The necessarily, wonderfully unsettled state of methodology in PER—a reflection

David Hammer, david.hammer@tufts.edu, Tufts University, USA

To appear in Taşar, F. M. & Heron, P., Eds,
International Handbook of Physics Education Research

Abstract

The field of Physics Education Research as described in this Handbook began about 50 years ago. Many of us initially trained and became enculturated in the much older discipline of physics, and some of our expectations and esthetics for rigor of thought and empirical research might not apply: PER is a much newer pursuit, and the phenomena of learning and instruction are highly complex and difficult to conceptualize. In this chapter, I reflect on the first studies, in the 1970's, which combined qualitative and quantitative methodologies, and on the much greater diversity of work taking place today. My purpose is to suggest that methodology in PER remains significantly unsettled, like and coupled with our theoretical frameworks. And I suggest we embrace it as a wonderful privilege, to be involved as we are during the field's formation.

This chapter is more a reflection than a review. Other chapters in this volume (REFS) present more systematic coverage of methodologies. My purpose is to consider the state of methodology in PER in the broadest sense to suggest it remains unsettled, in contrast to what many of us might expect from our backgrounds in physics.

In keeping with this volume, I will focus on PER. It is important to recognize, however, that research on learning in physics arose from and lives within research on learning and instruction much more broadly (National Research Council, 2000), which itself sits within still larger pursuits to understand mind and minds. What parts of that larger work are relevant depends significantly on what researchers see as "physics": a body of information, a set of reasoning skills, an evolving set of practices, a pursuit to understand. Each raises connections to other work, from research on learning in mathematics, history, and other disciplines (National Research Council, 2012) to research on how people develop motor skills or retain arbitrary information.

I say that as preface to emphasize that PER is part of and connects deeply to other areas. Many in PER see and promote it as a subfield of physics, with both substantive and political reasons. With respect to empirical methodology, however, my focus here, that positioning may be misleading. PER originated from and remains in rich interaction with other efforts to study human thinking and experience.

The first and most basic contribution of PER has been to show it is possible: Phenomena of learning and teaching can be studied through evidence and reasoning. That wasn't obvious, and it strains longstanding structures and expectations in higher education: Institutional policies and practices treat teaching and research as separate categories.[1] Physics education is a recent entrant into the set of areas humanity has begun learning how to study through deliberate, disciplined investigation.

Physicists have some sense of how that entry can happen. The tradition in Western curricula has been that physics began as an empirical science around 1600 in Europe with Galileo, but that was a re-introduction. There is clear historical evidence it started at least 600 years earlier in the Middle East, with al-Ḥasan ibn al-Haytham (known as "Alhazen"):

> In a critical treatise, *Aporias Against Ptolemy*, [Ibn al-Haytham] asserts that "Truth is sought for itself"—but "the truths," he warns, "are immersed in uncertainties" and the scientific authorities (such as Ptolemy, whom he greatly respected) are "not immune from error…." Nor, he said, is human nature itself: "Therefore, the seeker after the truth is not one who studies the writings of the ancients and, following his natural disposition, puts his trust in them, but rather the one who suspects his faith in them and questions what he gathers from them, the one who submits to argument and demonstration, and not to the sayings of a human being whose nature is fraught with all kinds of imperfection and deficiency. Thus the duty of the man who investigates the writings of scientists, if learning the truth is his goal, is to make himself an enemy of all that he reads, and, applying his mind to the core and margins of its content, attack it from every side. He should also suspect himself as he performs his critical examination of it, so that he may avoid falling into either prejudice or leniency." (Sabra, 2003).

This is the earliest expression I have seen of core values we aspire to uphold in science. I take it as a starting point here: Methodology is born out of the expectation that anyone can be wrong or ignorant, including ourselves. It takes discipline of mind to "submit to argument and demonstration," rather than to accept authority or what seems obvious. The emergence of PER, like physics earlier, is in part an emergence of disciplined humility.

That is, physics education is yet another area in which we cannot trust entirely in our traditions or ourselves. What, then, can we do to learn? How can we come to know? How can we check what we believe; how can we be sure we're not deceiving ourselves, or missing something? We devise methods for forming and assessing and refining knowledge. The added layer to consider is that methodology is also knowledge that reflects our imperfections and deficiencies. If we are "true seekers," we cannot simply study "the sayings of a human being" and put our trust in them, to know how to learn.

In this moment, of course, I am drawing on the writings of an ancient, ibn al-Haytham (by way of Abdelhamid Sabra), and I should pause to be "the enemy" of what I read. It is, for one, sexist—"the duty of the man" and all—something I do not want to perpetuate. Last semester (Spring 2022) I worked out a gender-neutral edit of the quote to include in an assignment for my course (General Physics I). But my teaching assistants convinced me that editing would do a different harm, of disguising a history that remains with us, and the quote went in as above.

