small controlled environment outside of the classroom.

2. Participants

Participants in cognitive research studies are typically students who know some physics, often paid volunteers who have taken one or two introductory courses in physics. The "experts" in expert-novice studies are often physics faculty or graduate teaching assistants who are experienced with teaching introductory courses.

3. Data sources and analysis

Research on cognitive processes often requires descriptive, qualitative sources of data, such as interviews or self-explanations, to infer students' reasoning processes as they engage in a task. Studies on attention might employ sophisticated data collection tools such as eye-tracking devices to record gaze patterns. Studies of problem categorization can utilize a variety of task formats, as described below.

- *Verbal reports from clinical interviews*. One common methodology for studying cognitive processes in physics education research, and many other disciplines as well, is to conduct interviews of subjects engaged in cognitive tasks, the hope being that verbal utterances reflect, at least in part, what the mind is doing [466–468]. There are differing views on what can be ascertained from subjects' verbal reports. One view is that subjects should simply be instructed to "think aloud" as they perform a task, since it is believed that doing so does not change the sequence of thoughts that would occur if the subject were to perform the task in silence [453]. Doing more, such as asking subjects why they make particular choices in performing a task [469] or asking subjects to self-explain while reading and attempting to understand a solution to a problem [470], may change the accuracy of observed performance in studies of expertise [467]. Others argue that cognitive clinical interviews are derivative of naturally occurring forms of interaction and are thus ecologically valid [466], or that the quality of the information obtained is partly dependent on students' perception of the nature of the discourse interaction [121]. In PER, interview techniques are often used that attempt to scaffold or cue students to previously learned knowledge (see Ref. [450] for examples and Refs. [5,471]); these interview techniques are useful for learning about ways of structuring effective instructional strategies, as in design research [472,473].
- *Self-explanations*. Asking subjects to explain to themselves out loud the important features of a problem solution as they study it has been used to study the impact of metacognition on learning [87,474].
- *Eye tracking*. Eye tracking methodologies (tracking eye gazes by capturing or recording position and duration during cognitive tasks) are quite prevalent in cognitive psychology and are used to study what people are paying attention to without disrupting reasoning processes. There is agreement that where the eyes look designates what people are paying attention to, so by following eye fixations one can unobtrusively observe what is being attended to during a cognitive task and thereby draw conclusions about what the mind is processing (see Ref. [475] for a good review; see also Ref. [476]).
- *Change events or flicker technique*. Methodologies are beginning to be used in PER that borrow from those used in visual cognition. For example, change event methodologies are used to study a phenomenon known as change blindness. Studies of change blindness [477] in psychology typically involve an unsuspecting subject engaged in a task (e.g., giving directions to another person, in this case the experimenter, on how to get to a location on a map), and then unbeknown to the subject something is changed (e.g., the experimenter holding a map is changed to another person holding the same map while the subject is momentarily distracted) to see if the subject notices. The flicker technique alternates between two images on a screen shown for 200 ms with a 100 ms black screen separating them; the two images are slightly different and the time to detect the change is measured to study saliency of the feature being changed. These methodologies attempt to ascertain what is considered salient and thus attended to, and what is considered nonsalient and neglected, or minimally processed, in a situation or task.
- *Problem categorization*. Problem categorization tasks are common in physics cognitive studies. These tasks ask subjects to identify problems that are solved with a similar approach. Three types of categorization tasks have been used: (a) problem sorting, where subjects place into piles problems that are solved similarly [84,478], (b) problem comparison, where subjects are shown either two or three problems together and asked to identify which are solved similarly [85] or are asked to rate how similar problems are on a Likert scale [139]. In the two-problem task, subjects are usually asked to state whether or not the problems are solved with a similar approach and to state a reason why. In the three-problem task, a model problem is used and two comparison problems, and the subject is asked to identify which of the two comparison problems is solved most like the model problem; this task allows experimenting with competing influences on categorization, such as matching a comparison problem to the

model problem on either surface attributes, solution principle or procedure, both, or none. The third categorization task is (c) multiple choice, where a problem is given followed by five choices for the major principle that would be used to solve it and the task is to select the appropriate one [117].

D. Findings

1. Knowledge and memory

- *Knowledge activation depends on framing and context.* Framing in anthropology and linguistics [479–482] and more recently science education [135,483] is used to describe how an individual makes sense out of context and decides what is important and what to pay attention to. For example, if framed in terms of a "school" context, then students tend to draw on formal knowledge, whereas if framed in terms of an out-of-school context, more intuitive or experiential knowledge is brought to bear [135]; also, nearly identical tasks can be framed in ways that lead to very different physics knowledge being brought to bear [447]. In short, transfer of *appropriate* knowledge depends highly on framing and context.
- *Knowledge is tied to specific examples.* Since it is difficult for people to think abstractly, a common approach in teaching physics and other sciences is to present the abstract concepts and then follow with a specific worked example to illustrate how the concepts are applied to solve problems. In cognitive science, Ross [131,484] has shown that if following a worked example students are given a subsequent problem to solve that is somewhat similar to the worked example, then the features of the example's context become a strong cue for students to use the same method or procedure that they learned in the worked example. That is, students' understanding of the concept becomes bound up in the particulars of the example that is used to illustrate the principle [132,133,164]. This phenomenon is known as the "specificity effect," and has been demonstrated in PER [485,486].
- *Expert-novice differences in knowledge organization and memory.* Experts chunk related knowledge in clusters, with chunks containing concepts as well as procedures for applying them and contexts or conditions under which concepts and procedures can be applied. When recalled, chunks are effective for reasoning and solving problems since they contain multiple knowledge, procedural, and contextual units. In contrast, novices' knowledge in memory is more amorphous and does not have the efficient chunked organization of experts' knowledge; when reasoning about a physics situation or solving a problem, novices recall units (e.g., equations, similar contexts) individually, making the search and thinking processes more hit or miss and simultaneously increasing memory load by creating high demands on short-term memory [78,79,81,83,92,93] (for a review, see Chap. 2 in Ref. [35]).

2. Attention

- *Attending to relevant features in physics diagrams.* Using a change blindness paradigm, Feil and Mestre [445] investigated whether novices and experts noticed surreptitious changes made to physics diagrams, finding that experts were able to identify changes that affected the underlying physics whereas novices did not. Other studies [141,487] have used eye tracking to explore how experts and novices allocate visual attention to physics diagrams in cases where critical information needed to answer a question was contained in the diagram. They found that experts spend more time looking at thematically relevant areas of diagrams than do novices. Other eye-tracking studies have touched on issues peripherally related to PER, such as investigations of troubleshooting of malfunctioning circuits [488], of comprehending malfunctioning mechanical devices [489], of comprehending how mechanical systems work [490], and of the relationship between spatial visualization and kinematics problem solving ability [491].
- *Attention in problem solving.* Additional eye-tracking studies using physics tasks are beginning to shed light on what students, and experts, attend to while performing cognitive tasks. One example is a study by Rosengrant, Thomson, and Mzoughi [492] in which a small number of novices and two experts answer questions about circuit diagrams while their eye gazes are being recorded. Experts showed a more "global" approach to processing information about the circuits whereas novices focused on "local" aspects such as individual resistors to combine in series or parallel. Another study [143] explored what novices look at while studying example problem solutions in introductory physics, comparing the time spent on mathematical information (equations) to the textual or conceptual information. Although students spent a considerable portion of time (about 40%) looking at textual or conceptual information, little conceptual information that students read was retained, suggesting that students may not have been prepared to learn conceptual knowledge and that this ability may lag ability to learn mathematical procedures for solving problems.

3. Reasoning and problem solving

- *Expert-novice differences in problem solving processes.* We refer the reader to the bullet, "Problem solving approaches," within Sec. III D 1 for the salient differences between expert and novice problem solving.

- *Problem categorization*. As discussed in the "Categorization" bullet within Sec. III D 1, experts categorize physics problems according to the major concept or principle that can be applied to solve them, whereas novices rely much more on the surface attributes of the problems to categorize them [84,85,145]. Another recent study showed contrasting results to those of Chi *et al.* by demonstrating a large overlap between the categorization performance of calculus-based introductory students and graduate students, suggesting that there is a wide distribution of expertise in mechanics among introductory and graduate students [493].
- *Learning from worked examples*. In the "Example structure and the worked-example effect" bullet in Sec. III D 2, we discussed the worked-example effect, namely, research that suggests that studying worked examples is more effective for learning to solve problems than actual practice in solving problems. There is evidence that studying worked examples can develop schemas better than problem-solving practice due to the lower levels of cognitive load that go with studying worked examples, which in turn leaves more resources in short-term memory to extract and make sense of solution strategies [494,495]. These findings have been replicated in other domains, such as statistics [496] and geometry [497]. The worked-examples effect holds promise for designing pedagogical interventions aimed at improving problem-solving skills in physics.

4. Learning

- *Self-explanations while studying worked examples*. Research conducted by Chi *et al.* [87] on how students study worked-out example solutions in textbooks to learn topics in mechanics found that successful students explain and justify solution steps to themselves (self-explain) to a greater extent than poor students. The quality of the explanations also differs; good students refer to general principles, concepts, or procedures which they read in an earlier part of the text, and examine how they are being instantiated in the current example [498].
- *Analogical reasoning*. Recent work suggests that use of analogies during instruction of electromagnetic waves helps students generate inferences, and that students taught with the help of analogies outperformed students taught traditionally [463,464]. Further, blending multiple analogies in instruction generated better student reasoning compared to instruction that did not use blends or that used standard abstract representations to convey the wave concepts [465].
- *Language and symbolic forms*. Research on the role of language showed that making its metaphorical nature transparent for the students helps them apply concepts and solve problems [499]. For example, instead of writing forces as W (weight) or T (tension), students benefit when labeling each force with two subscripts to identify two interacting objects—F_{EonO} (force exerted by Earth on object) or F_{RonO} (force exerted by rope on object). Another example is heat. To help students understand that heat is a process of energy transfer and of energy itself, the term heat can be substituted with "heating." Some PER-based curriculum materials are using this new language [319].

E. Strengths and limitations of cognitive research in PER

1. Strengths

- PER cognitive psychology research builds upon prior research from cognitive science and hence has a corpus of work from which to draw for methodologies and theoretical frameworks.
- This type of research helps us learn about human cognition in a complex domain (physics) that requires substantial prior knowledge, reasoning, and problem-solving skills. As such, this type of research can lead to instructional insights and the eventual design of effective instructional innovations.

2. Limitations

- The findings from many PER cognitive research studies often are not immediately applicable to improving classroom practice.
- Conducting these types of studies most often requires a pool of subjects who have taken at least one or several introductory courses since they need to possess some minimal knowledge of physics. This is to be contrasted with most cognitive psychology experiments that do not require domain knowledge.
- Classroom-based cognitive studies involving large numbers of subjects tend to be rare because it is much more difficult to control for extraneous variables that might affect outcomes in these settings; hence, the majority of cognitive PER studies are done in carefully controlled laboratory settings. This may mean that results from a lab-based study may yield somewhat different results if one attempts the same study in a realistic setting at a larger scale.

F. Areas for future study

There are many research areas in cognitive science whose potential has not been explored for learning about physics cognition. For example, psycholinguistics is a large and thriving field of study in cognitive science that draws heavily upon eye-tracking methodology, yet studies of students' understanding of the language of physics is only now beginning to be explored [461,462]; similar situations exist with areas such as memory and perception research, and with promising methodologies, such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), where virtually no PER that we are aware of exists (the exception being a fMRI study of physics misconceptions in Ref. [25]). Although subfields (e.g., visual cognition) and methodologies (e.g., eye tracking) from cognitive science are beginning to find their way into PER, the number of existing studies is extremely small. As PER matures and broadens its education of Ph.D. candidates to include more training in cognitive science, we will see an increase in cognitive studies in PER.

VII. ATTITUDES AND BELIEFS ABOUT TEACHING AND LEARNING, AND THEIR DIRECT AND INDIRECT IMPACT ON CURRICULUM CHOICES, CLASSROOM PRACTICES, AND STUDY HABITS

Students have attitudes, beliefs, and expectations about learning physics that can impact the way they behave and perform in a physics course [221,258,500]. For example, a common student belief is that physics is made up of several unrelated pieces of information. As a result, many students approach physics by memorizing formulas without connecting them to a broader understanding of the underlying concepts and principles. This section summarizes frameworks that have been developed to describe students' beliefs and multiple-choice attitude surveys to explore changes that result from instruction.

In addition to students' attitudes and beliefs, the beliefs that instructors (both professors and graduate teaching assistants) have about how students learn can impact their instructional decisions and classroom interactions. Research from PER and science education in general indicates that oftentimes instructors' beliefs and practices are inconsistent with each other. This section also summarizes research on physics instructors' beliefs, their decisions to adopt research-based curricula, and how PER-based curricula are actually implemented in the classroom. A discussion of research on general instructional practices such as traditional lectures or laboratories has not been included in this review; there is an abundance of articles on these topics in the *American Journal of Physics* dating back to the 1930s and the *Journal of Research in Science and Teaching* dating back to the 1960s. Finally, we include in this section curricular and instructional decisions by instructors as well as study decisions by students that, although perhaps not directly attributable to attitudes and beliefs, are at least indirectly influenced by them.

A. Research questions

Key areas of research on this topic include student attitudes and beliefs about learning physics, instructor beliefs about how students learn physics, instructors' decision-making processes, the ways in which instructors implement reformed curricula and instruction, and research on the attitudes and practices of teaching assistants.

1. Student attitudes and beliefs about learning physics

What attitudes, beliefs, and expectations do students have about learning physics? How do students' attitudes and beliefs change as a result of physics instruction? What instructional strategies are effective for promoting productive attitudes and beliefs? There exists research on students' beliefs about learning physics [42,501,502], and on the impact that their beliefs have on their performance as measured by concept inventory scores and course grades [221,500,503,504]. Research also exists on instructional strategies designed to explicitly promote productive attitudes and beliefs [224,505]. Below we review several surveys that have been developed to measure student attitudes and beliefs before and after instruction, including the Maryland Physics Expectation Survey (MPEX) [506], Views About Science Survey (VASS) [503,507], Epistemological Beliefs about Physical Science (EBAPS) [224,508], Colorado Learning Attitudes about Science Survey (CLASS) [509], and Attitudes and Approaches to Problem Solving survey (AAPS) [510].

2. Faculty beliefs and values about teaching and learning

How do instructors believe students learn physics (e.g., physics problem solving)? What prior knowledge or skills do instructors expect of their students [511]*? How do instructors view their role as a teacher? What factors contribute to their instructional decisions* [512]? Recent research has explored beliefs and values that faculty have about teaching and learning physics in general [512] and problem solving specifically [201,511,513], including instructors' reasons for choosing problem features [514]. (For a review, see Ref. [515].)

3. Instructor implementations of reformed curricula

How prevalent is the use of research-based instructional strategies by physics faculty in the United States? How is [research-based strategy X] typically implemented? What is the nature of faculty-student interactions during Peer Instruction? What instructional practices foster an interactive, collaborative classroom environment? Some very recent studies have explored faculty adoption of PER-based instructional strategies and materials [368,369,516], how PER-based strategies like *Peer Instruction* are actually implemented in the classroom [517,518], and student perceptions of classroom reforms.

4. Teaching Assistants

What attitudes and beliefs do teaching assistants have about student learning? How do teaching assistants view *their role as recitation and/or laboratory instructors? How do teaching assistants impact the success of reformed instruction?* A few (but growing number of) studies have explored the attitudes and beliefs that teaching assistants have about student learning, and how those relate to their "buy-in" to reformed instructional strategies and influences their instructional behaviors [519,520]. Other research explores how instructional style (specifically TA coaching) influences student learning during recitation [521].

B. Theoretical frameworks

Research studies of attitudes and beliefs about teaching and learning do not always explicitly identify theoretical frameworks. Studies might simply cite previous research findings from education research or psychology as a basis for their work. Some of the PER research on faculty beliefs takes this approach (see, for example, Ref. [511]). One exception is research on teaching assistants [519,520], which identifies "framing" theory from anthropology and linguistics as a basis for understanding how TAs make sense of their teaching experiences. The following examples are frameworks that have been developed and published by researchers in physics education.

1. Student attitudes and beliefs

Hammer [501,522] developed a framework for characterizing students' approaches to learning physics, which had its basis in education research on learning "orientations" in addition to interviews he conducted with physics students. The three categories of his framework are student beliefs about the structure of physics (made up of pieces versus coherent), beliefs about physics content (formulas versus concepts), and beliefs about learning physics (receiving knowledge versus actively making sense of new information). Hammer's framework was later used to design multiple-choice attitude surveys.

