# Selection, Generalization, and Theories of Cause in Physics Education Research: Connecting Paradigms and Practices

Amy D. Robertson, Sarah B. McKagan, and Rachel E. Scherr

Department of Physics, Seattle Pacific University, Seattle, WA, 98177

## Abstract:

In this paper, we use connections between (a) interviews with physics education researchers and (b) methodological literature in education research more broadly to answer questions about research practices in physics education research (PER). In particular, we define two paradigms in PER – case-oriented PER and recurrence-oriented PER – in terms of assumptions about the social world and about what counts as rigorous or trustworthy when it comes to research accounts. Case-oriented PER is predicated on the assumptions that (1) social actions are guided by the meanings that people are making of their local environments and that (2) reality is subjectively constructed. Recurrence-oriented research, on the other hand, is predicated on the assumptions that (1) human behavior is guided by predictable relationships between variables and that (2) real phenomena are reproducible. We will draw connections between (i) these two different sets of assumptions and (ii) differences in selection practices, generalization practices, and causal-claims-making in our field.

# 1. Introduction

This paper was inspired by my[a] move from the University of Washington (UW) to Seattle Pacific University (SPU) when I finished my doctoral work in 2011. My work at the UW[1-4] identified patterns in large numbers of student responses, speaking to common forms of incorrect reasoning about topics related to the particulate nature of matter. In 2011, at SPU, researchers were (among other things) selecting and analyzing short (3-5 minute) video episodes to make claims about teacher learning in the context of energy.[5,6] I was excited about joining the SPU team not only because I liked the people there and the substance of their work, but also because I did not understand their research practices and wanted to learn from them. As in: I thought there must be something rigorous and powerful – and ultimately generalizable – about, for example, carefully analyzing two minutes of video, but I did not know how to do it myself, nor did I know what purposes doing so ultimately served. When I arrived at SPU, I was immediately invited into the *practices* of selecting and analyzing short video episodes from K-12 teacher professional development, as part of the then-funded Energy Project.[5,7-12] What I found, though, was that I needed more overt framing. In particular, I wanted to know why I was doing what I was doing and toward what end. I could not answer those questions for myself, and I did not know what to ask to get the help I needed. I found myself paralyzed and at a loss for how to authentically engage, and the methodological overviews I read (e.g., Otero and Harlow[13] and Creswell[14]) offered me steps and processes, but not the framing I needed. Hence the birth of this paper: my co-authors, both early advisors of my work, encouraged me to interview other physics education researchers who engaged in micro-analytic video analysis, asking them how and why they did what they did, and, more broadly, how they conceptualized their own research.

This paper is organized around answering the questions I ultimately learned I had, drawing on the interviews I did and their connections to methodological literature in education research. It was through answering these questions that I became able to authentically engage in the kind of analysis that SPU was conducting. Central to my journey was coming to understand that SPU's research – and other, similar work in

[a] I (Robertson, first author) refer to myself in the first person here and throughout because this paper emerged from my journey as a researcher, and because the answers to the three questions on page 3 were generated through the connections I made between literature I read and interviews I did. However, the writing of this manuscript – including the organization of the ideas and the shape the manuscript took – was deeply shaped by my conversations with and the editorial support of Rachel Scherr and Sam McKagan, both co-authors on this paper. The paper will use the pronoun "I" when speaking autobiographically about my (Robertson's) journey and "we" when discussing conclusions agreed upon as a team.

PER – was predicated on *a different set of assumptions* than the ones I had become familiar with in my graduate work at the UW. In particular, case-oriented research (the kind of research being conducted at SPU when I started there in 2011) is predicated on the assumptions that (1) social actions are guided by the meanings that people are making of their local environments and that (2) reality is subjectively constructed. Recurrence-oriented research (the kind of research I conducted as a graduate student at the UW), on the other hand, is predicated on the assumptions that (1) human behavior is guided by predictable relationships between variables and that (2) real phenomena are reproducible. These two different sets of assumptions – sets of assumptions that we will refer to as "paradigms"[15] in this paper – inspire or are tied to very different research practices. My trouble, early on, was that I was trying to engage in and understand case-oriented practices using a recurrence-oriented framing.