That history, and helpful feedback from a colleague on my first draft, has me wonder also about the emphasis on "imperfection and deficiency." The message of human fallibility is important for *me* and others like me who are so often at risk of arrogance. There are more others who are often at risk of the opposite, living in a world

[1]Henderson & Dancy (2007, 2011) found that "[m]ost faculty work in institutions where structures have been set up to work well with traditional instruction." (2011, p. 7); Corbo *et al* (2016) argued that "the changes required for the systemic use of research-based teaching practices […] challenge existing norms and structures." My contention is similar, but with respect to the *conduct* of research on learning and instruction: It too challenges institutional norms and structures.

that has not expected they have something to contribute (Barthelemy et al., 2022; Prescod-Weinstein, 2020). Re-reading the text, I can see it as empowering: True seekers question "the writings of the ancients." Or: "In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual" (Galileo, 1632). Argument and demonstration can support ideas, including those that disrupt currently dominant views.

A reminder of methodology in physics

Part of the challenge for many of us in PER is that we are involved in it during its formation, typically after having been trained and enculturated in the much older discipline of physics. There we have had the privilege of centuries of progress, which like other privileges can be transparent. It will help to review, even briefly, some history of methodology in physics.

Time and again, there have been "revolutions" of mind (Kuhn, 1970) away from "truths" so obvious, or settled by authority, that no one was thinking to question them. Ibn al-Haytham led one, overturning settled ideas about light; Galileo another, about absolute rest and motion; Meitner another, about immutable elements. The revolutions often entailed showing some foundational aspect of thought to be an assumption—that objects have to be in contact to interact, that space and time are independent, that "elements of reality" (Einstein, Podolsky, & Rosen, 1935) exist without observation, and so on and on (Holton & Brush, 2001).

Changes in how physicists conceptualize physical phenomena have come with changes in methodology. The shifts from thinking of light as a particle to light as a wave in a medium, then dropping the medium, then to its quantization as photons, all drew from and affected experimental methodology. Those are examples of changes in *ontology*, that is in the kinds of entities a model takes to exist.

On a simple, classical view of phenomena, throughout the 19th century, methods of empirical research depended on reproducibility: The same initial conditions lead to the same outcomes, within measurement errors. Quantum mechanics introduced an ontological randomness, inherent and irreducible, which required a change in expectations of reproducibility, to *distributions* of outcomes. There were changes of research methods that many physicists, including Einstein, thought meant the end of physics as a science. Chaos theory introduced an epistemological "randomness," as another challenge to expectations of reproducibility: Even assuming classical determinism, tiny differences in initial conditions can lead to radically different outcomes, in systems as simple as a dripping water faucet or a double-pendulum.

The main points here are that (1) the empirical methodologies of research are deeply entangled with the tacit or explicit theoretical models, ontologies and epistemologies of a discipline, and (2) the methodologies develop over time, sometimes in ways earlier scholars could not have imagined. Over and over, the history of physics supports Ibn al-Haytham's insistence on human fallibility as well as the value of radically new thinking: Ideas that initially seem outlandish can take over the discipline.

In most areas of physics today, there are well-established theoretical foundations coupled with methods of research that have extensive track records of productivity. Those of us raised in those foundations and methods might forget that they were hundreds of years in formation, and that can have us impatient with what can seem like sloppy, undisciplined approaches to studies of learning and instruction. But disciplined study does not mean putting our trust in tradition. Methods we know from prior studies should inform how we approach research on learning, but we should not confuse adherence to authority with "rigor." Rigor should mean the ongoing pursuit of knowledge that is defensible, upon close examination and argumentation.

Physics Education Research is new, as a recognized pursuit, and it will remain new for our lifetimes. That gives us a different privilege, that of experiencing and contributing to its creation, which means—whether we enjoy it or not—our grappling with more uncertainties and ambiguities and possibilities than we would if our research were in (most areas of) physics. In the next section, I turn to an overview of that grappling.

Methodology in PER

PER has involved an eclectic, disparate mix of methodologies, reflecting and entangled with an eclectic, disparate mix of aims, theoretical frameworks, as well as proto-theoretical assumptions that we might not notice. How we conduct our studies depends on and affects what we're after, functionally, and how we conceptualize its form. In what follows, I will sample from studies in PER, from work in the early years and from work in recent years.