2. Instructional practices and conceptions

Dancy and Henderson [523] introduce a framework for describing instructional practices and conceptions. The framework was developed from research literature, discussions within the PER community, and interviews with physics faculty. The framework's ten categories of conceptions include learning view (transmissionist versus constructivist); beliefs about expertise; beliefs about knowledge; beliefs about the nature of physics; conceptions of their role as an instructor; the role of the school; views of students' capacity to learn; desired outcomes (goals) for students; diversity; and scientific literacy. The ten categories of practice identify behaviors typical of traditional instruction compared to alternative instruction and include interactivity (passive lecturing versus active student conversation); instructional decisions (made by teacher alone versus with students); knowledge sources (students receive versus construct knowledge); measures of student success (against preset standards versus individual improvement); learning mode (competitive versus cooperative); motivation (external versus internal); assessments (knowledge based versus process based); content (facts and principles only versus also teach thinking and problem solving skills); instructional design (knowledge driven or student driven); and problem solving (formulaic versus creative).

C. Methodology (data collection or sources and data analysis)

1. Contexts

Studies involving students, such as identifying students' attitudes and beliefs about learning physics and the development of attitude surveys, have primarily taken place in conjunction with introductory physics courses. In contrast, studies involving instructors (faculty or teaching assistants) often employ interviews that take place in a controlled experimental setting outside the classroom or surveys that are administered online.

2. Participants

Studies of students' attitudes and beliefs about learning physics have typically involved students enrolled in an introductory course. Studies of instructor's beliefs about learning and teaching physics often involve faculty (from a variety of institutional types) who are currently teaching a course or who have substantial experience teaching physics. A very small number of studies involve graduate teaching assistants in their first or second year of graduate school.

3. Data sources and analysis

- *Clinical interviews*. Research to identify the attitudes and beliefs of students and instructors typically begins with in-depth interviews that are audio and/or video recorded. The transcribed statements are coded by themes that are determined either by the researchers or from existing theoretical frameworks. These methods were used by Hammer [47,501,502] to describe students' beliefs about learning physics, and by several researchers investigating instructor's beliefs [201,511–514]. Sometimes these instructor interviews include specific *artifacts* such as sample student or instructor solutions that provide a basis for discussion.
- *Surveys*. Web-based surveys have been used in research on faculty knowledge and use of research-based instructional strategies [368,369]. The results are analyzed using descriptive statistical measures, such as reporting the average fraction of survey respondents who reported using a particular research-based instructional strategy.
- *Classroom observations*. Observations of instruction have been used in studies of teaching assistant

behaviors [519,520]. These videotaped observations are analyzed along several different measures, but typically attempt to characterize the nature of student-student and teacher-student interactions.

- *Analyzing students' written work*. In some studies, students' degree of "sophistication" in their beliefs about learning physics are analyzed by scoring their written responses to reflective questions [221,524]. After the researchers have obtained such a quantitative measure, it is analyzed further by statistical correlations with other course measures such as concept inventory scores, course exams, homework, or final course grades.

D. Findings

1. Student attitudes and beliefs about learning physics

This section discusses research on student attitudes and beliefs and the impact these have on learning physics. In these studies, a student's approach to learning is discussed using a variety of terms, such as attitudes, beliefs, values, expectations, views, personal interest, learning orientation, motivation, and epistemological beliefs. Our use of the terms "attitudes and beliefs" is intended to broadly encompass this wide range in terminology. When the term *epistemology* is used, it should be interpreted to refer to "beliefs about what constitutes knowledge in physics and how, as a student, one develops that knowledge" [501]. Some researchers have collectively referred to students' attitudes, beliefs, and assumptions about learning physics as their "expectations" upon starting a course [506]. For a more comprehensive review of research on epistemology and relationship to teaching and learning, see Refs. [525,526].

- *Epistemological beliefs in physics*. Hammer [42,501,502] developed a framework to characterize students' attitudes and beliefs about physics in terms of three scales: pieces coherence, formulas concepts, and by authority independent [205,502]. The *coherence* measure refers to a student's view of the structure of physics as consisting of several unrelated pieces of information (unfavorable) or as a connected system that is relevant to everyday experiences in the real world (favorable). A student's belief about *concepts* refers to whether they focus on memorizing and blindly applying equations (unfavorable) or attempt to understand the concepts that underlie the equations (favorable). A student with a favorable attitude on the *independence* measure actively attempts to construct their own understanding, whereas a student who just receives information from an authority (e.g., instructor or textbook) without question has an unfavorable learning attitude.

Research by Elby [527] extended Hammer's ideas to explore students' expectations for a physics course and the relationship between epistemological beliefs and study habits. He determined that although most students know what it means to understand physics deeply and how to go about learning physics in a deeper way, they do not engage in productive study habits because they do not see this level of understanding as essential for obtaining good grades in a physics class.

- *Surveys of student attitudes and beliefs.* In order to better characterize student attitudes and beliefs before, during, or after a course, researchers developed an assortment of attitude survey instruments. One of the first surveys developed was the MPEX, a 34-item Likert-scale questionnaire (disagree-agree) that probes six dimensions of students' expectations about learning physics [506]. The first three dimensions match those identified by Hammer [501]: independence, coherence, and concepts; and there are three additional dimensions: reality link (view physics as relevant to real life), math link (math is a way to represent physical phenomena, not just a tool for calculation), and effort (what students perceive they need to do to be successful in a physics class). Items are scored as being favorable if they agree with how an expert physicist would answer the item, or unfavorable if the response disagrees with expert responses on that item. Upon administering the test to physics students at six universities or schools, they consistently observed a decline in students' attitudes following a semester of instruction [506]. They suggest that some physics courses might actually reinforce students' inappropriate approaches to learning by placing an emphasis on memorization rather than a deeper conceptual understanding of the subject matter.

The VASS [503,507] is another survey instrument that probes students' views of learning science along six dimensions: three scientific dimensions (structure or coherence, methodology or use of mathematics, and validity of scientific knowledge) and three cognitive dimensions (learnability or personal effort, reflective thinking, and personal relevance). They use scores on the VASS to classify someone as having expertlike views, mixed (transitional) views, or folk views. For the populations studied, high school and college students scored similarly on the survey, and very few (no more than 10%) were classified as having expertlike views.

The EBAPS is formatted in a slightly different way than other surveys of attitudes and beliefs [224,508]. Although part one of the survey includes Likert-style statements (disagree-agree), part two of the survey includes multiple-choice questions in which students must decide which statement they agree with most. The five subscales of the EBAPS include structure of scientific knowledge, nature of knowing and learning, real-life applicability, evolving knowledge, and source of ability to learn (natural ability versus effort based). In contrast to the MPEX, the EBAPS survey attempts to only assess epistemological beliefs separately from expectations about learning physics.

The most recently developed attitude survey is a 42-item questionnaire adapted from previous work, called the CLASS [509]. It differs from previous surveys in that the dimensions are constructed from a factor analysis of students' scores rather than *a priori* expectations of question groupings, and it explicitly considers attitudes toward problem solving. The eight categories include real-world connection, personal interest, sense making or effort, conceptual connections, applied conceptual understanding, problem solving general, problem-solving confidence, and problem-solving sophistication. There are also six items that do not fit into these categories and are not scored. One study of scores on the CLASS [528] asked students to respond how they think a physicist would answer the questions, and how they personally would answer it. When answering from the perspective of a physicists, students were very accurate at selecting the "favorable" attitude responses; however, they did not answer in the same way when responding with their own personal beliefs.

Gire, Jones, and Price [529] used the CLASS survey to measure the epistemological development of physics majors. They administered the attitude survey to physics majors who were at different stages in their academic career, including students in their first, second, third, or fourth year of college, and graduate students in physics. They found that among students in their first year of college, the physics majors had more "expertlike" responses on the CLASS than did their introductory physics course peers, who were primarily composed of engineering students. Students in years one, two, and three had similar overall fraction of favorable scores on the CLASS, whereas the scores were slightly higher for students in their fourth year or in graduate school. Another study showed that students at various universities using the *Physics by Inquiry* curriculum showed a positive shift in CLASS scores [530], as did students in a large introductory course employing active learning [531].

A recently published measure assesses shifts in students' physics course expectations in response to SCALE-UP orientation and instruction. The assessment is called the Pedagogical Expectation Violation Assessment (PEVA) [532]. At the beginning of the course, most students expected to attend lectures in an amphitheater classroom, to attend a separate laboratory section, to read the text, and to memorize equations, all with limited opportunities to interact with instructors and peers. As a result of a brief orientation to the course, most students shifted their expectations to be closer to the actual design of SCALE-UP (decreased lecture, an integrated laboratory environment, and more interactions including collaborative group work). The students also reduced their expectation for memorizing at two of three institutions studied.

Mason and Singh [510] refined an instrument to measure attitudes and beliefs about problem solving, called the AAPS survey. They compared responses on the survey across several groups: introductory physics and astronomy students (algebra based and calculus based), graduate students, and faculty. The graduate students answered the survey twice, once from the perspective of solving introductory physics problems, and again from the perspective of solving graduate-level physics problems. In general, introductory students had beliefs about problem solving that were less expertlike than the beliefs of graduate students or faculty. When graduate students took the survey from the perspective of solving graduate-level physics problems, their responses were less expertlike than when they answered the survey for introductory problems. For example, graduate students indicated that they have difficulty checking whether their answers to graduate problems are correct, and they feel they need to seek help from others when they get stuck.

- *Relationships between epistemological sophistication and performance.* Some studies have explored the relationship between epistemological beliefs and understanding of physics concepts as measured by performance on concept inventories or other course measures [533]. Halloun [503] reported a significant correlation between scores on the VASS, FCI gain, and course grades. For example, students who were classified as an "expert" profile on the VASS survey were most likely to be high achievers in the physics class.

Perkins *et al.* [500] observed a statistically significant correlation between scores on some (but not all) categories of the CLASS survey and pre- and post-test scores on the FMCE. The categories of *conceptual understanding* and *math physics connection* were significantly correlated with FMCE (coefficients in the range of 0.20 to 0.30), whereas real-world connection, personal interest, and sense making or effort were not significant (correlation coefficients ranging from 0.02 to 0.17). In addition, when students were grouped by normalized learning gains on the FMCE, students in a high gain bin (>0.9) tended to have more favorable beliefs whereas students with the lowest gain (<0.2) had unfavorable beliefs that declined after the course.

Kortemeyer [524] investigated statistical relationships between several different measures: epistemology as coded from students' online discussion behavior associated with Web-based homework and scores on the MPEX survey; and physics learning as measured by the FCI, final exam, and course grades. The study found significant correlations between FCI post-test scores, FCI gain, course grades, and the extent to which students' online discussion posts were physics oriented (positive) as compared to solution oriented (negative). These correlations were not observed for the MPEX, suggesting that online discussions might be a useful (and more authentic) diagnostic tool for assessing students' approaches to learning physics than survey instruments.

May and Etkina [221] required introductory physics students to submit weekly reports in which they reflected on how they learned particular physics topics in electricity and magnetism. The researchers compared the quality of students' reflections to their performance on several concept inventories: the FCI, MBT, and CSEM. They found that students with high pre-post normalized gains had a higher level of epistemological sophistication, such as asking insightful questions and trying to make sense of the material rather than saying they learned formulas.

Lising and Elby [534] conducted a case study on a single student "Jan" to investigate how epistemology affects learning. They observed that Jan had a separation between formal, classroom reasoning and everyday reasoning which contributed to her difficulties with learning physics. Although she could reason in both ways, she often did not attempt to connect them and/or reconcile inconsistencies. They suggest that research-based curricula can be made even more effective by making epistemology an explicit part of instruction, especially through reflective questions on assignments or during class discussions.

- *Instructional strategies to improve students' epistemological beliefs*. Elby [224] and Hammer and Elby [505] developed a set of instructional materials and strategies to address students' epistemological beliefs, which were found to result in significant, favorable shifts on the MPEX survey. Effective instructional practices included assigning essay questions in which students must argue for or against multiple perspectives, asking students to reflect on whether their answers agree with their "intuition" during laboratory activities, and submitting journal-like paragraphs in which students reflect about the strategies they use to learn physics (memorization, summarizing the text, solving problems, etc.). They acknowledge that implementing these approaches was at the expense of reduced content coverage, they only minimally used the textbook, and they attempted to keep lesson plans flexible to allow time to address students' difficulties.
- *Gender differences on attitude surveys*. One study of scores on CLASS indicate that females have less expertlike beliefs in statements related to categories of real-world connections, personal interest, problem-solving confidence, and problem-solving sophistication than their male peers [509]. However, their responses to questions in the sense-making or effort category are slightly more expertlike than those of male students.

2. Faculty beliefs and values about teaching and learning

Faculty beliefs and values about how students learn physics influence their decisions about what and how to teach, but due to external factors these beliefs may not be reflected in their actual teaching practices [511,512]. An important first step to understanding instructional decisions is to identify a common set of beliefs held by instructors and the situational factors that limit their use of reformed curricula or teaching methods. This information can then be taken into consideration by curriculum designers and leaders of professional development programs when disseminating research-based instructional tools.

- *Faculty conceptions and instructional practices*. Henderson and Dancy [512] report the results of semistructured interviews with five experienced, tenured physics faculty from four different institutions. They analyzed the faculties' conceptions about teaching and learning and self-reported instructional practices according to a framework outlined in Dancy and Henderson [523]. The ten categories used to rate the conceptions focus on views of student learning and how the instructor views their teaching role, which were scored on a scale from being traditional (transmissionist) in nature to more alternative (constructivist). Similarly, the ten categories used to rate self-described instructional practices considers whether an instructor describes their actions as being more consistent with alternative instruction (e.g., active, cooperative, creative, with process-based assessments) or traditional instruction (e.g., passive, individualistic, formulaic, with knowledge-based assessments). Henderson and Dancy [512] found that most of the faculty had views about teaching and learning that were rated as semialternative or a mix between alternative (reformed) views and traditional views. In contrast, their descriptions of their own teaching practices were rated as more traditional in nature. The study concluded that even though these faculty members were familiar with research-based instructional methods, agreed with reform approaches, and had access to materials, they often did not implement them because of external factors beyond their control. These barriers included a need to cover content (and limited time in which to do so), lack of time to prepare for teaching, inappropriate class size and room layout, departmental "norms" for how classes are taught, and student factors (student resistance or poor student attitudes).
- *Beliefs and values about teaching and learning of problem solving*. Problem solving is a primary component to most university physics courses, and a substantial portion of students' grades are often dependent upon their ability to solve problems as homework and on exams. Students are provided with worked-out examples in their textbooks and might observe additional demonstrations of problem solving during class; however, the problem features, example solutions, and problem-solving procedures used by

different instructors may vary widely. Yerushalmi *et al.* [511] and Henderson *et al.* [513] used structured interviews with six physics faculty to explore factors that contributed to their instructional decisions related to the teaching and learning of problem solving. The interviews took place with faculty at a research university and utilized specific "artifacts" as a basis for discussion, including different formats for the same problem statement, multiple student solutions, and multiple versions of an instructor solution [513]. During the interviews, the participants were prompted to comment on their format preferences for each type of artifact and indicate what they typically use in the classes they teach and why.

The researchers concluded that these physics instructors had unstable, often conflicting beliefs that were constructivist in nature, while their actions in the classroom reflected a traditional model of transmitting information [511]. For example, the faculty expressed a belief that students must be reflective learners and solve a lot of problems on their own to gradually build up an understanding of physics (an inquiry-based, constructivist view), but they often provided explicit guidance in course materials such as using problems that are broken into parts (parts a, b, c, etc.) to direct students through a problem-solving procedure. In addition, the faculty experienced a conflict between their "physicist" view that places value on compact, concise problem solutions and a "teacher" value of wanting students to communicate their reasoning. As a result, the instructors were unwilling to penalize a student who wrote a very sparse answer that could be interpreted as a correct result, but would penalize students who wrote incorrect reasoning that led to a correct numerical result [201]. The faculty who were interviewed acknowledged that their preferences do not match what they use in their courses, often because it takes too much time and expertise to construct high-quality problems and problem solutions, and they also do not want to overwhelm students with complicated or overly detailed problem solutions.

A later study extended this research to include interviews with 30 physics faculty at a wide range of institutions, and focused on instructors' choice of features when writing or selecting physics problems for their class [514]. During interviews, faculty indicated their preferences for or against particular problem features, including problems that were qualitative (no calculation), multiple choice, broken into parts, had a real-world context, were wordy (extra unnecessary words), included a drawing, or were complex (required multiple principles for a solution). The interview statements were coded along three dimensions: whether a particular problem feature supported or hindered an instructor's teaching goals, whether the feature was used by a faculty member in their courses in any context, and whether the feature was used on exams. The researchers concluded that for four of these features, the instructors' values were in conflict with their practices. Although faculty believed conceptual questions promote student understanding and a real-world context provides necessary motivation for students, they chose not to use problems with those features in class. In contrast, they stated that problems that are broken into parts and include a drawing hinder their goal that students will learn general problem-solving skills, but they frequently use these features to make a problem clear, especially on exams.

3. Instructor implementations of reformed curricula

This section reviews findings from studies related to faculty adoption of PER-based instructional strategies and materials, how PER-based strategies are actually implemented in the classroom, and student perceptions of classroom reforms.