The beginnings of PER

The first empirical paper of modern PER,[2] to my knowledge, was Reif, Larkin, & Brackett (1976), based on Larkin's dissertation and titled "Teaching general learning and problem-solving skills." It described two studies, one of the effects of teaching a "general learning skill" for learning "any new relation" (p. 212) and the other of the effects of teaching a "simple problem-solving strategy." The methods were a mix of qualitative and quantitative. Describing the study of general learning skills, the authors wrote,

> Our most important assessment method consisted of detailed observations of individual students. Such observations are essential to elucidate how students learn and to obtain the information needed for improving the teaching materials and the underlying models upon which they are based. (p. 214)

Describing their study of problem-solving strategies, the authors specified how they posed students problems, citing methodology of protocol analysis from research on problem solving in cognitive science (Newell & Simon, 1972). Their protocols showed

> that many students in an introductory physics course approach problems in very haphazard and ineffective ways…. Thus, even when students know all the relevant facts and principles necessary for the solution of a problem, they may be unable to solve it because they lack any systematic strategy for guiding them to apply such facts and principles. (p. 216)

The quantitative parts of their study consisted of randomly dividing students to receive "special" or "ordinary" instruction, then testing the outcomes.

At roughly the same time, Laurence Viennot was working on her dissertation (Viennot, 1977), studying "spontaneous reasoning in elementary dynamics." She developed a pencil-and-paper questionnaire and administered it to several hundred

[2] David Meltzer showed me research from earlier, "a few dozen research studies were published from 1910-1945" (Meltzer, 2015; see also Meltzer & Otero, 2015). The work I cite here, as the beginnings of modern PER, was evidently an independent reemergence.

students, mostly from Belgian and British universities. Today we would recognize her questions as valid probes of conceptual understanding, but at the time Viennot heard objections that they were “traps” for students. She argued it was essential to pose questions different from those students had become familiar with in conventional teaching. In this she was in line with widely accepted expectations of research, that to make a phenomenon apparent it is not always sufficient simply to observe. Researchers construct investigations, and that should not be different for research on learning (Viennot, 1977, p. 5).

The quantitative results—tallies of correct and incorrect answers—was the first evidence presented within PER of students not learning concepts in ways instructors assumed. Qualitative analyses of students’ written explanations showed several patterns of reasoning, multiple “notions” of force that students use, “depending on the question asked” (Viennot, 1979). Viennot drew both on cognitive psychology, mainly Piaget (1973), as well as on the history and philosophy of science, mainly Koyré (1966).

Trowbridge and McDermott (1980) followed shortly after with a study of students’ reasoning in kinematics. They began with an “individual demonstration interview,” which they described as “like the ‘clinical interview’ pioneered by Jean Piaget”: “While the questioning follows a regular format, it allows for exploring any particular aspect of the student's thinking that may be of interest. Each interview lasts from 20 to 30 min and is audiotaped, or occasionally videotaped. The dialog is transcribed and analyzed in detail.”

They conducted over 300 of these interviews, initially drawing tasks from Piaget’s research, and eventually tailoring their own speed comparison tasks that they used in pre- and post-course interviews. They identified a particular difficulty, in students’ confusion of speed with position, which enabled them to craft written questions they could include for students in course examinations.

Their analyses led them to conclude that “prior to instruction the student typically has a repertoire of procedures, vocabulary, associations, and analogies for interpreting motion in the real world,” which they described as “a set of protoconcepts which antedate understandings of the concepts of kinematics” (p. 1027). They also found, just as Viennot argued in defense of her methods, that

> The ability to solve conventional problems on examinations or to pass the usual types of "mastery" tests does not always indicate conceptual understanding. Only certain types of questions can probe for the ability to resolve concepts from one another and to apply them to real situations. (p. 1028)

Right away, in the first few years of PER, there were multiple aims for research and methodologies. Reif, Larkin, & Brackett (1976) wanted to understand how to help students learn effective skills, with an expectation that these skills are independent of domain. They saw the target phenomena of reasoning as complex enough to need detailed observations for study. They started from observations of where students began, and from those observations saw students as “haphazard,” *lacking* in skills and strategies. Their experiment was to impart skills and strategies, in the form of explicit, step-by-step instructions. Seeing these skills and strategies as domain general, their methods involved testing for outcomes in both physics and accounting.

Viennot’s (1977, 1979) aim was to reveal and study phenomena of student reasoning that she also saw as complex, which also motivated methods of close, qualitative analyses of students’ explanations for their answers. In other respects, her research was quite different: The phenomena she aimed to study concerned student conceptual reasoning about force; Reif *et al* (1976) were focused on how students learn. Viennot's analyses led her to see students as *having* multiple, mutually inconsistent notions of force they would apply depending on the question. They were not reasoning in ways physics instructors would hope, but Viennot saw substance and structure relevant for physics education, reflecting a constructivist epistemology.

Trowbridge's and McDermott's (1980) methods started like Reif *et al*'s (1976). They too saw the phenomena as needing qualitative study, and they drew methodology from cognitive psychology. Like Viennot, they focused on conceptual understanding, and like her saw students' extant understanding as important to understand. While they emphasized "student difficulties," and designated them as the focus of later instructional intervention, they attributed to students protoconcepts important for their learning. Insight into those protoconcepts supported the researchers' designing specific, written probes.

It is striking in retrospect how these early studies were a mix of qualitative and quantitative methods, with the former more about initial exploration of the phenomena, and the latter to quantify some particular aspect. Methodology was significantly informed by much more general research on learning, from the cognitive sciences in the case of Reif *et al* (1976), significantly driven by computational models of mind. In line with emphases in that body of work, Reif *et al* (1976) focused their investigations on learning and problem-solving skills. Viennot (1977) and Trowbridge & McDermott (1980) drew on Piagetian constructivism, which was not at the time influenced by computational modeling, and they focused on conceptual knowledge.

PER would go on to focus much more attention to conceptual understanding as the modal target of research, beginning mainly around phenomena of motion and force in the domain of Newtonian mechanics. These two early studies presented similar ontologies of conceptual knowledge, attributing to students multiple "notions" (Viennot, 1977) or "protoconcepts" (Trowbridge & McDermott, 1980). In this they were similar to ideas diSessa (1979, 1982) was formulating, a view he would later present as "knowledge in pieces" (diSessa, 1988).

Other studies presented a unitary ontology, attributing to students a "stable, alternative view" (Clement, 1982), or "a rich accumulation of interrelated ideas that constitute a personal system of common-sense beliefs" (Champagne, Klopfer, & Anderson, 1980) that operates as something like a "paradigm" in competition with Newtonian mechanics. This ontology, or forms of it, became more common in PER for quite some time, and reflected in most accounts of misconceptions.

One reason for the prevalence of unitary ontologies may have been—may be—their methodological tractability: Stable properties are much easier to investigate. Unlike the other early studies I have described, Champagne *et al* (1980) conducted their research by developing instruments to measure students' pre- and post-instruction conceptual knowledge, abilities for logical reasoning, and mathematical skills, which allowed statistical analyses of relationships.

They were drawing on disciplinary practices in psychology, not Piagetian clinical interviews but quantitative methods that resonate well with, and may have been influenced by, disciplinary practices in physics: "In physical science," Lord Kelvin wrote, "the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it." That stance has been widely adopted by psychologists, of quantifying qualities of mind and minds, including, significantly, through the development of instruments.

A great deal of PER has happened through the development of instruments (Ding, 2019), perhaps for this resonance between the disciplines of psychology and physics, a resonance of epistemology or esthetics. These methods support and are supported by conceptualizations of ontology, attributions of stable properties to individuals, such as of conceptions, levels of cognitive development, attitudes or beliefs. Qualitative methods, in contrast, more easily afford conceptualization of situated dynamics, short-lived states of reasoning.

Along the way, the community has come to see a divide between qualitative and quantitative (Ding & Liu, 2012; Otero & Harlow, 2009; Robertson, McKagan, & Scherr, 2018). It has also mattered that PER as a community, and as an enterprise, sits within and in close contact with the educational systems of K-12 and higher education, with its conceptualizations, values, and practices. Notions of efficient, objective assessment—such as in standardized tests—are powerfully attractive within those systems.

Some researchers developing instruments have cautioned against their use for individual attributions—to "use considerable care in applying the results of a limited probe such as our survey to a single student" (Redish, Steinberg, & Saul, 1998), noting "survey results for an individual student may be misleading," while still having considerable value as measures of the statistical ensemble. Redish *et al* (1998) delivered that caution for having recognized the need: When an instrument produces a number for a student, a researcher or educator may be inclined to attribute it as a property.

A sampling of recent work

Since those beginnings, the goals and methods of PER have only expanded in variety and complexity. The targets of study have gone beyond conceptual knowledge to include disciplinary practices of learning; attitudes, expectations, epistemologies; affect and emotions, effects of contexts and community (Docktor & Mestre, 2014). Theoretical frameworks have expanded as well, and diversified, with the dynamics growing ever more complex (Brown, 2014), drawing on and leading to new developments in methodology.

I refer readers to other chapters in this volume (REFS) for more systematic reviews of current work. Here, I mean only to call attention to the rich, fertile diversity of research taking place, from my own readings in the literature as well as from suggestions by colleagues.[3]

There are studies of student inquiry in "naturalistic" settings (if a physics course is part of nature!). Euler, Radahl, & Gregorcic (2019) focused close attention on 2.5 minutes of physics reasoning by a pair of introductory students working to make sense of binary-star dynamics, showing the students' advancing by dancing and gesturing, finding a range of ways of thinking to support physics sense-making. What Euler et al presented generalizes, not necessarily as true of all students at all times, but in identifying and characterizing possible dynamics of student inquiry.