- *Self-reported knowledge and use of reformed instruction*. Henderson and Dancy [368] conducted a large-scale online survey of physics faculty in the U.S. Of the 722 faculty who responded, most of them had heard of at least one research-based instructional strategy (87.1%) and about half were familiar with six or more strategies. In terms of use, approximately half of the faculty reported using reformed instruction in their teaching, but often with significant modifications. From the list of 24 research-based instructional strategies, the most common instructional strategy was *peer instruction*, with 63.5% of faculty reporting they know about the strategy and 29.2% of them saying they use it in some form. Other strategies that were familiar to more than 40% of faculty include Physlets, cooperative group problem solving, workshop physics, just-in-time teaching, tutorials, and interactive lecture demonstrations. Although this survey shows that there is high knowledge of PER-based instruction among faculty, situational factors such as limited time or lack of support deter faculty from using them [368,512]. During interviews with five physics faculty, Henderson and Dancy [516] also found that several faculty expressed dissatisfaction with their interactions with educational researchers. They suggest that PER should change its model for disseminating information and materials to better help faculty adapt the strategies to their specific situation.
- *Implementations of Peer Instruction*. Turpen and Finkelstein [517] used detailed classroom observations to identify variations in the ways that physics instructors implement *Peer Instruction* [213]. They found that the basic elements of *Peer Instruction* were present in all six classes studied: the presentation of conceptual information (usually in the form of questions) and opportunities for students to discuss physics with their classmates. The instructors differed, however, in the average number of conceptual questions asked per class period (range from 3 to 8), how they

interacted with students (both during the "voting" time and discussing the responses), and the time spent explaining the solution (range from 1 minute to nearly 4 minutes). One of the most prominent differences was in the average number of explanations heard from students during a class period, which was 4 or 5 statements for three of the instructors, around 2 or 3 for one instructor, and between 0 and 1 for the remaining two instructors. In a related study, Turpen and Finkelstein [518] administered surveys to students to explore the relationship between these instructor differences and student perceptions of their classroom experiences. They found that students who were in a class with more opportunities for discussion felt comfortable asking questions and speaking during class, with both their peers and the instructor. A higher fraction of these students also stated that it was important to understand the reasoning for an answer, not just to know the correct answer.

4. Teaching assistants

Graduate and undergraduate students are frequently employed to teach laboratory and/or recitation sections of introductory physics courses, especially at large institutions. The preparation and instructional support these TAs receive varies widely across institutions, and to date limited research has been done on teaching assistants' beliefs and practices.

- *Teaching assistants' pedagogical beliefs and their influence on instructor-student interactions*. Goertzen, Scherr, and Elby [519,520] and Spike and Finkelstein [535] used interviews and videotaped observations of teaching assistants to analyze their classroom interactions while using *Tutorials* curriculum at the University of Colorado–Boulder (CU) and the University of Maryland (UM). Goertzen *et al.* found that a lack of "buy-in" from the TA resulted in behaviors that conflicted with the tutorial developers' intentions, such as giving direct answers to students' questions and increasing the use of equations as a reasoning tool. Spike and Finkelstein concluded that the beliefs and behaviors of teaching assistants had a profound impact on the attitudes of students in the class, such as their willingness to engage in group work. Additional studies have investigated the pedagogical beliefs of graduate students and learning assistants about problem solving [536], studio style classrooms [537], interactive-engagement courses [538], or an inquiry-based course for future elementary school teachers [539]. These studies have generally concluded that there is a wide range in teaching assistant behaviors, which result in a variety of different experiences for students in those classes.
- *The impact of instructional styles on student learning*. Koenig, Endorf, and Braun [521] studied recitation classes taught by teaching assistants using one unit from the *Tutorials in Introductory Physics* curriculum and four different versions of implementation. The topic studied was energy and momentum, and the four instructional modes were lecture presentation of the materials, students working individually on the tutorial, students working in a group of 3 or 4 students, and students working in a group with instructor coaching (Socratic dialog). Student learning in each of these classes was assessed using pretest–post-test questions designed by the authors, which were written to assess concepts addressed during the tutorial unit. They found that students who experienced the fourth style of instruction scored significantly higher on the post-test than other groups (cooperative groups with instructor coaching). Surprisingly, the students in the third style (cooperative learning without instructor dialogue) scored at the same level as students in the lecture or individual course sections, suggesting that the TA interactions with students had a substantial influence on learning.

E. Strengths and limitations of research on attitudes and beliefs

1. Strengths

- There has been at least one large-scale survey research study to identify what fraction of faculty know about PER-based instructional strategies and how many claim to use them [368,369]. This gives a baseline indication of the success attained by dissemination efforts in PER and where future efforts should be focused. This, as well as interview studies of adoption of PER-informed instructional strategies, provides insights into what is needed to accelerate the adoption of effective curricula in teaching undergraduate physics.
- The prevalence of student attitude surveys (like the MPEX and CLASS) has given instructors an easy way to monitor students' attitudes before, during, and after physics instruction.

2. Limitations

- Qualitative studies involving video- and audiotaped interviews or observations are time consuming and difficult to analyze, so there is often a delay in the availability of results from these studies.
- Some attitude surveys, like the VASS and EBAPS, do not have published documentation about the instrument's development and score validation process.
- Research into faculty adoption of reformed curricula is in its infancy stage and a lot more work is needed to understand ways of helping faculty change their attitudes and adopt more effective instructional strategies, as well as ways of helping departments

implement reward structures for implementing evidence-based reforms effectively.

F. Areas for future study

- *Varied implementations of reformed curricula*. Although this section summarizes research studies that describe instructors' variations in their implementation of certain interactive-engagement strategies (such as *Peer Instruction*), there is a need for studies that explore how instructors modify and implement other research-based instructional strategies. There is also research needed on sociological and infrastructure factors that interfere with adoption of reformed curricula and instructional practices for faculty, undergraduate learning or teaching assistants, and graduate teaching assistants.
- *Teaching assistants*. There is a need for additional research on the attitudes and beliefs of TAs, how physics TAs impact the success of instructional reforms, and the implications for professional development of TAs. There is little guidance on how to prepare teaching assistants, in terms of both the components of a teaching assistant orientation program and ongoing professional development opportunities during teaching.
- *Institutional barriers*. There is a need for research on approaches that research universities can take to encourage and reward the implementation of research-based instructional strategies. Advancement in research universities for faculty (e.g., achieving tenure and promotion, receiving merit raises, being respected among peers) largely hinges on professors' research prowess, not teaching prowess. Better instruction would likely result when reward structures align with teaching accomplishments that reflect effective student learning, but it is not clear how to change institutional structures and traditions in ways that are acceptable to administrators, faculty, and institutional goals and values.

VIII. SUMMARY AND CONCLUSIONS

1. Summary of research in physics education from an historical perspective

Even with this nonexhaustive review of PER at the undergraduate level, it is evident that there has been a considerable corpus of research done to date. One characterization of PER is that it had its origin in some instructors observing stubborn conceptual difficulties experienced by their students in topics that seemed to the instructors to be simple physics ideas. This interest in digging deeper into students' conceptual learning difficulties led to hundreds of studies on common misconceptions, which have been documented for nearly every physics topic ranging from mechanics concepts like force, motion, momentum, and energy to topics in electricity and magnetism, thermal physics, light and optics, and modern physics. Research to document students' conceptual difficulties in the 1970s and 1980s led to the development of multiple-choice concept inventories that became widely available in the early 1990s, and to instructional interventions aimed at helping students overcome conceptual difficulties.

The availability of concept inventories served as a catalyst for discussions among physicists about the nature of learning and of conceptual difficulties among their students. Developers of early concept inventories, as well as PER researchers, invited colleagues to administer those inventories to their students following instruction; most refused, considering it to be a waste of time since they thought that surely their students would not hold these incorrect notions about fundamental and easy (to the professors) physics ideas—after all, they were good instructors who presented ideas clearly. Those few who initially took up the offer were surprised, perhaps even dismayed, by what they found. What later research showed was that quality of lecturing or instructor charisma had little to do with helping students learn concepts about which they held deeply rooted beliefs that contradicted physics laws.

Parallel with research into conceptual difficulties was interest in problem solving, given how central it is to physics. Initial research studies in the late 1970s and early 1980s focused on describing "expert-novice" differences in problem solving, by contrasting the processes used by beginning physics students to the principle-based approaches used by experienced solvers. These studies led to the development of instructional strategies and curricula to promote the use of expertlike approaches, which continue to be active topics of research today. It has also not gone unnoticed with physics instructors that teaching problem-solving skills to students in physics is a challenging endeavor. Indeed, nearly all physics instructors have experienced students coming to them and stating that they are "A students" in nearly all subjects but that they are doing poorly in physics and pleading for some prescriptive guidance for how to do well on exams.

Thus, research on problem solving combined with research on conceptual understanding has given rise to research-based and research-informed concept inventories, curricula, and instructional strategies. Research-based instructional strategies have become collectively referred to as "interactive engagement" methods in contrast to traditional, passive modes of instruction. For example, the use of classroom polling technologies or "clickers" and interactive demonstrations have become relatively widespread in introductory physics courses, and classroom environments structured to support increased interactions are growing in popularity (e.g., workshop, studio, and SCALE-UP classrooms). Several curricular "packages" were identified in Sec. IV, and some of these have been in existence for several decades. In recent years the design of curricula has expanded to include computer-based and Web-based instruction, including online homework systems, animations or simulations of phenomena, multimedia presentations, and computer tools for laboratory data collection and analysis.

As some of the more obvious things to try have already been done in PER, the decade from 2000 to 2010 has seen increased work in more interdisciplinary areas, such as applications of cognitive psychology to physics learning, argumentation in physics from a linguistic or psycholinguistic perspective, and student or faculty attitudes about teaching and learning, including opportunities and obstacles for instructors' adoption of research-based or informed instructional strategies.

Besides looking at the expansion of PER, another way to characterize the growth and interest in PER is to observe the historical development of the Physics Education Research Conference. Prior to 2001, PER conferences were a "cottage industry" with sporadic conferences held whenever a senior member of the PER community decided to organize one. There was one such conference organized by Beichner at NC State University in 1994 attracting about 43 attendees (24 PER faculty, with the remaining19 being graduate students and local faculty); the venue was not to discuss research, but issues such as the job market, what graduates of PER programs should know, how they should learn it, what the requirements of a PER graduate program should be, the need for publication venues, future conferences, research funding, and educating the physics community. The next PER conference was organized by Fuller at the University of Nebraska in 1998 attracting 83 participants. Starting in 2001, the Physics Education Research Conference (PERC) became a yearly event attached to the end of the American Association of Physics Teacher's summer meeting; a peer-reviewed conference proceeding also became an ancillary part of the conference. Typical attendance at the PERC during the last few years has been over 200.

2. What distinguishes PER from other DBER Fields?

For us, it is hard to characterize PER as one thing since it is diverse and evolving, drawing from disciplines such as cognitive science, sociology, education, linguistics, psycholinguistics, assessment, and measurement.

It should be noted that there are some elements of physics that distinguish it from other natural sciences, which may have implications for how PER differs from other discipline-based educational research. One is that the content of introductory physics (i.e., classical physics) has changed little in more than a century. This is certainly not the case in the biological sciences, astronomy, and geology, where the content of textbooks from 50 years ago has undergone major transformations compared to the content of today's textbooks; in contrast, other than the addition of color and other cosmetic changes, introductory physics textbooks have changed little. Physics instruction also places a strong emphasis on quantitative problem solving, which is not the case in other science disciplines (e.g., biology and introductory astronomy). Engineering places a major emphasis on design, while field experiences and a systems approach are very prominent in geology; not so in traditional physics instruction. These differences are likely to have important implications for the type and direction of STEM discipline-based education research.

In summary, we believe PER has a strong research basis in the first four areas addressed in this review: students' difficulties learning physics concepts and solving problems, the development and evaluation of instructional strategies and curricula (particularly "interactive engagement" methods), and the design and analysis of concept inventories. The final two areas of cognitive psychology and attitudes and beliefs are less developed but growing. The Supplemental Material [3] for this synthesis includes suggestions and speculations about future directions for PER that could hold promise, or that at least would further inform physics teaching and learning.

In conclusion, physics education research as a discipline has a rich history, makes a wide array of interdisciplinary connections, and yet has many promising avenues for future research.

ACKNOWLEDGMENTS

We thank the National Research Council, and in particular Natalie Nielsen and Susan Singer from the NRC's Committee on the Status, Contributions and Future Directions of Discipline-Based Education Research, for asking us to develop the original PER synthesis from which this publication emerged and for their very helpful discussions that helped shape the structure of this synthesis. We also thank Charles Henderson for encouraging us to update the original NRC white paper for publication in PRST-PER. Additionally, we thank the anonymous reviewers for their helpful suggestions. We also thank Paula Heron, Michael Wittmann, and Rachel Scherr for inviting us to present a summary of our white paper at the 2011 Frontiers and Foundations of Physics Education Research Conference, and Suzanne Brahmia for helping to summarize the suggestions made by attendees at this conference; we found the interactions and suggestions made by the attendees of this conference extremely helpful. Finally, we thank Jeanne Brunner, for her help in updating our review of the literature. This material is based on work supported by the National Science Foundation under Grant No. DUE-1347722.

[1] J. Docktor and J. Mestre, *A Synthesis of Discipline-based Education Research in Physics* (National Academy Press, Washington, DC, 2011), http://sites.nationalacademies.org/dbasse/bose/dbasse_080124.

[2] S. R. Singer, N. R. Nielsen, and H. A. Schweingruber, *Discipline-based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (National Academy Press, Washington DC, 2012).

[3] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevSTPER.10.020119 for the authors' subjective views of promising future areas for PER research.

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

[5] L. C. McDermott, Research on conceptual understanding in mechanics, Phys. Today **37**, 24 (1984).

[6] *Proceedings of the International Seminar on Misconceptions in Science and Mathematics*, edited by H. Helm and J. D. Novak (Department of Education, Cornell University, Ithaca, NY, 1983).

[7] *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Vols. I, II, and III*, edited by J. D. Novak (Department of Education, Cornell University, Ithaca, NY, 1987).

[8] S. Flores-Garcia, L. L. Alfaro-Avena, J. E. Chavez-Pierce, J. Luna-Gonzalez, and M. D. Gonzalez-Quezada, Students' difficulties with tension in massless strings, Am. J. Phys. **78**, 1412 (2010).

[9] R. A. Streveler, T. A. Litzinger, R. L. Miller, and P. S. Steif, Learning conceptual knowledge in the engineering sciences: Overview and future research directions, J. Eng. Educ. **97**, 279 (2008).

[10] L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics education research, Am. J. Phys. **67**, 755 (1999).

[11] B. S. Ambrose, P. R. L. Heron, S. Vokos, and L. C. McDermott, Student understanding of light as an electromagnetic wave: Relating the formalism to physical phenomena, Am. J. Phys. **67**, 891 (1999).

[12] L. C. McDermott and P. S. Shaffer, Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding, Am. J. Phys. **60**, 994 (1992).

[13] R. J. Beichner, Testing student interpretation of kinematics graphs, Am. J. Phys. **62**, 750 (1994).

[14] D. Hestenes and M. Wells, A mechanics baseline test, Phys. Teach. **30**, 159 (1992).

[15] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, Phys. Teach. **30**, 141 (1992).

[16] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).

[17] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula, Am. J. Phys. **66**, 338 (1998).

[18] M. T. H. Chi and J. D. Slotta, The ontological coherence of intuitive physics, Cognit. Instr. **10**, 249 (1993).

[19] M. T. H. Chi, J. D. Slotta, and N. de Leeuw, From things to processes: A theory of conceptual change for learning science concepts, Learn. Instr. **4**, 27 (1994).

[20] A. A. diSessa, N. Gillespie, and J. Esterly, Coherence vs. fragmentation in the development of the concept of force, Cogn. Sci. **28**, 843 (2004).

[21] A. Gupta, D. Hammer, and E. F. Redish, The case for dynamic models of learners' ontologies in physics, J. Learn. Sci. **19**, 285 (2010).

[22] C. Ioannides and S. Vosniadou, The changing meanings of force, *Cognit. Sci. Q.* **2**, 5 (2002).

[23] J. D. Slotta, M. T. H. Chi, and E. Joram, Assessing students' misclassifications of physics concepts: An ontological basis for conceptual change, Cognit. Instr. **13**, 373 (1995).

[24] G. Chiou and O. R. Anderson, A study of undergraduate physics students' understanding of heat conduction based on mental model theory and an ontology-process analysis, Sci. Educ. **94**, 825 (2010).

[25] K. N. Dunbar, J. A. Fugeslsang, and C. Stein, in *Thinking with Data*, edited by M. C. Lovett and P. Shah (Erlbaum, Mahwah, NJ, 2007), pp. 197–205.

[26] C. Camp and J. Clement, *Preconceptions in Mechanics: Lessons Dealing with Students' Conceptual Difficulties* (Kendall-Hunt, Dubuque, IA, 1994).