Other studies that have focused on close examinations of particular instances include Suárez (2020) studying how a group of bilingual students drew on different language resources in thinking about electricity in circuits; Odden & Russ (2019) finding

[3] I am grateful for suggestions by Amy Robertson and by an anonymous reviewer.

affective and linguistic markers of *vexation* as initiating and sustaining inquiry; Sayre & Irving (2015) identifying markers of metacognition in students' brief, spontaneous interjections; and Kapon, Laherto, & Levrini (2018) examining student inquiry in two settings to explore tensions and connections between goals of authenticity to physics and of personal relevance to students.

There are also studies based on interviews. One example is by Moshfeghyeganeh & Hazari (2021), who set out to understand a phenomenon evident in large-scale patterns: Women are much better represented as physics students in Iran than in the United States, and the pattern obtains across other Muslim Majority (MM) countries as well. Why is that? The researchers interviewed seven faculty members in physics, all women who emigrated from MM countries, asking them to recount experiences over their lives and to reflect on expectations in their communities. Their work led to several possible hypotheses, such as that "expressions of femininity emphasize modesty" in MM countries, "rather than physical attractiveness," are a closer match to values in the physics community.

Surveys of course remain a prominent and important approach to research at larger scales (Madsen, McKagan, & Sayre, 2017). For one recent example, Deslauriers *et al* (2019) used a randomized experimental approach to compare "passive lectures with active learning." They used two instruments, one a test of conceptual understanding and the other a Likert scale survey of students' agreement or disagreement with the statement "I feel like I learned a great deal from this class." The findings showed students' feelings of learning were anticorrelated with their scores on the conceptual test.

PER began with studies that worked across qualitative and quantitative studies, and it seems like a valuable heuristic for the field: Look for evidence in multiple forms, as a community if not as individual researchers. Hypotheses from Moshfeghyeganeh & Hazari (2021) could inform the development of larger-scale surveys; aggregate correlations in Deslauriers *et al* (2019) could motivate focused small-N study of the phenomena. Little *et al* (2019) took a novel approach to studying an established idea, Dweck's (2013) construct of *mindset*, by coding interview dialogue, finding nuance they would have missed using the usual Likert surveys. John & Allie designed a series of studies specifically to connect findings across methods, starting with a multiple-choice instrument (John & Allie, 2017a), then free writing responses (John & Allie, 2017b), and finally naturalistic "micro-episodes" (John & Allie, 2019), all to study the contextuality and complexity of student reasoning around DC circuits.

In PER as in other fields, new technologies afford new methodologies; the possibility of video recording had a powerful on research in the learning sciences (Derry et al., 2010). More recent developments have scholars using eye-tracking (Ibrahim & Ding, 2021; Wu & Liu, 2021) and imaging of brain activity (Allaire-Duquette et al., 2021), connecting the evidence from these sensors to findings from other modes of research. PER has begun to draw on methods of machine learning and data science (Aiken, De Bin, Lewandowski, & Caballero, 2021; Yang et al., 2020), including to support qualitative data analysis (Çınar, Ince, Gezer, & Yılmaz, 2020; Sherin, Kersting, & Berland, 2018). These approaches may provide new ways to bridge research across the qualitative/particular and the quantitative/general. (On the last, I am uneasy that the dynamics of our communities will rush to put such tools into practical implementation—automatic grading in courses. Here I am expressing interest in their use for research.)

Looking forward

PER remains a young area of study. We have made progress, no question, adapting methods of research including interviews, observations, and surveys that will continue to shape our work. But the precise forms of these methods, and our methodologies more generally, remain significantly unsettled, in tandem with our theoretical foundations. This unsettled state might be discomforting, especially for those of who were first trained in physics, where we had the benefit of working within long established frameworks.

Of course, we will and should draw on ideas from other disciplines and areas of research. Physics and psychology have been our two leading source fields,; in recent years PER has looked to sociology (e.g. Goffman, 1974) and to critical race and gender studies (Rodriguez, Barthelemy, & McCormick, 2022; Traxler, Cid, Blue, & Barthelemy, 2016). My colleagues and I recently argued for drawing from research in ecology (Hammer, Gouvea, & Watkins, 2018).[4] The subfield of community ecology in particular has also struggled with how to handle the difficulties of complexity and idiosyncrasy, in the phenomena they study. But there may well be more—we are, after all, studying *humans*. (Perhaps we should see PER as a subfield, not of physics, but of biology?)

Wherever we get our ideas, we should take care not to treat them as dogma. It is easy and tempting to settle back on authority, on "what everyone knows," on what seems clear and obvious. In many areas of physics, there are genuinely well-established methodologies, but physics as an empirical discipline has been around for at least 1000 years. And the moral from studying physics is that even the most foundational conceptualizations can change. The empirical study of learning and teaching in physics has only been around for decades: There should be no illusion that there are permanent "gold standards" of empirical scholarship (Cartwright, 2007).

My main objective in this essay has been to argue that we work as "true seekers," remembering along the way to "suspect [ourselves] as imperfect and deficient" (Sabra, 2003, p. 54). That means both staying humble, in our own scholarship and in assessing others', and embracing the possibility of novelty. It seems to be an occupational hazard for scholars, investing deeply into a point of view, and committing to it firmly, that we (myself certainly included!) can find ourselves policing the community for adherence. If we're doing that in our scholarship, as reviewers or editors or grant panelists, or even simply in positions of influence in the community, we can end up preventing new ideas from getting explored, considered, published, or funded.

That is not to say that we should let anything go. To the contrary, we should continue to question each other and ourselves, ancient (1980's?) and recent. My colleague Leema Berland and I published a critique to challenge some common and widely accepted practices in qualitative research (Hammer & Berland, 2014). In recent years for research quite broadly, quantitative methods have been challenged for problems of p-hacking. But we should take care to construct arguments and evidence, open to the possibility of something different. The unsettled state means we can and should expect, welcome, and engage with diverse ideas, including those that depart from "truths" that seem like obvious common sense. Novel methods will need novel consideration, and PER is full of novel methods.[5]

---

[4] I am grateful to Julia Svoboda Gouvea for seeing that connection.

[5] During my work on this chapter, Robertson & Hairston (2022) published a novel method of analyzing the "whiteness" inherent in an "ordinary" interaction among students in introductory physics. It drew unscholarly, intolerant reactions, dismissing the work without genuine consideration, a striking case in

I have argued PER needs to be in an unsettled state, as a new field taking on very difficult, very complex problems. I am also suggesting we see it as a wonderful privilege, to be working in PER when the field is so new and dynamic and evolving. That things are unsettled allows for scholarly invention and imagination; there remains lots of unexplored intellectual terrain.

Teaching introductory physics, I try to convince students to embrace confusion and uncertainty. I tell them that physicists are professional learners, that being confused for a physicist is like breathing hard for a runner: It's what you're supposed to feel. It is difficult: They have grown up and still live in an educational system filled with messages that confusion and uncertainty are bad things to avoid. I was motivated to write this essay for my sense that we too, the PER community, feel systemic pressures, to present clear, simple findings, to be too sure too soon. I hope this essay might help in some small way.

Acknowledgements

First, I thank Paula Heron and Fatih Tasar for their gracious invitation to suggest a topic, and for accepting this one. I am most grateful to David Meltzer, Amy Robertson, Roger Tobin, and two anonymous reviewers for their helpful, critical feedback.

References

Aiken, J. M., De Bin, R., Lewandowski, H. J., & Caballero, M. D. (2021). Framework for evaluating statistical models in physics education research. *Physical Review Physics Education Research, 17*(2). doi:10.1103/PhysRevPhysEducRes.17.020104

Allaire-Duquette, G., Brault Foisy, L. M., Potvin, P., Riopel, M., Larose, M., & Masson, S. (2021). An fMRI study of scientists with a Ph.D. in physics confronted with naive ideas in science. *NPJ Sci Learn, 6*(1), 11. doi:10.1038/s41539-021-00091-x

Barthelemy, R. S., Swirtz, M., Garmon, S., Simmons, E. H., Reeves, K., Falk, M. L., . . . Atherton, T. J. (2022). LGBT+ physicists: Harassment, persistence, and uneven support. *Physical Review Physics Education Research, 18*(1). doi:10.1103/PhysRevPhysEducRes.18.010124

Brown, D. E. (2014). Students' Conceptions as Dynamically Emergent Structures. *Science & Education, 23*(7), 1463-1483. doi:10.1007/s11191-013-9655-9

Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties, 2*(1), 11-20.

Champagne, A. B., Klopfer, L. E., & Anderson, J. H. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics, 48*, 1074-1079.

Çınar, A., Ince, E., Gezer, M., & Yılmaz, Ö. (2020). Machine learning algorithm for grading open-ended physics questions in Turkish. *Education and Information Technologies, 25*(5), 3821-3844. doi:10.1007/s10639-020-10128-0

Clement, J. (1982). Student preconceptions in introductory mechanics. *American Journal of Physics, 50*, 66.

point. Henderson and Thoennessen (2022) responded with an editorial, arguing for the importance of "thoughtful presentation and debate of ideas in a welcoming environment" and "open discussion of sometimes controversial ideas."