[27] L. C. McDermott, P. S. Shaffer, and The Physics Education Group at the University of Washington, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).

[28] J. Slotta and M. T. H. Chi, Helping students understand challenging topics in science through ontology training, Cognit. Instr. **24**, 261 (2006).

[29] D. R. Sokoloff and R. K. Thornton, in *Proceedings of ICUPE*, edited by E. F. Redish and J. S. Rigden (American Institute of Physics, Melville, NY, 1997), pp. 1061–1074.

[30] M. S. Steinberg and C. L. Wainwright, Using models to teach electricity—The CASTLE project, Phys. Teach. **31**, 353 (1993).

[31] K. Wosilait, P. R. L. Heron, P. S. Shaffer, and L. C. McDermott, Development and assessment of a research-based tutorial on light and shadow, Am. J. Phys. **66**, 906 (1998).

[32] E. D. Corpuz and N. S. Rebello, Investigating students' mental models and knowledge construction of microscopic friction. II. Implications for curriculum design and development, Phys. Rev. ST Phys. Educ. Res. **7**, 020103 (2011).

[33] G. Posner, K. Strike, P. Hewson, and W. Gerzog, Accommodation of a scientific conception: Toward a theory of conceptual change, Sci. Educ. **66**, 211 (1982).

[34] E. Etkina, J. Mestre, and A. O'Donnell, in *The Cognitive Revolution in Educational Psychology*, edited by J. M. Royer (Information Age Publishing, Greenwich, CT, 2005), pp. 119–164.

[35] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience, and School* (National Academy Press, Washington, DC, 1999).

[36] S. Carey, in *Conceptual Development: Piaget's Legacy*, edited by E. K. Scholnick, K. Nelson, and P. Miller (Lawrence Erlbaum Assoc., Mahwah, NJ, 1999), pp. 293–326.

[37] G. M. Sinatra and P. R. Pintrich, in *Intentional Conceptual Change*, edited by G. M. Sinatra and P. R. Pintrich (Lawrence Erlbaum Assoc., Mahwah, NJ, 2003), pp. 1–18.

[38] K. A. Strike and G. J. Posner, Conceptual change and science teaching, Eur. J. Sci. Educ. **4**, 231 (1982).

[39] K. A. Strike and G. J. Posner, in *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice*, edited by R. A. Duschl and R. J. Hamilton (State University of New York Press, New York, 1992), pp. 147–176.

[40] A. A. diSessa, Toward an epistemology of physics, Cognit. Instr. **10**, 105 (1993).

[41] A. A. diSessa and B. L. Sherin, What changes in conceptual change? Int. J. Sci. Educ. **20**, 1155 (1998).

[42] D. Hammer, Misconceptions or p-prims: How might alternative perspectives on cognitive structure influence instructional perceptions and intentions, J. Learn. Sci. **5**, 97 (1996).

[43] Harvard-Smithsonian Center for Astrophysics, A Private Universe, videotape, http://www.learner.org/resources/series28.html, retrieved Sept. 15, 2010.

[44] D. Hammer, Student resources for learning introductory physics, Am. J. Phys. **68**, S52 (2000).

[45] J. Minstrell, in *Proceedings of the International Workshop on Research in Physics Education: Theoretical Issues and Empirical Studies*, edited by R. Duit, F. Goldberg, and H. Niedderer (Institut fur die Padagogik der Naturwissenschaften, Kiel, Germany, 1992), pp. 110–128.

[46] J. Smith, A. diSessa, and J. Roschelle, Misconceptions reconceived: A constructivist analysis of knowledge in transition, J. Learn. Sci. **3**, 115 (1994).

[47] D. Hammer, More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research, Am. J. Phys. **64**, 1316 (1996).

[48] D. Hammer and A. Elby, in *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, edited by B. K. Hofer and P. R. Pintrich (Erlbaum, Mahwah, NJ, 2002), pp. 169–190.

[49] R. J. Dufresne, W. J. Leonard, and W. J. Gerace, Making sense of students' answers to multiple choice questions, Phys. Teach. **40**, 174 (2002).

[50] R. N. Steinberg and M. S. Sabella, Performance on multiple-choice diagnostics and complementary exam problems, Phys. Teach. **35**, 150 (1997).

[51] D. E. Brown and D. Hammer, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, New York, 2008), pp. 127–154.

[52] M. T. H. Chi, in *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science*, edited by R. Giere (University of Minnesota Press, Minneapolis, MN, 1992), pp. 129–186.

[53] M. T. H. Chi, in *Conceptual Structures, and Processes: Emergence, Discovery, and Change*, edited by T. B. War, S. M. Smith, and J. Vaid (American Psychological Association, Washington, DC, 1997), pp. 209–234.

[54] F. C. Keil, *Semantic and Conceptual Development: An Ontological Perspective* (Harvard University Press, Cambridge, MA, 1979).

[55] G. Zhu and C. Singh, Improving students' understanding of quantum measurement. I. Investigation of difficulties, Phys. Rev. ST Phys. Educ. Res. **8**, 010117 (2012).

[56] G. Zhu and C. Singh, Improving students' understanding of quantum measurement. II. Development of research-based learning tools, Phys. Rev. ST Phys. Educ. Res. **8**, 010118 (2012).

[57] A. Strauss and J. Corbin, *Basics of Qualitative Research* (Sage, Thousand Oaks, CA, 1998).

[58] D. Brown and J. Clement, Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction, Instr. Sci. **18**, 237 (1989).

[59] J. Clement, Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics, J. Res. Sci. Teach. **30**, 1241 (1993).

[60] Harvard-Smithsonian Center for Astrophysics, Minds of Our Own: Lessons from Thin Air, videotape, http://www.learner.org/resources/series26.html, retrieved Sept. 15, 2010.

[61] M. T. H. Chi, Common sense conceptions of emergent processes: Why some misconceptions are robust, J. Learn. Sci. **14**, 161 (2005).

[62] M. Chiu, C. Chou, and C. Liu, Dynamic processes of conceptual changes: Analysis of constructing mental models of chemical equilibrium, J. Res. Sci. Teach. **39**, 688 (2002).

[63] D. B. Clark, Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring, Cognit. Instr. **24**, 467 (2006).

[64] M. Ferrari and M. T. H. Chi, The nature of naïve explanations of natural selection, Int. J. Sci. Educ. **20**, 1231 (1998).

[65] C. Furio and J. Guisasola, Difficulties in learning the concept of electric field, Sci. Educ. **82**, 511 (1998).

[66] E. Kikas, Teachers' conceptions and misconceptions concerning three natural phenomena, J. Res. Sci. Teach. **41**, 432 (2004).

[67] Y. Lee and N. Law, Explorations in promoting conceptual change in electrical concepts via ontological category shift, Int. J. Sci. Educ. **23**, 111 (2001).

[68] L. M. Tyson, G. J. Venville, A. G. Harrison, and D. F. Treagust, A multidimensional framework for interpreting conceptual change events in the classroom, Sci. Educ. **81**, 387 (1997).

[69] G. J. Venville and D. F. Treagust, Exploring conceptual change in genetics using a multidimensional interpretive framework, J. Res. Sci. Teach. **35**, 1031 (1998).

[70] L. Deslauriers and C. Wieman, Learning and retention of quantum concepts with different teaching methods, Phys. Rev. ST Phys. Educ. Res. **7**, 010101 (2011).

[71] F. Reif and L. Scott, Teaching scientific thinking skills: Students and computers coaching each other, Am. J. Phys. **67**, 819 (1999).

[72] F. Reif and M. St. John, Teaching physicists' thinking skills in the laboratory, Am. J. Phys. **47**, 950 (1979).
[73] A. Van Heuvelen, Learning to think like a physicist: A review of research-based strategies, Am. J. Phys. **59**, 891 (1991).
[74] D. P. Maloney, in *Handbook of Research on Science Teaching and Learning*, edited by D. L. Gabel (MacMillan, New York, 1994), pp. 327–354.
[75] L. Hsu, E. Brewe, T. M. Foster, and K. A. Harper, Resource letter RPS-1: Research in problem solving, Am. J. Phys. **72**, 1147 (2004).
[76] K. Cummings, *A Developmental History of Physics Education Research* (National Academy Press, Washington, DC, 2011), http://sites.nationalacademies.org/dbasse/bose/dbasse_080124.
[77] E. Bagno and B. Eylon, From problem solving to knowledge structure: An example from the domain of electromagnetism, Am. J. Phys. **65**, 726 (1997).
[78] M. T. H. Chi, R. Glaser, and E. Rees, in *Advances in the Psychology of Human Intelligence*, edited by R. J. Sternberg (Lawrence Erlbaum Associates, Hillsdale, NJ, 1982), Vol. 1, pp. 7–75.
[79] B. Eylon and F. Reif, Effects of knowledge organization on task performance, Cognit. Instr. **1**, 5 (1984).
[80] J. Larkin, Processing information for effective problem solving, *Eng. Educ.* **70**, 285 (1979).
[81] J. Larkin, Cognition of learning physics, Am. J. Phys. **49**, 534 (1981).
[82] J. Larkin, in *Cognitive Skills and Their Acquisition*, edited by J. R. Anderson (Lawrence Erlbaum Associates, Hillsdale, NJ, 1981), pp. 311–344.
[83] J. H. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, Science **208**, 1335 (1980).
[84] M. T. H. Chi, P. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, Cogn. Sci. **5**, 121 (1981).
[85] P. T. Hardiman, R. Dufresne, and J. P. Mestre, The relationship between problem categorization and problem solving among experts and novices, Mem. Cogn. **17**, 627 (1989).
[86] E. R. Savelsbergh, T. de Jong, and M. G. M. Ferguson-Hessler, Choosing the right solution approach: The crucial role of situational knowledge in electricity and magnetism, Phys. Rev. ST Phys. Educ. Res. **7**, 010103 (2011).
[87] M. T. H. Chi, M. Bassok, M. W. Lewis, P. Reimann, and R. Glaser, Self-explanations: How students study and use examples in learning to solve problems, Cogn. Sci. **13**, 145 (1989).
[88] M. G. M. Ferguson-Hessler, and T. de Jong, Studying physics texts: Differences in study processes between good and poor performers, Cognit. Instr. **7**, 41 (1990).
[89] E. Cohen, A. Mason, C. Singh, and E. Yerushalmi, Identifying differences in diagnostic skills between physics students: Students' self-diagnostic performance given alternative scaffolding, AIP Conf. Proc. **1064**, 99 (2008).
[90] E. Yerushalmi, A. Mason, E. Cohen, and C. Singh, Effects of self-diagnosis on subsequent problem solving performance, AIP Conf. Proc. **1064**, 53 (2008).
[91] M. Ward and J. Sweller, Structuring effective worked examples, Cognit. Instr. **7**, 1 (1990).
[92] J. I. Heller and F. Reif, Prescribing effective human problem-solving processes: Problem description in physics, Cognit. Instr. **1**, 177 (1984).
[93] J. H. Larkin, in *Mental Models*, edited by D. Gentner and A. L. Stevens (Lawrence Erlbaum Assoc., Hillsdale, NJ, 1983), pp. 75–98.
[94] B. Ibrahim and N. S. Rebello, Representational task formats and problem solving strategies in kinematics and work, Phys. Rev. ST Phys. Educ. Res. **8**, 010126 (2012).
[95] D. Rosengrant, A. Van Heuvelen, and E. Etkina, Free-body diagrams: Necessary or sufficient? AIP Conf. Proc. **790**, 177 (2005).
[96] A. Van Heuvelen and X. Zou, Multiple representations of work-energy processes, Am. J. Phys. **69**, 184 (2001).
[97] M. De Cock, Representation use and strategy choice in physics problem solving, Phys. Rev. ST Phys. Educ. Res. **8**, 020117 (2012).
[98] P. B. Kohl and N. D. Finkelstein, Student representational competence and self-assessment when solving problems, Phys. Rev. ST Phys. Educ. Res. **1**, 010104 (2005).
[99] D. E. Meltzer, Relation between students' problem-solving performance and representational format, Am. J. Phys. **73**, 463 (2005).
[100] T. J. Bing and E. F. Redish, Epistemic complexity and the journeyman-expert transition, Phys. Rev. ST Phys. Educ. Res. **8**, 010105 (2012).
[101] M. Bodin, Mapping university students' epistemic framing of computational physics using network analysis, Phys. Rev. ST Phys. Educ. Res. **8**, 010115 (2012).
[102] E. T. Torigoe and G. E. Gladding, Connecting symbolic difficulties with failure in physics, Am. J. Phys. **79**, 133 (2011).
[103] B. L. Sherin, The symbolic basis of physical intuition: A study of two symbol systems in physics instruction, Ph.D. thesis, University of California, Berkeley, 1996 (unpublished).
[104] B. L. Sherin, How students understand physics equations, Cognit. Instr. **19**, 479 (2001).
[105] E. Torigoe, What kind of math matters? A study of the relationship between mathematical ability and success in physics, Ph.D. thesis, University of Illinois at Urbana-Champaign, 2008 (unpublished).
[106] J. Tuminaro and E. F. Redish, Elements of a cognitive model of physics problem solving: Epistemic games, Phys. Rev. ST Phys. Educ. Res. **3**, 020101 (2007).
[107] N. L. Nguyen and D. E. Meltzer, Initial understanding of vector concepts among students in introductory physics courses, Am. J. Phys. **71**, 630 (2003).
[108] L. Cui, N. S. Rebello, and A. G. Bennett, College students' transfer from calculus to physics, AIP Conf. Proc. **818**, 37 (2006).
[109] E. Cohen and S. E. Kanim, Factors influencing the algebra "reversal error", Am. J. Phys. **73**, 1072 (2005).
[110] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, Am. J. Phys. **60**, 637 (1992).

[111] J. Mestre, Probing adults' conceptual understanding and transfer of learning via problem posing, J. Appl. Dev. Psychol. **23**, 9 (2002).
[112] T. L. O'Kuma, S. P. Maloney, and C. J. Hieggelke, *Ranking Task Exercises in Physics* (Prentice-Hall, Upper Saddle River, NJ, 2000).
[113] A. Van Heuvelen, Experiment problems for mechanics, Phys. Teach. **33**, 176 (1995).
[114] A. Van Heuvelen and D. Maloney, Playing physics jeopardy, Am. J. Phys. **67**, 252 (1999).
[115] D. S. Wright and C. D. Williams, A WISE strategy for introductory physics, Phys. Teach. **24**, 211 (1986).
[116] C. Becerra-Labra, A. Gras-Martí, and J. Martínez Torregrosa, The effects of a problem-based structure of physics contents on conceptual learning and the ability to solve problems, Int. J. Sci. Educ. **34**, 1235 (2012).
[117] W. J. Leonard, R. J. Dufresne, and J. P. Mestre, Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems, Am. J. Phys. **64**, 1495 (1996).
[118] A. Van Heuvelen, Overview, case study physics, Am. J. Phys. **59**, 898 (1991).
[119] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, Am. J. Phys. **60**, 627 (1992).
[120] G. Kortemeyer, E. Kashy, W. Benenson, and W. Bauer, Experiences using the open-source learning content management and assessment system LON-CAPA in introductory physics courses, Am. J. Phys. **76**, 438 (2008).
[121] Y.-J. Lee, D. J. Palazzo, R. Warnakulassoriya, and D. E. Pritchard, Measuring student learning with item response theory, Phys. Rev. ST Phys. Educ. Res. **4**, 010102 (2008).
[122] A. Newell, J. D. Shaw, and H. A. Simon, Elements of a theory of human problem solving, Psychol. Rev. **65**, 151 (1958).
[123] A. Newell and H. A. Simon, *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, NJ, 1972).
[124] H. A. Simon, in *Handbook of Learning and Cognitive Processes*, edited by W. K. Estes (Erlbaum, Hillsdale, NJ, 1978), Vol. 5, pp. 271–295.
[125] M. L. Gick, Problem-solving strategies, Educ. Psychol. **21**, 99 (1986).
[126] J. Sweller, Cognitive load during problem solving: Effects on learning, Cogn. Sci. **12**, 257 (1988).
[127] D. Gentner, Structure-mapping: A theoretical framework for analogy, Cogn. Sci. **7**, 155 (1983).
[128] D. Gentner, in *Similarity and Analogical Reasoning*, edited by S. Vosniadou and A. Ortony (Cambridge University Press, Cambridge, England, 1989), pp. 199–241.
[129] K. Holyoak, in *The Psychology of Learning and Motivation*, edited by G. H. Bower (Academic Press, San Diego, CA, 1985), Vol. 19.
[130] K. J. Holyoak and K. Koh, Surface and structural similarity in analogical transfer, Mem. Cogn. **15**, 332 (1987).
[131] B. H. Ross, This is like that: The use of earlier problems and the separation of similarity effects, J. Exp. Psychol. Learn. Mem. Cogn. **13**, 629 (1987).
[132] B. H. Ross, Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems, J. Exp. Psychol. Learn. Mem. Cogn. **15**, 456 (1989).
[133] L. M. Reeves and R. W. Weisberg, The role of content and abstract information in analogical transfer, Psychol. Bull. **115**, 381 (1994).
[134] S. Y. Lin and C. Singh, Using isomorphic problems to learn introductory physics, Phys. Rev. ST Phys. Educ. Res. **7**, 020104 (2011).
[135] D. Hammer, A. Elby, R. Scherr, and E. F. Redish, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. Mestre (Information Age Publishing, Greenwich, CT, 2005), pp. 89–119.
[136] J. Tuminaro, A cognitive framework for analyzing and describing introductory students use and understanding of mathematics in physics, Ph.D. thesis, University of Maryland, 2004 (unpublished).
[137] J. S. Brown, A. Collins, and P. Duguid, Situated cognition and the culture of learning, Educ. Res. **18**, 32 (1989).
[138] A. Mason and C. Singh, Revisiting categorization, in *Proceedings of the 2009 National Association for Research in Science Teaching (NARST) Annual Meeting* (Garden Grove, CA, 2009) Symposium S2.2, pp. 3–21.
[139] F. Mateycik, D. Jonassen, and N. S. Rebello, Using Similarity Rating Tasks to Assess Case Reuse in *Problem Solving, Proceedings of the Physics Education Research Conference* (Ann Arbor, MI, 2010).[*AIP Conference Proceedings* **1179**, 201, (2009)].
[140] S. F. Wolf, D. P. Dougherty, and G. Kortemeyer, Empirical approach to interpreting card-sorting data, Phys. Rev. ST Phys. Educ. Res. **8**, 010124 (2012).
[141] A. Carmichael, A. Larson, E. Gire, L. Loschky, and N. S. Rebello, How does visual attention differ between experts and novices on physics problems?, AIP Conf. Proc. **1289**, 93 (2010).
[142] D. Rosengrant, A. Van Heuvelen, and E. Etkina, Do students use and understand free-body diagrams?, Phys. Rev. ST Phys. Educ. Res. **5**, 010108 (2009).
[143] A. D. Smith, J. P. Mestre, and B. H. Ross, Eye gaze patterns as students study worked out examples in mechanics, Phys. Rev. ST Phys. Educ. Res. **6**, 020118 (2010).
[144] T. de Jong and M. G. M. Ferguson-Hessler, Cognitive structures of good and poor novice problem solvers in physics, J. Educ. Psychol. **78**, 279 (1986).
[145] J. L. Docktor, J. P. Mestre, and B. H. Ross, Impact of a short intervention on novices' categorization criteria, Phys. Rev. ST Phys. Educ. Res. **8**, 020102 (2012).
[146] S. F. Wolf, D. P. Dougherty, and G. Kortemeyer, Rigging the deck: Selecting good problems for expert-novice card-sorting experiments, Phys. Rev. ST Phys. Educ. Res. **8**, 020116 (2012).
[147] M. Finegold and R. Mass, Differences in the process of solving physics problems between good problem solvers and poor problem solvers, Res. Sci. Technol. Educ. **3**, 59 (1985).
[148] L. N. Walsh, R. G. Howard, and B. Bowe, Phenomenographic study of students' problem solving approaches in

physics, Phys. Rev. ST Phys. Educ. Res. **3**, 020108 (2007).

[149] A. Pawl, A Barrantes, and D. E. Pritchard, What do seniors remember from freshman physics?, Phys. Rev. ST Phys. Educ. Res. **8**, 020118 (2012).

[150] M. T. H. Chi, in *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. A. Ericsson, N. Charness, R. R. Hoffman, and P. J. Feltovich (Cambridge University Press, New York, 2006), pp. 21–30.

[151] F. Reif and J. I. Heller, Knowledge structure and problem solving in physics, Educ. Psychol. **17**, 102 (1982).

[152] C. Singh, When physical intuition fails, Am. J. Phys. **70**, 1103 (2002).

[153] M. G. M. Ferguson-Hessler and T. de Jong, On the quality of knowledge in the field of electricity and magnetism, Am. J. Phys. **55**, 492 (1987).

[154] C. Henderson and K. A. Harper, Quiz corrections: Improving learning by encouraging students to reflect on their mistakes, Phys. Teach. **47**, 581 (2009).

[155] D. Rosengrant and T. Mzoughi, Preliminary study of impulse-momentum diagrams, AIP Conf. Proc. **1064**, 187 (2008).

[156] C. J. De Leone and E. Gire, Is instructional emphasis on the use of nonmathematical representations worth the effort?, AIP Conf. Proc. **818**, 45 (2006).

[157] R. Ploetzner, E. Fehse, C. Kneser, and H. Spada, Learning to relate qualitative and quantitative problem representations in a model-based setting for collaborative problem solving, J. Learn. Sci. **8**, 177 (1999).

[158] D. Rosengrant, A. Van Heuvelen, and E. Etkina, Case study: Students' use of multiple representations in problem solving, AIP Conf. Proc. **818**, 49 (2006).

[159] P. B. Kohl and N. D. Finkelstein, Patterns of multiple representation use by experts and novices during physics problem solving, Phys. Rev. ST Phys. Educ. Res. **4**, 010111 (2008).

[160] J. M. Aguirre and G. Erickson, Students' conceptions about the vector characteristics of three physics concepts, J. Res. Sci. Teach. **21**, 439 (1984).

[161] S. Flores, S. E. Kanim, and C. H. Kautz, Student use of vectors in introductory mechanics, Am. J. Phys. **72**, 460 (2004).

[162] P. S. Shaffer and L. C. McDermott, A research-based approach to improving student understanding of the vector nature of kinematical concepts, Am. J. Phys. **73**, 921 (2005).

[163] K. Christianson, J. P. Mestre, and S. G. Luke, Practice makes (nearly) perfect: Solving 'students-and-professors'-type algebra word problems, Appl. Cogn. Psychol. **26**, 810 (2012).

[164] M. Bassok and K. J. Holyoak, Interdomain transfer between isomorphic topics in algebra and physics, J. Exp. Psychol. Learn. Mem. Cogn. **15**, 153 (1989).

[165] D. C. Meredith and K. A. Marrongelle, How students use mathematical resources in an electrostatics context, Am. J. Phys. **76**, 570 (2008).

[166] M. S. Sabella and E. F. Redish, Knowledge organization and activation in physics problem solving, Am. J. Phys. **75**, 1017 (2007).

[167] T. J. Bing and E. F. Redish, Analyzing problem solving using math in physics: Epistemological framing via warrants, Phys. Rev. ST Phys. Educ. Res. **5**, 020108 (2009).

[168] E. C. Sayre and M. C. Wittmann, Plasticity of intermediate mechanics students' coordinate system choice, Phys. Rev. ST Phys. Educ. Res. **4**, 020105 (2008).

[169] C. A. Ogilvie, Changes in students' problem-solving strategies in a course that includes context-rich, multifaceted problems, Phys. Rev. ST Phys. Educ. Res. **5**, 020102 (2009).

[170] A. Van Heuvelen, L. D. Allen, and P. Mihas, Experiment problems for electricity and magnetism, Phys. Teach. **37**, 482 (1999).

[171] D. P. Maloney, Ranking tasks: A new type of test item, *J. Coll. Sci. Teach.* **16**, 510 (1987).

[172] See http://relate.mit.edu/teaching/real-world-problems/.

[173] See http://www.physics.umd.edu/perg/abp/think/.

[174] L. Ding, N. Reay, A. Lee, and L. Bao, Exploring the role of conceptual scaffolding in solving synthesis problems, Phys. Rev. ST Phys. Educ. Res. **7**, 020109 (2011).

[175] D. Huffman, Effect of explicit problem solving instruction on high school students' problem-solving performance and conceptual understanding of physics, J. Res. Sci. Teach. **34**, 551 (1997).

[176] J. H. P. Van Weeren, F. F. M. de Mul, M. J. Peters, H. Kramers-Pals, and H. J. Roossink, Teaching problem-solving in physics: A course in electromagnetism, Am. J. Phys. **50**, 725 (1982).

[177] G. Pólya, *How to Solve It* (Princeton University Press, Princeton, NJ, 1945).

[178] K. Heller and P. J. Heller, *Competent Problem Solver—Calculus Version* (McGraw-Hill, New York, 2000).

[179] F. Reif, *Understanding Basic Mechanics, Text and Workbook* (Wiley, New York, 1995).

[180] R. Gautreau and L. Novemsky, Concepts first—A small group approach to physics learning, Am. J. Phys. **65**, 418 (1997).

[181] R. J. Dufresne, W. J. Gerace, P. T. Hardiman, and J. P. Mestre, Constraining novices to perform expertlike problem analyses: Effects on schema acquisition, J. Learn. Sci. **2**, 307 (1992).

[182] J. P. Mestre, R. J. Dufresne, W. J. Gerace, P. T. Hardiman, and J. S. Touger, Promoting skilled problem-solving behavior among beginning physics students, J. Res. Sci. Teach. **30**, 303 (1993).

[183] A. Van Heuvelen, *ALPS: Mechanics* (Hayden-McNeil Publishing, Plymouth, MI 1994), Vol. 1.

[184] A. Van Heuvuelen, *ALPS: Electricity and Magnetism* (Hayden-McNeil Publishing, Plymouth, MI, 1994), Vol. 2.

[185] R. Taconis, M. G. M. Ferguson-Hessler, and H. Broekkamp, Teaching science problem-solving: An overview of experimental work, J. Res. Sci. Teach. **38**, 442 (2001).

[186] D. Kane and B. Sherwood, A computer-based course in classical mechanics, Comput. Educ. **4**, 15 (1980).

[187] B. A. Sherwood, Free-body diagrams (a PLATO lesson), Am. J. Phys. **39**, 1199 (1971).

[188] S. G. Smith and B. A. Sherwood, Educational uses of the PLATO computer system, Science **192**, 344 (1976).
[189] A. S. Gertner and K. VanLehn, in *Proceedings of the 5th International Conference, ITS 2000*, edited by G. Gautheier, C. Frasson, and K. VanLehn (Springer, New York, 2000), pp. 133–142.
[190] Q. Ryan, E. Frodermann, K. Heller, L. Hsu, and B. Aryal, Designing a useful problem solving coach: Usage and usability studies, in *Proceedings of the Physics Education Research Conference 2013, Portland, OR, 2014* edited by P. V. Engelhardt, A. D. Churukian, D. L. Jones pp. 309–312.
[191] E. Kashy, B. M. Sherrill, Y. Tsai, D. Thaler, D. Weinshank, M. Engelmann, and D. J. Morrissey, CAPA, an integrated computer assisted personalized assignment system, Am. J. Phys. **61**, 1124 (1993).
[192] E. Kashy, S. J. Gaff, N. H. Pawley, W. L. Stretch, S. L. Wolfe, D. J. Morrissey, and Y. Tsai, Conceptual questions in computer-assisted assignments, Am. J. Phys. **63**, 1000 (1995).
[193] A. M. Pascarella, CAPA (Computer-Assisted Personalized Assignments) in a Large University Setting, Ph.D. thesis, University of Colorado, Boulder, 2002 (unpublished).
[194] S. W. Bonham, D. L. Deardorff, and R. J. Beichner, Comparison of student performance using web and paper-based homework in college level physics, J. Res. Sci. Teach. **40**, 1050 (2003).
[195] K. K. Cheng, B. A. Thacker, R. L. Cardenas, and C. Crouch, Using an online homework system enhances students' learning of physics concepts in an introductory physics course, Am. J. Phys. **72**, 1447 (2004).
[196] J. Mestre, D. M. Hart, K. A. Rath, and R. Dufresne, The effect of web-based homework on test performance in large enrollment introductory physics courses, *J. Comput. Math. Sci. Teach.* **21**, 229 (2002).
[197] E.-S. Morote and D. E. Pritchard, What course elements correlate with improvement on tests in introductory Newtonian mechanics?, Am. J. Phys. **77**, 746 (2009).
[198] K. VanLehn, C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill, The Andes physics tutoring system: Lessons learned, *Int. J. Artif. Intell. Educ.* **15**, 147 (2005).
[199] E. Gire, Between the poles: Locating physics majors in the expert-novice continuum Ph.D. thesis, University of California, San Diego, 2007 (unpublished).
[200] K. A. Harper, Investigating the development of problem solving skills during a freshman physics sequence Ph.D. thesis, The Ohio State University, 2001 (unpublished).
[201] C. Henderson, E. Yerushalmi, V. Kuo, P. Heller, and K. Heller, Grading student problem solutions: The challenge of sending a consistent message, Am. J. Phys. **72**, 164 (2004).
[202] W. K. Adams and C. E. Wieman, Problem solving skill evaluation instrument—Validation studies, AIP Conf. Proc. **883**, 18 (2006).
[203] J. D. Marx and K. Cummings, in Development of a survey instrument to gauge students' problem-solving abilities, AIP Conf. Proc. **1289** 221 (2010).
[204] J. L. Docktor, Development and validation of a physics problem-solving assessment rubric, Ph.D. thesis, University of Minnesota, 2009 (unpublished).
[205] E. F. Redish, *Teaching Physics with the Physics Suite* (John Wiley & Sons, Hoboken, NJ, 2003).
[206] A. B. Arons, *A Guide to Introductory Physics Teaching* (John Wiley & Sons, New York, 1990).
[207] R. D. Knight, *Five Easy Lessons: Strategies for Successful Physics Teaching* (Addison-Wesley, Glenview, IL, 2002).
[208] D. E. Meltzer and R. K. Thornton, Resource letter ALIP-1: Active-learning instruction in physics, Am. J. Phys. **80**, 478 (2012).
[209] 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).
[210] D. E. Meltzer and K. Manivannan, Transforming the lecture-hall environment: The fully interactive physics lecture, Am. J. Phys. **70**, 639 (2002).
[211] L. Ding, N. W. Reay, N. Lee, and L. Bao, Are we asking the right questions? Validating clicker question sequences by student interviews, Am. J. Phys. **77**, 643 (2009).
[212] R. J. Dufresne, W. J. Gerace, W. J. Leonard, J. P. Mestre, and L. Wenk, Classtalk: A classroom communication system for active learning, J. Comput. High Educ. **7**, 3 (1996).
[213] E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
[214] D. R. Sokoloff and R. K. Thornton, *Interactive Lecture Demonstrations, Active Learning in Introductory Physics* (John Wiley & Sons, Hoboken, NJ, 2004).
[215] G. M. Novak, E. T. Patterson, A. D. Gavrin, and W. Christian, *Just in Time Teaching* (Prentice-Hall, Upper Saddle River, NJ, 1999).
[216] Z. Chen, T. Stelzer, and G. Gladding, Using multimedia modules to better prepare students for introductory physics lecture, Phys. Rev. ST Phys. Educ. Res. **6**, 010108 (2010).
[217] T. Stelzer, D. T. Brookes, G. Gladding, and J. P. Mestre, Impact of multimedia learning modules on an introductory course on electricity and magnetism, Am. J. Phys. **78**, 755 (2010).
[218] T. Stelzer, G. Gladding, J. P. Mestre, and D. T. Brookes, Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content, Am. J. Phys. **77**, 184 (2009).
[219] E. Etkina, Weekly reports: A two-way feedback tool, Sci. Educ. **84**, 594 (2000).
[220] K. A. Harper, E. Etkina, and Y. Lin, Encouraging and analyzing student questions in a large physics course: Meaningful patterns for instructors, J. Res. Sci. Teach. **40**, 776 (2003).
[221] D. B. May and E. Etkina, College physics students' epistemological self-reflection and its relationship to conceptual learning, Am. J. Phys. **70**, 1249 (2002).
[222] H. G. Close and P. R. L. Heron, Student understanding of the angular momentum of classical particles, Am. J. Phys. **79**, 1068 (2011).

[223] S. J. Pollock and N. D. Finkelstein, Sustaining educational reforms in introductory physics, Phys. Rev. ST Phys. Educ. Res. **4**, 010110 (2008).

[224] A. Elby, Helping physics students learn how to learn, Am. J. Phys. **69**, S54 (2001).

[225] M. C. Wittmann, R. N. Steinberg, E. F. Redish, and the University of Maryland Physics Education Research Group, *Activity-Based Tutorials Volume 1: Introductory Physics, The Physics Suite* (John Wiley & Sons, New York, 2004).

[226] M. C. Wittmann, R. N. Steinberg, E. F. Redish, and the University of Maryland Physics Education Research Group, *Activity-Based Tutorials Volume 2: Modern Physics* (John Wiley & Sons, New York, 2004).

[227] D. R. Sokoloff, R. K. Thornton, and P. W. Laws, *RealTime Physics Active Learning Laboratories Module 1: Mechanics*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2004).

[228] D. R. Sokoloff, R. K. Thornton, and P. W. Laws, *RealTime Physics Active Learning Laboratories Module 2: Heat and Thermodynamics*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2004).

[229] D. R. Sokoloff, R. K. Thornton, and P. W. Laws, *RealTime Physics Active Learning Laboratories Module 3: Electric Circuits*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2004).