Corbo, J. C., Reinholz, D. L., Dancy, M. H., Deetz, S., & Finkelstein, N. (2016). Framework for transforming departmental culture to support educational innovation. *Physical Review Physics Education Research, 12*(1), 010113.

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., . . . Sherin, B. L. (2010). Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics. *Journal of the Learning Sciences, 19*(1), 3-53. doi:10.1080/10508400903452884

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1821936116

Ding, L. (2019). Theoretical perspectives of quantitative physics education research. *Physical Review Physics Education Research, 15*(2). doi:10.1103/PhysRevPhysEducRes.15.020101

Ding, L., & Liu, X. (2012). Getting started with quantitative methods in physics education research. In C. Henderson & K. A. Harper (Eds.), *Getting Started in PER, Reviews in PER Vol. 2* (Vol. 2). College Park, MD: American Association of Physics Teachers.

diSessa, A. A. (1979). Learnable representations of knowledge. In *Cognitive Process Instruction*. Philadelphia: Franklin Institute Press.

diSessa, A. A. (1982). Unlearning Aristotelian physics: a study of knowledge-based learning. *Cognitive Science, 6*, 37-75.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Putall (Eds.), *Constructivism in the Computer Age* (pp. 49-70). Hillsdale, NJ: Erlbaum.

Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics-Physics Education Research, 10*(2). doi:10.1103/PhysRevSTPER.10.020119

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development*: Psychology press.

Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical review, 47*(10), 777.

Euler, E., Radahl, E., & Gregorcic, B. (2019). Embodiment in physics learning: A social-semiotic look. *Physical Review Physics Education Research, 15*(1). doi:10.1103/PhysRevPhysEducRes.15.010134

Goffman, E. (1974). *Frame analysis : an essay on the organization of experience*. Cambridge, Mass.: Harvard University Press.

Hammer, D., & Berland, L. K. (2014). Confusing Claims for Data: A Critique of Common Practices for Presenting Qualitative Research on Learning. *Journal of the Learning Sciences, 23*, 37-46. doi:10.1080/10508406.2013.802652

Hammer, D., Gouvea, J., & Watkins, J. (2018). Idiosyncratic cases and hopes for general validity: what education research might learn from ecology. *Infancia y Aprendizaje, 41*(4), 625-673. doi:10.1080/02103702.2018.1504887

Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research, 3*(2), 020102.

Henderson, C., & Dancy, M. H. (2011). *Increasing the impact and diffusion of STEM education innovations*. Paper presented at the Invited paper for the National Academy of Engineering, Center for the Advancement of Engineering Education

Forum, Impact and Diffusion of Transformative Engineering Education Innovations.
Henderson, C., & Thoennessen, M. (2022). Editorial: Research on Advancing Equity Is Critical for Physics. *Physical Review Physics Education Research, 18*(1), 010001. doi:10.1103/PhysRevPhysEducRes.18.010001
Holton, G. J., & Brush, S. G. (2001). *Physics, the human adventure : from Copernicus to Einstein and beyond* ([3rd ed.). New Brunswick, N.J.: Rutgers University Press.
Ibrahim, B., & Ding, L. (2021). Sequential and simultaneous synthesis problem solving: A comparison of students' gaze transitions. *Physical Review Physics Education Research*, *17*(1), 010126.
John, I., & Allie, S. (2017a). DC circuits: I. Evidence for fine grained contextual dependence. *European Journal of Physics, 38*(1). doi:10.1088/0143-0807/38/1/015701
John, I., & Allie, S. (2017b). DC circuits: II. Identification of foothold ideas in DC circuits. *European Journal of Physics, 38*(1). doi:10.1088/0143-0807/38/1/015702
John, I., & Allie, S. (2019). DC circuits: III. The complex terrain of sense-making. *European Journal of Physics, 40*(5). doi:10.1088/1361-6404/ab3315
Kapon, S., Laherto, A., & Levrini, O. (2018). Disciplinary authenticity and personal relevance in school science. *Science Education*, *102*(5), 1077-1106. doi:10.1002/sce.21458
Koyré, A. (1966). *Études galiléennes* (Vol. 15): Hermann Paris.
Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
Little, A. J., Humphrey, B., Green, A., Nair, A., & Sawtelle, V. (2019). Exploring mindset's applicability to students' experiences with challenge in transformed college physics courses. *Physical Review Physics Education Research, 15*(1). doi:10.1103/PhysRevPhysEducRes.15.010127
Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Resource Letter RBAI-1: Research-Based Assessment Instruments in Physics and Astronomy. *American Journal of Physics, 85*(4), 245-264. doi:10.1119/1.4977416
Meltzer, D. E. (2015). *Early Efforts at Assessment of Learning in Evidence-Based Physics Instruction,*. Paper presented at the 2015 Frontiers and Foundations in Physics Education Research, Bar Harbor, ME. http://physicseducation.net/talks/FFPER_2015_final.pdf
Meltzer, D. E., & Otero, V. K. (2015). A brief history of physics education in the United States. *American Journal of Physics*, *83*(5), 447-458.
Moshfeghyeganeh, S., & Hazari, Z. (2021). Effect of culture on women physicists' career choice: A comparison of Muslim majority countries and the West. *Physical Review Physics Education Research, 17*(1). doi:10.1103/PhysRevPhysEducRes.17.010114
National Research Council. (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: The National Academies Press.
National Research Council. (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. Washington, DC: The National Academies Press.
Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104): Prentice-hall Englewood Cliffs, NJ.
Odden, T. O. B., & Russ, R. S. (2019). Vexing questions that sustain sensemaking. *International Journal of Science Education, 41*(8), 1052-1070. doi:10.1080/09500693.2019.1589655

Otero, V. K., & Harlow, D. B. (2009). Getting started with quantitative methods in physics education research. In C. Henderson & K. A. Harper (Eds.), *Getting Started in PER, Reviews in PER Vol. 2* (Vol. 2). College Park, MD: American Association of Physics Teachers.

Piaget, J. (1973). *La formation de la notion de force* (Vol. 29): Presses universitaires de France.

Prescod-Weinstein, C. (2020). Making Black women scientists under white empiricism: the racialization of epistemology in physics. *Signs: Journal of Women in Culture and Society, 45*(2), 421-447.

Redish, E. F., Steinberg, R. N., & Saul, J. M. (1998). Student expectations in introductory physics. *American Journal of Physics, 66*(3), 212-224.

Reif, F., Larkin, J. H., & Brackett, G. C. (1976). Teaching general learning and problem-solving skills. *American Journal of Physics, 44*(3), 212-217.

Robertson, A. D., & Hairston, W. T. (2022). Observing whiteness in introductory physics: A case study. *Physical Review Physics Education Research, 18*(1). doi:10.1103/PhysRevPhysEducRes.18.010119

Robertson, A. D., McKagan, S. B., & Scherr, R. E. (2018). Selection, Generalization, and Theories of Cause in Physics Education Research: Connecting Paradigms and Practices. In C. Henderson & K. A. Harper (Eds.), *Getting Started in PER, Reviews in PER Vol. 2* (Vol. 2). College Park, MD: American Association of Physics Teachers.

Rodriguez, M., Barthelemy, R., & McCormick, M. (2022). Critical race and feminist standpoint theories in physics education research: A historical review and potential applications. *Physical Review Physics Education Research, 18*(1), 013101.

Sabra, A. I. (2003). Ibn al-Haytham: Brief life of an Arab mathematician. *Harvard Magazine, September-October*, 54-55.

Sayre, E. C., & Irving, P. W. (2015). Brief, embedded, spontaneous metacognitive talk indicates thinking like a physicist. *Physical Review Special Topics-Physics Education Research, 11*(2). doi:10.1103/PhysRevSTPER.11.020121

Sherin, B., Kersting, N., & Berland, M. (2018). Learning analytics in support of qualitative analysis. In: International Society of the Learning Sciences, Inc.[ISLS].

Suarez, E. (2020). "Estoy Explorando Science": Emergent bilingual students problematizing electrical phenomena through translanguaging. *Science Education, 104*(5), 791-826. doi:10.1002/sce.21588

Traxler, A. L., Cid, X. C., Blue, J., & Barthelemy, R. (2016). Enriching gender in physics education research: A binary past and a complex future. *Physical Review Physics Education Research, 12*(2), 020114.

Trowbridge, D. E., & McDermott, L. C. (1980). Investigation of student understanding of the concept of velocity in one dimension. *American Journal of Physics, 48*(3), 1020-1028.

Viennot, L. (1977). *Le Raisonnement Spontane en Dynamique Elementaire.* Universite Paris VII,

Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education, 1*(2), 205-221.

Wu, C.-J., & Liu, C.-Y. (2021). Eye-movement study of high- and low-prior-knowledge students' scientific argumentations with multiple representations. *Physical Review Physics Education Research, 17*(1). doi:10.1103/PhysRevPhysEducRes.17.010125

Yang, J., DeVore, S., Hewagallage, D., Miller, P., Ryan, Q. X., & Stewart, J. (2020). Using machine learning to identify the most at-risk students in physics classes. *Physical Review Physics Education Research, 16*(2). doi:10.1103/PhysRevPhysEducRes.16.020130