[230] D. R. Sokoloff, R. K. Thornton, and P. W. Laws, *RealTime Physics Active Learning Laboratories Module 4: Light and Optics*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2004).

[231] R. J. Beichner, The impact of video motion analysis on kinematics graph interpretation skills, Am. J. Phys. **64**, 1272 (1996).

[232] E. Etkina, and A. Van Heuvelen, Investigative Science Learning Environment – A science process approach to learning physics, in *Reviews in PER Volume 1: Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007).

[233] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. E. Hmelo-Silver, Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories, J. Learn. Sci. **19**, 54 (2010).

[234] L. McCullough, The effect of introducing computers into an introductory problem-solving laboratory, Ph.D. thesis, University of Minnesota, 2000 (unpublished).

[235] R. Lippmann Kung, Teaching the concepts of measurement: An example of a concept-based laboratory course, Am. J. Phys. **73**, 771 (2005).

[236] B. J. Duch, S. E. Groh, and D. E. Allen, *The Power of Problem-Based Learning* (Stylus Publishing, Sterling, VA, 2001).

[237] M. A. Pease and D. Kuhn, Experimental analysis of the effective components of problem-based learning, Sci. Educ. **95**, 57 (2011).

[238] L. C. McDermott and The Physics Education Group at the University of Washington, *Physics by Inquiry* (John Wiley & Sons, New York, 1996), Vol. I.

[239] L. C. McDermott and The Physics Education Group at the University of Washington, *Physics by Inquiry* (John Wiley & Sons, New York, 1996), Vol. II.

[240] P. Laws, Calculus-based physics without lectures, Phys. Today **44**, 24 (1991).

[241] P. W. Laws, *Workshop Physics Activity Guide* (John Wiley & Sons, New York, 2004), Vols. 1–4.

[242] K. Cummings, J. Marx, R. Thornton, and D. Kuhl, Evaluating innovation in studio physics, Am. J. Phys. **67**, S38 (1999).

[243] C. Hoellwarth, M. Moelter, and R. Knight, A direct comparison of conceptual learning and problem solving ability in traditional and studio style classrooms, Am. J. Phys. **73**, 459 (2005).

[244] P. B. Kohl and H. V. Kuo, Chronicling a successful secondary implementation of Studio Physics, Am. J. Phys. **80**, 832 (2012).

[245] C. M. Sorensen, A. D. Churukian, S. Maleki, and D. A. Zollman, The New Studio format for instruction of introductory physics, Am. J. Phys. **74**, 1077 (2006).

[246] J. Wilson, The CUPLE physics studio, Phys. Teach. **32**, 518 (1994).

[247] R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, D. L. Deardorff, R. J. Allain, S. W. Bonham, M. H. Dancy, and J. S. Risley, The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project, in, *Reviews in PER Volume 1: Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007).

[248] J. Belcher, Studio physics at MIT, MIT Physics Annual 2001, **58** (2001), http://web.mit.edu/physics/papers/Belcher_physicsannual_fall_01.pdf, retrieved October 6, 2010.

[249] Y. J. Dori and J. Belcher, How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts?, J. Learn. Sci. **14**, 243 (2005).

[250] Y. J. Dori, J. Belcher, M. Bessette, M. Danziger, A. McKinney, and E. Hult, Technology for active learning, Mater. Today **6**, 44 (2003).

[251] K. Cummings, P. Laws, E. Redish, and P. Cooney, *Understanding Physics* (John Wiley & Sons, Hoboken, NJ, 2004).

[252] R. W. Chabay and B. A. Sherwood, *Matter and Interactions*, 3rd ed. (John Wiley & Sons, Hoboken, NJ, 2009).

[253] W. K. Adams, S. Reid, R. LeMaster, S. B. McKagan, K. K. Perkins, M. Dubson, and C. E. Wieman, A study of educational simulations Part I: Engagement and learning, *J. Interact. Learn. Res.* **19**, 397 (2008).

[254] W. K. Adams, S. Reid, R. LeMaster, S. B. McKagan, K. K. Perkins, M. Dubson, and C. E. Wieman, A study of educational simulations Part II: Interface design, *J. Interact. Learn. Res.* **19**, 551 (2008).

[255] N. D. Finkelstein, W. K. Adams, C. J. Keller, P. B. Kohl, K. K. Perkins, N. S. Podolefsky, S. Reid, and R. LeMaster, When learning about the real world is better done virtually: A study of substituting computer

simulations for laboratory equipment, Phys. Rev. ST Phys. Educ. Res. **1**, 010103 (2005).
[256] R. Chabay and B. Sherwood, Computational physics in the introductory calculus-based course, Am. J. Phys. **76**, 307 (2008).
[257] E. F. Redish and J. M. Wilson, Student programming in the introductory physics course: M.U.P.P.E.T., Am. J. Phys. **61**, 222 (1993).
[258] I. Halloun and D. Hestenes, The initial knowledge state of college physics students, Am. J. Phys. **53**, 1043 (1985).
[259] W. J. Leonard, R. J. Dufresne, W. J. Gerace, and J. P. Mestre, *Minds-On Physics: Activities and Reader* (Kendall-Hunt Publishing, Dubuque, IA, 2000), Vols. 1–6.
[260] E. Brewe, Modeling theory applied: Modeling Instruction in introductory physics, Am. J. Phys. **76**, 1155 (2008).
[261] D. H. Jonassen and S. M. Land, *Theoretical Foundations of Learning Environments* (Lawrence Erlbaum Assoc., Mahweh, NJ, 2000).
[262] L. W. Trowbridge, R. W. Bybee, and J. C. Powell, *Teaching Secondary School Science: Strategies for Developing Scientific Literacy*, 7th ed. (Prentice-Hall, Upper Saddle River, NJ, 2000).
[263] E. Von Glasersfeld, Cognition, construction of knowledge, and teaching,, Synthese **80**, 121 (1989).
[264] J. Dewey, *Experience and Education*, 60th anniversary ed. (Kappa Delta Pi, Indianapolis, IN, 1998).
[265] D. Jonassen, Designing constructivist learning environments, in *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory*, edited by C. M. Reigeluth (Lawrence Erlbaum Assoc., Mahwah, NJ, 1999), Vol. II, pp. 217–240.
[266] R. Karplus, Science teaching and the development of reasoning, J. Res. Sci. Teach. **14**, 169 (1977).
[267] D. Zollman, in *Proceedings of ICUPE, College Park, MD, 1996*, edited by E. F. Redish and J. S. Rigden (American Institute of Physics, Melville, NY, 1997), pp. 1137–1149.
[268] C. A. Manogue, P. J. Siemens, J. Tate, K. Browne, M. L. Niess, and A. J. Wolfer, Paradigms in Physics: A new upper-division curriculum, Am. J. Phys. **69**, 978 (2001).
[269] D. E. Trowbridge and L. C. McDermott, Investigation of student understanding of the concept of velocity in one dimension, Am. J. Phys. **48**, 1020 (1980).
[270] P. S. Shaffer and L. C. McDermott, Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies, Am. J. Phys. **60**, 1003 (1992).
[271] D. E. Trowbridge and L. C. McDermott, Investigation of student understanding of the concept of acceleration in one dimension, Am. J. Phys. **49**, 242 (1981).
[272] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).
[273] C. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).
[274] N. Lasry, E. Mazur, and J. Watkins, Peer Instruction: From Harvard to the two-year college, Am. J. Phys. **76**, 1066 (2008).
[275] I. D. Beatty, W. J. Gerace, W. J. Leonard, and R. J. Dufresne, Designing effective questions for classroom response system teaching, Am. J. Phys. **74**, 31 (2006).
[276] C. Keller, N. D. Finkelstein, K. Perkins, S. J. Pollock, C. Turpen, and M. Dubson, Research-based practices for effective clicker use, AIP Conf. Proc. **951**, 128 (2007).
[277] M. C. James, The effect of grading incentive on student discourse in Peer Instruction, Am. J. Phys. **74**, 689 (2006).
[278] For more information, see http://www.iclicker.com/dnn/; http://www.einstruction.com/.
[279] J. E. Caldwell, Clickers in the large classroom: Current research and best-practice tips, CBE Life Sci. Educ. **6**, 9 (2007).
[280] N. Lasry, Clickers or flashcards: Is there really a difference?, Phys. Teach. **46**, 242 (2008).
[281] C. H. Crouch, A. P. Fagen, J. P. Callan, and E. Mazur, Classroom demonstrations: Learning tools or entertainment?, Am. J. Phys. **72**, 835 (2004).
[282] S. P. Formica, J. L. Easley, and M. C. Spraker, Transforming common-sense beliefs into Newtonian thinking through Just-In-Time Teaching, Phys. Rev. ST Phys. Educ. Res. **6**, 020106 (2010).
[283] See http://research.physics.illinois.edu/per/prelectures.html.
[284] H. R. Sadaghiani, Using multimedia learning modules in a hybrid-online course in electricity and magnetism, Phys. Rev. ST Phys. Educ. Res. **7**, 010102 (2011).
[285] H. R. Sadaghiani, Controlled study on the effectiveness of multimedia learning modules for teaching mechanics, Phys. Rev. ST Phys. Educ. Res. **8**, 010103 (2012).
[286] M. L. Rosenquist and L. C. McDermott, A conceptual approach to teaching kinematics, Am. J. Phys. **55**, 407 (1987).
[287] L. C. McDermott, M. L. Rosenquist, and E. H. van Zee, Student difficulties in connecting graphs and physics: Examples from kinematics, Am. J. Phys. **55**, 503 (1987).
[288] L. C. McDermott, P. S. Shaffer, and M. D. Somers, Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine, Am. J. Phys. **62**, 46 (1994).
[289] R. A. Lawson and L. C. McDermott, Student understanding of the work-energy and impulse-momentum theorems, Am. J. Phys. **55**, 811 (1987).
[290] T. O'Brien Pride, S. Vokos, and L. C. McDermott, The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems, Am. J. Phys. **66**, 147 (1998).
[291] F. M. Goldberg and L. C. McDermott, An investigation of student understanding of the real image formed by a converging lens or concave mirror, Am. J. Phys. **55**, 108 (1987).
[292] B. S. Ambrose, P. S. Shaffer, R. N. Steinberg, and L. C. McDermott, An investigation of student understanding of single-slit diffraction and double-slit interference, Am. J. Phys. **67**, 146 (1999).

[293] S. Vokos, P. S. Shaffer, B. S. Ambrose, and L. C. McDermott, Student understanding of the wave nature of matter: Diffraction and interference of particles, Am. J. Phys. **68**, S42 (2000).
[294] K. Wosilait, P. R. L. Heron, P. S. Shaffer, and L. C. McDermott, Addressing student difficulties in applying a wave model to the interference and diffraction of light, Am. J. Phys. **67**, S5 (1999).
[295] C. H. Kautz, P. R. L. Heron, M. E. Loverude, and L. C. McDermott, Student understanding of the ideal gas law, Part I: A macroscopic perspective, Am. J. Phys. **73**, 1055 (2005).
[296] C. H. Kautz, P. R. L. Heron, P. S. Shaffer, and L. C. McDermott, Student understanding of the ideal gas law, Part II: A microscopic perspective, Am. J. Phys. **73**, 1064 (2005).
[297] M. E. Loverude, C. H. Kautz, and P. R. L. Heron, Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas, Am. J. Phys. **70**, 137 (2002).
[298] M. J. Cochran and P. R. L. Heron, Development and assessment of research-based tutorials on heat engines and the second law of thermodynamics, Am. J. Phys. **74**, 734 (2006).
[299] M. E. Loverude, P. R. L. Heron, and C. H. Kautz, Identifying and addressing student difficulties with hydrostatic pressure, Am. J. Phys. **78**, 75 (2010).
[300] L. G. Ortiz, P. R. L. Heron, and P. S. Shaffer, Student understanding of static equilibrium: Predicting and accounting for balancing, Am. J. Phys. **73**, 545 (2005).
[301] H. G. Close and P. R. L. Heron, Research as a guide for improving student learning: An example from momentum conservation, Am. J. Phys. **78**, 961 (2010).
[302] R. N. Steinberg, M. C. Wittmann, and E. F. Redish, Mathematical tutorials in introductory physics, AIP Conf. Proc. **399**, 1075 (1997).
[303] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, Phys. Rev. ST Phys. Educ. Res. **1**, 010101 (2005).
[304] H. V. Mauk and D. Hingley, Student understanding of induced current: Using tutorials in introductory physics to teach electricity and magnetism, Am. J. Phys. **73**, 1164 (2005).
[305] E. F. Redish, J. M. Saul, and R. N. Steinberg, On the effectiveness of active-engagement microcomputer-based laboratories, Am. J. Phys. **65**, 45 (1997).
[306] T. I. Smith and M. C. Wittmann, Comparing three methods for teaching Newton's third law, Phys. Rev. ST Phys. Educ. Res. **3**, 020105 (2007).
[307] D. W. Johnson, R. T. Johnson, and E. J. Holubec, *Cooperation in the Classroom* (Interaction Book Company, Edina, MN, 1988).
[308] D. W. Johnson, R. T. Johnson, and K. A. Smith, *Active Learning: Cooperation in the College Classroom* (Interaction Book Company, Edina, MN, 1991).
[309] H. Brasell, The effect of real-time laboratory graphing on learning graphic representations of distance and velocity, J. Res. Sci. Teach. **24**, 385 (1987).
[310] R. K. Thornton and D. R. Sokoloff, Learning motion concepts using real-time microcomputer-based laboratory tools, Am. J. Phys. **58**, 858 (1990).
[311] R. Beichner, The effect of simultaneous motion presentation and graph generation in a kinematics lab, J. Res. Sci. Teach. **27**, 803 (1990).
[312] J. Brungardt and D. Zollman, Influence of interactive videodisc instruction using simultaneous-time analysis on kinematics graphing skills of high school physics students, J. Res. Sci. Teach. **32**, 855 (1995).
[313] See http://groups.physics.umn.edu/physed/Research/CPSL/MBL.html.
[314] R. K. Thornton and D. R. Sokoloff, RealTime Physics: Active learning laboratory, AIP Conf. Proc. **399**, 1101 (1997).
[315] D. Zollman and R. Fuller, Teaching and learning physics with interactive video, Phys. Today **47**, 41 (1994).
[316] L. T. Escalada and D. A. Zollman, An investigation of the effects of using interactive digital video in a physics classroom on student learning and attitudes, J. Res. Sci. Teach. **34**, 467 (1997).
[317] E. Etkina, S. Murthy, and X. Zou, Using introductory labs to engage students in experimental design, Am. J. Phys. **74**, 979 (2006).
[318] W. M. Reid and D. F. Arseneau, Labs of unlimited scope, Am. J. Phys. **39**, 271 (1971).
[319] A. Van Heuvelen and E. Etkina, *The Physics Active Learning Guide* (Addison-Wesley, San Francisco, 2006).
[320] See also http://www.islephysics.net/.
[321] A. Karelina and E. Etkina, Acting like a physicist: Student approach study to experimental design, Phys. Rev. ST Phys. Educ. Res. **3**, 020106 (2007).
[322] See http://groups.physics.umn.edu/physed/Research/PSL/pslintro.html.
[323] See http://umdperg.pbworks.com/Scientific-Community-Labs.
[324] For more information, see http://www.udel.edu/inst.
[325] B. J. Duch, Problem-based learning in physics: The power of students teaching students, *J. Coll. Sci. Teach.* **15**, 326 (1996).
[326] B. J. Duch, Problem-based learning in physics: Making connections with the real world, AIP Conf. Proc. **399**, 557 (1997).
[327] See also http://www.phys.washington.edu/groups/peg/pbi.html.
[328] P. R. L. Heron, M. E. Loverude, P. S. Shaffer, and L. C. McDermott, Helping students develop an understanding of Archimedes' principle. II. Development of research-based instructional materials, Am. J. Phys. **71**, 1188 (2003).
[329] B. Thacker, E. Kim, K. Trefz, and S. M. Lea, Comparing problem solving performance of physics students in inquiry-based and traditional introductory physics courses, Am. J. Phys. **62**, 627 (1994).
[330] See http://physics.dickinson.edu/~wp_web/WP_homepage.html.
[331] See http://physics.dickinson.edu/~eip_web/EiP_homepage.html.
[332] See http://www.ncsu.edu/per/scaleup.html.
[333] See http://scaleup.ncsu.edu/FAQs.html.

[334] J. Gaffney, E. Richards, M. B. Kustusch, L. Ding, and R. Beichner, Scaling up education reform, *J. Coll. Sci. Teach.* **37**, 48 (2008).

[335] Y. J. Dori and J. Belcher, in *Visualization in Science Education*, edited by J. K. Gilbert (Springer, Amsterdam, 2005), pp. 187–216.

[336] R. D. Knight, *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics and Mastering Physics* (Addison-Wesley, San Francisco, 2004).

[337] T. A. Moore, *Six Ideas That Shaped Physics* (McGraw-Hill, New York, 2003).

[338] See http://www.physics.pomona.edu/sixideas/sisuc.html.

[339] R. W. Chabay and B. A. Sherwood, Bringing atoms into first-year physics, Am. J. Phys. **67**, 1045 (1999).

[340] R. Chabay and B. Sherwood, Modern mechanics, Am. J. Phys. **72**, 439 (2004).

[341] R. Chabay and B. Sherwood, Restructuring the introductory electricity and magnetism course, Am. J. Phys. **74**, 329 (2006).

[342] M. A. Kohlmyer, M. D. Caballero, R. Catrambone, R. W. Chabay, L. Ding, M. P. Haugan, M. J. Marr, B. A. Sherwood, and M. F. Schatz, Tale of two curricula: The performance of 2000 students in introductory electromagnetism, Phys. Rev. ST Phys. Educ. Res. **5**, 020105 (2009).

[343] M. D. Caballero, E. F. Greco, E. R. Murray, K. R. Bujak, and M. J. Marr, Comparing large lecture mechanics curricula using the Force Concept Inventory: A five thousand student study, Am. J. Phys. **80**, 638 (2012).

[344] M. H. Dancy and R. Beichner, Impact of animation on assessment of conceptual understanding in physics, Phys. Rev. ST Phys. Educ. Res. **2**, 010104 (2006).

[345] S. Yeo, R. Loss, M. Zadnik, A. Harrison, and D. Treagust, What do students really learn from interactive multimedia? A physics case study, Am. J. Phys. **72**, 1351 (2004).

[346] R. N. Steinberg, Computers in teaching science: To simulate or not to simulate?, Am. J. Phys. **68**, S37 (2000).

[347] J. M. Casperson and M. C. Linn, Using visualizations to teach electrostatics, Am. J. Phys. **74**, 316 (2006).

[348] See, for example, ActivPhysics OnLine, http://www.aw-bc.com/knight.

[349] K. Perkins, W. Adams, M. Dubson, N. Finkelstein, S. Reid, C. Wieman, and R. LeMaster, PhET: Interactive simulations for teaching and learning physics, Phys. Teach. **44**, 18 (2006).

[350] S. B. McKagan, K. K. Perkins, M. Dubson, C. Malley, S. Reid, R. LeMaster, and C. E. Wieman, Developing and researching PhET simulations for teaching quantum mechanics, Am. J. Phys. **76**, 406 (2008).

[351] Special double issue of the *American Journal of Physics* on incorporating computation into the physics curriculum and using computation to enhance physics teaching [*Am. J. Phys.* **76**, Issue No. 4 (2008)].

[352] L. Bao, S. R. Stonebraker, and H. Sadaghiani, A flexible homework method, Am. J. Phys. **76**, 878 (2008).

[353] See an overview of the research project and sample materials at http://cpucips.sdsu.edu/petpset.html.

[354] V. K. Otero and K. E. Gray, Attitudinal gains across multiple universities using the Physics and Everyday Thinking curriculum, Phys. Rev. ST Phys. Educ. Res. **4**, 020104 (2008).

[355] See http://www.iat.com/courses/high-school-science/active-physics/.

[356] See http://cpucips.sdsu.edu/interactionsinfo/.

[357] D. H. McIntyre, J. Tate, and C. A. Manogue, Integrating computational activities into the upper-level Paradigms in Physics curriculum at Oregon State University, Am. J. Phys. **76**, 340 (2008).

[358] B. S. Ambrose, Investigating student understanding in intermediate mechanics: Identifying the need for a tutorial approach to instruction, Am. J. Phys. **72**, 453 (2004).

[359] S. V. Chasteen and S. J. Pollock, A researched-based approach to assessing student learning issues in upper-division electricity and magnetism, AIP Conf. Proc. **1179**, 7 (2009).

[360] S. V. Chasteen, S. J. Pollock, R. E. Pepper, and K. K. Perkins, Thinking like a physicist: A multi-semester case study of junior-level electricity and magnetism, Am. J. Phys. **80**, 923 (2012).

[361] S. V. Chasteen, S. J. Pollock, R. E. Pepper, and K. K. Perkins, Transforming the junior level: Outcomes from instruction and research in E&M, Phys. Rev. ST Phys. Educ. Res. **8**, 020107 (2012).

[362] S. J. Pollock, S. V. Chasteen, M. Dubson, and K. K. Perkins, The use of concept tests and peer instruction in upper-division physics, AIP Conf. Proc. **1289**, 261 (2010).

[363] S. Goldhaber, S. Pollock, M. Dubson, P. Beale, and K. Perkins, Transforming upper-division quantum mechanics: Learning goals and assessment, AIP Conf. Proc. **1179**, 145 (2009).

[364] L. D. Carr and S. B. McKagan, Graduate quantum mechanics reform, Am. J. Phys. **77**, 308 (2009).

[365] V. Otero, S. Pollock, and N. Finkelstein, A physic department's role in preparing physics teachers: The Colorado learning assistant model, Am. J. Phys. **78**, 1218 (2010).

[366] R. M. Goertzen, E. Brewe, L. H. Kramer, L. Wells, and D. Jones, Moving toward change: Institutionalizing reform through implementation of the Learning Assistant model and Open Source Tutorial, Phys. Rev. ST Phys. Educ. Res. **7**, 020105 (2011).

[367] See http://www.aapt.org/Conferences/newfaculty/nfw.cfm.

[368] C. Henderson and M. H. Dancy, Impact of physics education research on the teaching of introductory quantitative physics in the United States, Phys. Rev. ST Phys. Educ. Res. **5**, 020107 (2009).

[369] M. Dancy and C. Henderson, Pedagogical practices and instructional change of physics faculty, Am. J. Phys. **78**, 1056 (2010).

[370] C. Singh, Student understanding of quantum mechanics, Am. J. Phys. **69**, 885 (2001).

[371] E. F. Redish and D. Hammer, Reinventing college physics for biologists: Explicating an epistemological curriculum, Am. J. Phys. **77**, 629 (2009).

[372] M. Plomer, K. Jessen, G. Rangelov, and M. Meyer, Teaching physics in a physiologically meaningful manner, Phys. Rev. ST Phys. Educ. Res. **6**, 020116 (2010).

[373] J. L. Campbell, A quantitative comparison of the composition and linguistic properties of introductory physics textbooks, Ph.D. thesis, University of Arkansas, 2008 (unpublished).

[374] *Cognitive Load Theory*, edited by J. L. Plass, R. Moreno, and R. Brunken (Cambridge University Press, New York, 2010).

[375] J. Sweller, in *The Psychology of Learning, and Motivation: Cognition in Education*, edited by J. P. Mestre and B. H. Ross (Elsevier, San Diego, 2011).

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

[377] See a list at http://www.ncsu.edu/per/TestInfo.html.

[378] P. V. Engelhardt, in *Reviews in PER Vol. 2: Getting Started in Physics Education Research*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2009), http://www.per-central.org/document/ServeFile.cfm?ID=8807, retrieved August 10, 2010.

[379] D. Treagust, Evaluating students' misconceptions by means of diagnostic multiple choice items, Res. Sci. Educ. **16**, 199 (1986).

[380] R. S. Lindell, E. Peak, and T. M. Foster, Are they all created equal? A comparison of different concept inventory development methodologies, AIP Conf. Proc. **883**, 14 (2007).

[381] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, Phys. Rev. ST Phys. Educ. Res. **5**, 020103 (2009).

[382] L. Ding, N. W. Reay, A. Lee, and L. Bao, Effects of testing conditions on conceptual survey results, Phys. Rev. ST Phys. Educ. Res. **4**, 010112 (2008).

[383] V. P. Coletta, J. A. Phillips, and J. J. Steinert, Interpreting force concept inventory scores: Normalized gain and SAT scores, Phys. Rev. ST Phys. Educ. Res. **3**, 010106 (2007).

[384] V. P. Coletta and J. A. Phillips, Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, Am. J. Phys. **73**, 1172 (2005).

[385] D. E. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores, Am. J. Phys. **70**, 1259 (2002).

[386] E. Kuo, M. M. Hull, A. Gupta, and A. Elby, How students blend conceptual and formal mathematical reasoning in solving physics problems, Sci. Educ. **97**, 32 (2013).

[387] R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx, Comparing the force and motion conceptual evaluation and the force concept inventory, Phys. Rev. ST Phys. Educ. Res. **5**, 010105 (2009).

[388] L. Bao, T. Cai, K. Koenig, K. Fang, J. Han, J. Wang, Q. Liu, L. Ding, L. Cui, Y. Luo, Y. Luo, Y. Wang, L. Li, and N. Wu, Learning and scientific reasoning, Science **323**, 586 (2009).

[389] L. Bao, K. Fang, T. Cai, J. Wang, L. Yang, L. Cui, J. Han, L. Ding, and Y. Luo, Learning of content knowledge and development of scientific reasoning ability: A cross culture comparison, Am. J. Phys. **77**, 1118 (2009).

[390] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5**, 010101 (2009).

[391] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. **74**, 118 (2006).

[392] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?, Phys. Rev. ST Phys. Educ. Res. **3**, 010107 (2007).

[393] G. Kortemeyer, Gender differences in the use of an online homework system in an introductory physics course, Phys. Rev. ST Phys. Educ. Res. **5**, 010107 (2009).

[394] C. T. Richardson and B. W. O'Shea, Assessing gender differences in response system questions for an introductory physics course, Am. J. Phys. **81**, 231 (2009).

[395] H. T. Hudson and C. K. Hudson, Suggestions on the construction of multiple-choice tests, Am. J. Phys. **49**, 838 (1981).

[396] M. Scott, T. Stelzer, and G. Gladding, Evaluating multiple-choice exams in large introductory physics courses, Phys. Rev. ST Phys. Educ. Res. **2**, 020102 (2006).

[397] M. M. Hull, E Kuo, A. Gupta, and A. Elby, Problem-solving rubrics revisited: Attending to blending of informal conceptual and formal mathematical reasoning, Phys. Rev. ST Phys. Educ. Res. **9**, 010105 (2013).

[398] J. D. Marx and K. Cummings, Development of a survey instrument to gauge students' problem-solving ability, AIP Conf. Proc. **1289**, 221 (2010).

[399] 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).

[400] D. E. Pritchard, Y.-J. Lee, and L. Bao, Mathematical learning models that depend on prior knowledge and instructional strategies, Phys. Rev. ST Phys. Educ. Res. **4**, 010109 (2008).

[401] A. F. Heckler and E. C. Sayre, What happens between pre- and post-tests: Multiple measurements of student understanding during an introductory physics course, Am. J. Phys. **78**, 768 (2010).

[402] E. C. Sayre and A. F. Heckler, Peaks and decays of student knowledge in an introductory E&M course, Phys. Rev. ST Phys. Educ. Res. **5**, 013101 (2009).

[403] See http://www.teststandards.org.

[404] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing* (American Educational Research Association, Washington, DC, 1999).

[405] P. Heller and D. Huffman, Interpreting the Force Concept Inventory: A reply to Hestenes and Halloun, Phys. Teach. **33**, 503 (1995).

[406] D. Hestenes and I. Halloun, Interpreting the Force Concept Inventory: A response to March 1995 critique by Huffman and Heller, Phys. Teach. **33**, 502 (1995).

[407] D. Huffman and P. Heller, What does the Force Concept Inventory actually measure?, Phys. Teach. **33**, 138 (1995).

[408] S. Ramlo, Validity and reliability of the force and motion conceptual evaluation, Am. J. Phys. **76**, 882 (2008).
[409] C. Singh and D. Rosengrant, Multiple-choice test of energy and momentum concepts, Am. J. Phys. **71**, 607 (2003).
[410] P. V. Engelhardt and R. J. Beichner, Students' understanding of direct current resistive electrical circuits, Am. J. Phys. **72**, 98 (2004).
[411] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 010105 (2006).
[412] P. Nieminen, A. Savinainen, and J. Viiri, Force concept inventory-based multiple-choice test for investigating students' representational consistency, Phys. Rev. ST Phys. Educ. Res. **6**, 020109 (2010).
[413] S. B. McKagan, K. K. Perkins, and C. E. Wieman, Design and validation of the Quantum Mechanics Conceptual Survey, Phys. Rev. ST Phys. Educ. Res. **6**, 020121 (2010).
[414] S. V. Chasteen, R. E. Pepper, M. D. Caballero, S. J. Pollock, and K. K. Perkins, Colorado Upper-Division Electrostatics diagnostic: A conceptual assessment for the junior level, Phys. Rev. ST Phys. Educ. Res. **8**, 020108 (2012).
[415] R. Rosenblatt and A. F. Heckler, Systematic study of student understanding of the relationships between the direction of force, velocity, and acceleration in one dimension, Phys. Rev. ST Phys. Educ. Res. **7**, 020112 (2011).
[416] A. E. Lawson, Development and validation of the classroom test of formal reasoning, J. Res. Sci. Teach. **15**, 11 (1978).
[417] L. Bao, Theoretical comparisons of average normalized gain calculations, Am. J. Phys. **74**, 917 (2006).
[418] J. D. Marx and K. Cummings, Normalized change, Am. J. Phys. **75**, 87 (2007).
[419] G. A. Morris, L. Branum-Martin, N. Harshman, S. D. Baker, E. Mazur, S. Dutta, T. Mzoughi, and V. McCauley, Testing the test: Item response curves and test quality, Am. J. Phys. **74**, 449 (2006).
[420] J. Wang and L. Bao, Analyzing force concept inventory with item response theory, Am. J. Phys. **78**, 1064 (2010).
[421] T. I. Smith and M. C. Wittmann, Applying a resources framework to analysis of the Force and Motion Conceptual Evaluation, Phys. Rev. ST Phys. Educ. Res. **4**, 020101 (2008).
[422] J. Stewart, M. Miller, C. Audo, and G. Stewart, Using cluster analysis to identify patterns in students' responses to contextually different conceptual problems, Phys. Rev. ST Phys. Educ. Res. **8**, 020112 (2012).
[423] M. Planinic, L. Ivanjek, and A. Susac, Rasch model based analysis of the Force Concept Inventory, Phys. Rev. ST Phys. Educ. Res. **6**, 010103 (2010).
[424] L. Bao and E. F. Redish, Concentration analysis: A quantitative assessment of student states, Am. J. Phys. **69**, S45 (2001).
[425] C. Henderson, Common concerns about the Force Concept Inventory, Phys. Teach. **40**, 542 (2002).
[426] S. Bonham, Reliability, compliance, and security in web-based course assessments, Phys. Rev. ST Phys. Educ. Res. **4**, 010106 (2008).
[427] J. Stewart, H. Griffin, and G. Stewart, Context sensitivity in the force concept inventory, Phys. Rev. ST Phys. Educ. Res. **3**, 010102 (2007).
[428] B. W. Frank and R. E. Scherr, Interactional processes for stabilizing conceptual coherences in physics, Phys. Rev. ST Phys. Educ. Res. **8**, 020101 (2012).
[429] A. B. Champagne and L. E. Klopfer, A causal model of students' achievement in a college physics course, J. Res. Sci. Teach. **19**, 299 (1982).
[430] A. B. Champagne, L. E. Klopfer, and J. H. Anderson, Factors influencing the learning of classical mechanics, Am. J. Phys. **48**, 1074 (1980).
[431] J. Docktor and K. Heller, Gender differences in both force concept inventory and introductory physics performance, AIP Conf. Proc. **1064**, 15 (2008).
[432] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory physics performance: The influence of high school physics preparation and affective factors, Sci. Educ. **91**, 847 (2007).
[433] P. M. Sadler and R. H. Tai, Success in introductory college physics: The role of high school preparation, Sci. Educ. **85**, 111 (2001).
[434] R. H. Tai and P. M. Sadler, Gender differences in introductory undergraduate physics performance: University physics versus college physics in the USA, Int. J. Sci. Educ. **23**, 1017 (2001).
[435] L. E. Kost-Smith, S. J. Pollock, and N. D. Finkelstein, Gender disparities in second-semester college physics: The incremental effects of a "smog of bias", Phys. Rev. ST Phys. Educ. Res. **6**, 020112 (2010).
[436] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, Reducing the gender achievement gap in college science: A classroom study of values affirmation, Science **330**, 1234 (2010).
[437] W. Fakcharoenphol, E. Potter, and T. Stelzer, What students learn when studying physics practice exam problems, Phys. Rev. ST Phys. Educ. Res. **7**, 010107 (2011).
[438] D. J. Palazzo, Y.-L. Lee, R. Warnakulasooriya, and D. E. Pritchard, Patterns, correlates, and reduction of homework copying, Phys. Rev. ST Phys. Educ. Res. **6**, 010104 (2010).
[439] K. Cummings and J. D. Marx, Beta-test data on an assessment of textbook problem solving ability: An argument for right/wrong grading?, AIP Conf. Proc. **1289**, 113 (2010).
[440] J. Stewart and S. Ballard, Effect of written presentation on performance in introductory physics, Phys. Rev. ST Phys. Educ. Res. **6**, 020120 (2010).
[441] M. A. McDaniel, H. L. Roediger III, and K. B. McDermott, Generalizing test-enhanced learning from the laboratory to the classroom, Psychon. Bull. Rev. **14**, 200 (2007).
[442] For a discussion of the relationship between discipline-based education research and other research areas such as educational psychology and cognitive science, refer to the introduction of Ref. [2].

[443] M. L. Gick and K. J. Holyoak, Schema induction and analogical transfer, Cogn. Psychol. **15**, 1 (1983).
[444] S. K. Reed, in *Transfer on Trial: Intelligence, Cognition, and Instruction*, edited by D. K. Detterman and R. J. Sternberg (Ablex, Norwood, NJ, 1993), pp. 39–67.
[445] A. Feil and J. Mestre, Change blindness as a means of studying expertise in physics, J. Learn. Sci. **19**, 480 (2010).
[446] J. Mestre, Is transfer ubiquitous or rare? New paradigms for studying transfer, AIP Conf. Proc. **790**, 3 (2005).
[447] R. Dufresne, J. P. Mestre, T. Thaden-Koch, W. Gerace, and W. Leonard, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Publishing, Greenwich, CT, 2005), pp. 155–215.
[448] C. Singh, Transfer of learning in quantum mechanics, AIP Conf. Proc. **790**, 23 (2005).
[449] T. J. Nokes-Malach and J. P. Mestre, Toward a model of transfer as sense-making, Educ. Psychol. **48**, 184 (2013).
[450] N. S. Rebello, D. A. Zollman, A. R. Allbaugh, P. V. Engelhardt, K. E. Gray, Z. Hrepic, and S. F. Itza-Ortiz, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Publishing, Greenwich, CT, 2005).
[451] J. Mestre, "Transfer of learning: Issues and research agenda," National Science Foundation Report No. NSF03-212, 2003, www.nsf.gov/pubs/2003/nsf03212/nsf03212.pdf, retrieved October 4, 2010.
[452] Special issue on transfer strand [J. Learn. Sci. **15**, Issue No. 4, 431 (2006)].
[453] *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. A. Ericsson, N. Charness, P. Feltovich, and R. Hoffman (Cambridge University Press, Cambridge, England, 2006).
[454] *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Publishing, Greenwich, CT, 2005).
[455] J. H. Flavell, Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry, Am. Psychol. **34**, 906 (1979).
[456] G. Taasoobshirazi and J. Farley, Construct validation of the Physics Metacognitive Inventory, Int. J. Sci. Educ. **35**, 447 (2013).
[457] D. L. Medin, Concepts and conceptual structure, Am. Psychol. **44**, 1469 (1989).
[458] D. L. Medin and L. J. Rips, in *The Cambridge Handbook of Thinking and Reasoning*, edited by K. J. Holyoak and R. G. Morrison (Cambridge University Press, New York, 2005), pp. 37–72.
[459] B. L. Whorf, in *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, edited by J. B. Carroll (The MIT Press, Cambridge, MA, 1956).
[460] C. R. Sutton, Figuring out a scientific understanding, J. Res. Sci. Teach. **30**, 1215 (1993).
[461] D. T. Brookes and E. Etkina, Using conceptual metaphor and functional grammar to explore how language used in physics affects student learning, Phys. Rev. ST Phys. Educ. Res. **3**, 010105 (2007).
[462] D. T. Brookes and E. Etkina, Force, ontology, and language, Phys. Rev. ST Phys. Educ. Res. **5**, 010110 (2009).
[463] N. S. Podolefsky and N. D. Finkelstein, Use of analogy in learning physics: The role of representations, Phys. Rev. ST Phys. Educ. Res. **2**, 020101 (2006).
[464] N. S. Podolefsky and N. D. Finkelstein, Analogical scaffolding and the learning of abstract ideas in physics: An example from electromagnetic waves, Phys. Rev. ST Phys. Educ. Res. **3**, 010109 (2007).
[465] N. S. Podolefsky and N. D. Finkelstein, Analogical scaffolding and the learning of abstract ideas in physics: Empirical studies, Phys. Rev. ST Phys. Educ. Res. **3**, 020104 (2007).
[466] A. A. diSessa, An interactional analysis of clinical interviewing, Cognit. Instr. **25**, 523 (2007).
[467] K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports as Data (revised edition)* (Bradford Books, Cambridge, MA, 1993).
[468] V. R. Lee, R. S. Russ, and B. Sherin, in *Proceedings of the 30th Annual Meeting of the Cognitive Science Society, Austin, TX, 2008*, edited by V. Sloutsky, B. Love, and K. McRae (Cognitive Science Society, 2008), pp. 1723–1728, http://csjarchive.cogsci.rpi.edu/Proceedings/2008/pdfs/p1723.pdf.
[469] R. E. Nisbett and T. D. Wilson, Telling more than we can know: Verbal reports on mental processes, Psychol. Rev. **84**, 231 (1977).
[470] M. T. H. Chi, N. De Leeuw, M.-H. Chiu, and C. Lavancher, Eliciting self-explanations improves understanding, Cogn. Sci. **18**, 439 (1994).
[471] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 1992).
[472] J. Confrey, in *The Cambridge Handbook of the Learning Sciences*, edited by R. K. Sawyer (Cambridge University Press, Cambridge, England, 2006), pp. 135–152.
[473] T. Lamberg and J. A. Middleton, Design research perspectives on transitioning from individual microgenetic interviews to a whole-class teaching experiment, Educ. Res. **38**, 233 (2009).
[474] M. T. H. Chi, in *Advances in Instructional Psychology: Educational Design and Cognitive Science*, edited by R. Glaser (Lawrence Erlbaum Assoc., Mahwah, NJ, 2000), Vol. 5, pp. 161–238.
[475] A. B. Hoffman and B. Rehder, The costs of supervised classification: The effect of learning task on conceptual flexibility, J. Exp. Psychol. **139**, 319 (2010).
[476] K. Rayner, Eye movements in reading and information processing: 20 years of research, Psychol. Bull. **124**, 372 (1998).
[477] D. J. Simons and D. T. Levin, Failure to detect changes to people during a real-world interaction, Psychon. Bull. Rev. **5**, 644 (1998).
[478] C. Singh, Categorization of problems to assess and improve proficiency as teachers and learners, Am. J. Phys. **77**, 73 (2009).
[479] G. Bateson, in *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and*

Epistemology, edited by G. Bateson (Ballantine, New York, 1972), pp. 177–193.
[480] E. Goffman, *Frame Analysis: An Essay on the Organization of Experience* (Northeastern University Press, Boston, MA, 1986).
[481] G. L. MacLachlan and I. Reid, *Framing and Interpretation* (Melbourne University Press, Portland, OR, 1994).
[482] D. Tannen, *Framing in Discourse* (Oxford University Press, New York, 1993).
[483] E. F. Redish, A theoretical framework for physics education research: Modeling student thinking, in *Proceedings of the International School of Physics "Enrico Fermi," Course CLVI, Varenna, Italy, 2003*, edited by E. Redish, C. Tarsitani, and M. Vicentini (IOS Press, Amsterdam, 2004).
[484] B. H. Ross, Remindings and their effects in learning a cognitive skill, Cogn. Psychol. **16**, 371 (1984).
[485] D. T. Brookes, B. H. Ross, and J. P. Mestre, Specificity, transfer, and the development of expertise, Phys. Rev. ST Phys. Educ. Res. **7**, 010105 (2011).
[486] J. P. Mestre, B. H. Ross, D. T. Brookes, A. D. Smith, and T. J. Nokes, in *Fostering Scientific Habits of Mind: Pedagogical Knowledge and Best Practices in Science Education*, edited by I. M. Saleh and M. S. Khine (Sense Publishers, Rotterdam, Netherlands, 2009), pp. 145–171.
[487] A. M. Madsen, A. M. Larson, L. C. Loschky, and N. S. Rebello, Differences in visual attention between those who correctly and incorrectly answer physics problems, Phys. Rev. ST Phys. Educ. Res. **8**, 010122 (2012).
[488] T. Van Gog, F. Paas, and J. J. G. Van Merriënboer, Uncovering expertise-related differences in troubleshooting performance: Combining eye movement data and concurrent verbal protocol data, Appl. Cogn. Psychol. **19**, 205 (2005).
[489] A. C. Graesser, S. Lu, B. A. Olde, E. Cooper-Pye, and S. Whitten, Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down, Mem. Cogn. **33**, 1235 (2005).
[490] M. Hegarty, in *Eye Movements and Visual Cognition*, edited by K. Rayner (Springer-Verlag, New York, 1992), pp. 428–443.
[491] M. Kozhevnikov, M. A. Motes, and M. Hegarty, Spatial visualization in physics problem solving, Cogn. Sci. **31**, 549 (2007).
[492] D. Rosengrant, C. Thomson, and T. Mzoughi, Comparing experts and novices in solving electrical circuit problems with the help of eye-tracking, AIP Conf. Proc. **1179**, 249 (2009).
[493] A. Mason and C. Singh, Assessing expertise in introductory physics using categorization task, Phys. Rev. ST Phys. Educ. Res. **7**, 020110 (2011).
[494] R. K. Atkinson, S. J. Derry, A. Renkl, and D. W. Wortham, Learning from examples: Instructional principles from the worked examples research, Rev. Educ. Res. **70**, 181 (2000).
[495] J. Sweller, J. J. G. Van Merriënboer, and F. G. W. C. Paas, Cognitive architecture and instructional design, Educ. Psychol. Rev. **10**, 251 (1998).
[496] F. Paas, Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach, J. Educ. Psychol. **84**, 429 (1992).
[497] F. G. W. C. Paas and J. J. G. Van Merriënboer, Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach, J. Educ. Psychol. **86**, 122 (1994).
[498] M. T. H. Chi and K. VanLehn, The content of physics self-explanations, J. Learn. Sci. **1**, 69 (1991).
[499] D. T. Brookes, The role of language in learning physics, Ph.D. thesis (Rutgers, The State University of New Jersey, 2006 (unpublished).
[500] K. K. Perkins, W. K. Adams, S. J. Pollock, N. D. Finkelstein, and C. E. Wieman, Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey, AIP Conf. Proc. **790**, 61 (2005).
[501] D. Hammer, Epistemological beliefs in introductory physics, Cognit. Instr. **12**, 151 (1994).
[502] D. Hammer, Epistemological considerations in teaching introductory physics, Sci. Educ. **79**, 393 (1995).
[503] I. Halloun, Views about science, and physics achievement: The VASS story, AIP Conf. Proc. **399**, 605 (1997).
[504] U. Wutchana and N. Emarat, Student effort expectations and their learning in first-year introductory physics: A case study in Thailand, Phys. Rev. ST Phys. Educ. Res. **7**, 010111 (2011).
[505] D. Hammer and A. Elby, Tapping epistemological resources for learning physics, J. Learn. Sci. **12**, 53 (2003).
[506] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66**, 212 (1998).
[507] I. Halloun and D. Hestenes, Interpreting VASS dimensions and profiles for physics students, Sci. Educ. **7**, 553 (1998).
[508] B. White, A. Elby, J. Frederiksen, and C. Schwarz, in *Proceedings of the American Education Research Association, Montreal, 1999* (unpublished).
[509] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, Phys. Rev. ST Phys. Educ. Res. **2**, 010101 (2006).
[510] A. Mason and C. Singh, Surveying graduate students' attitudes and approaches to problem solving, Phys. Rev. ST Phys. Educ. Res. **6**, 020124 (2010).
[511] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I: Mapping the common core, Phys. Rev. ST Phys. Educ. Res. **3**, 020109 (2007).
[512] C. Henderson and M. H. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics, Phys. Rev. ST Phys. Educ. Res. **3**, 020102 (2007).

[513] C. Henderson, E. Yerushalmi, V. H. Kuo, K. Heller, and P. Heller, Physics faculty beliefs and values about the teaching and learning of problem solving. II: Procedures for measurement and analysis, Phys. Rev. ST Phys. Educ. Res. **3**, 020110 (2007).

[514] E. Yerushalmi, E. Cohen, K. Heller, P. Heller, and C. Henderson, Instructors' reasons for choosing problem features in a calculus-based introductory physics course, Phys. Rev. ST Phys. Educ. Res. **6**, 020108 (2010).

[515] C. Henderson, A. Beach, and N. Finkelstein, Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature, J. Res. Sci. Teach. **48**, 952 (2011).

[516] C. Henderson and M. H. Dancy, Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations, Am. J. Phys. **76**, 79 (2008).

[517] C. Turpen and N. D. Finkelstein, Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction, Phys. Rev. ST Phys. Educ. Res. **5**, 020101 (2009).

[518] C. Turpen and N. D. Finkelstein, The construction of different classroom norms during Peer Instruction: Students perceive differences, Phys. Rev. ST Phys. Educ. Res. **6**, 020123 (2010).

[519] R. M. Goertzen, R. E. Scherr, and A. Elby, Accounting for tutorial teaching assistants' buy-in to reform instruction, Phys. Rev. ST Phys. Educ. Res. **5**, 020109 (2009).

[520] R. M. Goertzen, R. E. Scherr, and A. Elby, Tutorial teaching assistants in the classroom: Similar teaching behaviors are supported by varied beliefs about teaching and learning, Phys. Rev. ST Phys. Educ. Res. **6**, 010105 (2010).

[521] K. M. Koenig, R. J. Endorf, and G. A. Braun, Effectiveness of different tutorial recitation teaching methods and its implications for TA training, Phys. Rev. ST Phys. Educ. Res. **3**, 010104 (2007).

[522] D. Hammer, Two approaches to learning physics, Phys. Teach. **27**, 664 (1989).

[523] M. Dancy and C. Henderson, Framework for articulating instructional practices and conceptions, Phys. Rev. ST Phys. Educ. Res. **3**, 010103 (2007).

[524] G. Kortemeyer, Correlations between student discussion behavior, attitudes, and learning, Phys. Rev. ST Phys. Educ. Res. **3**, 010101 (2007).

[525] B. K. Hofer, Personal epistemology research: Implications for learning and teaching, J. Educ. Psychol. Rev. **13**, 353 (2001).

[526] B. K. Hofer and P. R. Pintrich, The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning, Rev. Educ. Res. **67**, 88 (1997).

[527] A. Elby, Another reason that physics students learn by rote, Am. J. Phys. **67**, S52 (1999).

[528] K. E. Gray, W. K. Adams, C. E. Wieman, and K. K. Perkins, Students know what physicists believe, but they don't agree: A study using the CLASS survey, Phys. Rev. ST Phys. Educ. Res. **4**, 020106 (2008).

[529] E. Gire, B. Jones, and E. Price, Characterizing the epistemological development of physics majors, Phys. Rev. ST Phys. Educ. Res. **5**, 010103 (2009).

[530] B. A. Lindsey, L. Hsu, H. Sadaghiani, J. W. Taylor, K. Cummings, Positive attitudinal shifts with the Physics by Inquiry curriculum across multiple implementations, Phys. Rev. ST Phys. Educ. Res. **8**, 010102 (2012).

[531] M. Milner-Bolotin, T. Antimirova, A. Noack, and A. Petrov, Attitudes about science and conceptual physics learning in university introductory physics courses, Phys. Rev. ST Phys. Educ. Res. **7**, 020107 (2011).

[532] J. D. H. Gaffney, A. L. Housley Gaffney, and R. J. Beichner, Do they see it coming? Using expectancy violation to gauge the success of pedagogical reforms, Phys. Rev. ST Phys. Educ. Res. **6**, 010102 (2010).

[533] M. Bodin and M. Winberg, Role of beliefs and emotions in numerical problem solving in university physics education, Phys. Rev. ST Phys. Educ. Res. **8**, 010108 (2012).

[534] L. Lising and A. Elby, The impact of epistemology on learning: A case study from introductory physics, Am. J. Phys. **73**, 372 (2005).

[535] B. T. Spike and N. D. Finkelstein, Examining the beliefs and practice of teaching assistants: Two case studies, AIP Conf. Proc. **1289**, 309 (2010).

[536] S.-Y. Lin, C. Henderson, W. Mamudi, C. Singh, and E. Yerushalmi, Teaching assistants' beliefs regarding example solutions in introductory physics, Phys. Rev. ST Phys. Educ. Res. **9**, 010120 (2013).

[537] G. DeBeck, S. Settelmeyer, S. Li, and D. Demaree, TA beliefs in a SCALE-UP style classroom, AIP Conf. Proc. **1289**, 121 (2010).

[538] E. A. West, C. A. Paul, C. Webb, and W. H. Potter, Variation of instructor-student interactions in an introductory interactive physics course, Phys. Rev. ST Phys. Educ. Res. **9**, 010109 (2013).

[539] M. J. Volkmann and M. Zgagacz, Learning to teach physics through inquiry: The lived experience of a graduate teaching assistant, J. Res. Sci. Teach. **41**, 584 (2004).