In the remainder of the paper, we will use connections between interviews with physics education researchers and methodological literature to answer questions about three specific research practices in case-oriented PER, and we will tie these answers to the *paradigms* we articulate above. I selected these practices because they were the ones that I most struggled to understand. Specifically:

1. **Selection:** How do researchers engaging in case-oriented PER select episodes for analysis? How do they respond to the inherent subjectivity in selection?

2. **Generalization:** How do researchers engaging in case-oriented PER generalize from single cases? What purposes do such generalizations serve?

3. **Theories of cause:** How do researchers engaging in case-oriented PER make causal claims from single cases? What purposes do such claims serve?

As we answer these questions about case-oriented PER, we will answer companion questions about recurrence-oriented PER, though with less emphasis,[b] since this is the lens from which I was trying to understand case-oriented research.

Though I have framed case- and recurrence-oriented research in terms of my personal history, I believe that this work has relevance beyond myself. In other words, I think that my characterizations of these two paradigms – and their connectedness to particular practices in PER – extend beyond the work I have done at the UW and SPU.

[b] My goal in this paper is to support people like my former self, who want to understand case-oriented research but have legitimate, substantive questions about how to make sense of various practices therein. Therefore, I do not attempt to represent case- and recurrence-oriented research with equal weight and attention, nor do I try to comprehensively capture research paradigms in PER.

I hope this paper can help us understand our field – and one another – better. To be clear, I am not saying that these are the *only* two paradigms in PER. Nor am I saying that they are mutually exclusive; researchers may take up both recurrence-oriented and case-oriented PER, sequentially or simultaneously, and often to study the same phenomenon.[c] My point is to highlight that (at least some of) the diversity that we see in the practices of selection, generalization, and causal-claims-making in our field may be *motivated by* differences in assumptions about how the social world works, and/or what counts as rigorous or trustworthy when it comes to making claims about social phenomena. In other words, I mean to make more visible that people are often doing what they are doing for good reason and to illustrate plausible links between paradigms and practices.

This work is important to me in part because it is what I needed to know in order to participate in a kind of research I wanted to learn to do, and I could not find anything that directly addressed the questions I had. I want to offer what I have learned to others who may be in the same position. It is also important to me because understanding one another – and believing the best of one another's research – is one of my own deepest personal values. There are many ways in which this paper is about the "deep story"[16] that permeates our work – *why* we make certain choices in certain moments. Of course these deep stories have multiple facets and strands; I do not mean to reduce them to a single dimension. But this dimension – that of paradigms and how they tie to practices – is central to how we think about the legitimacy of research, "[a]nd matters of legitimacy have to do with matters of publication, appointment, retention, and promotion,"[17] all of which tie to belongingness and identity.

The rest of the paper sets us up to answer our three questions: first by sharing how I generated my characterizations of paradigms and my answers to the three questions (section 2), then by briefly articulating the central assumptions of case- and recurrence-oriented PER (section 3), and then by answering questions about selection, generalization, and theories of cause in turn (sections 4-6, respectively). We close with a brief discussion in section 7.

## 2. Research methods

Our primary effort in this paper is to answer questions about selection, generalization, and causal-claims-making in PER, and to show how differences in these practices may be connected to different paradigms. We adopt Greene and Caracelli's definition of a

[c] In fact, a number of the physics education researchers we interviewed do just that. For more on this, see: https://arxiv.org/abs/1307.4135.

paradigm as a set of assumptions about "knowledge, our social world, our ability to know that world, and our reasons for knowing it" that frame and guide a particular orientation toward research, "including what questions to ask, what methods to use, what knowledge claims to strive for, and what defines high-quality work."[15] Our answers to the three questions and our characterizations of case- and recurrence-oriented research grew out of connections we made between (1) researchers' own descriptions of their work during interviews and (2) methodological literature in the social sciences, including education research. In a sense, what we say here is not new; methodologists already know that research assumptions motivate research practices. On the other hand, our work takes a careful look at what this looks like in one field – PER – including how it is expressed in researcher talk. It also pulls together multiple lines of thinking in a way we have not seen elsewhere (and in a way that I needed as a novice participant in case-oriented research). In this section, we describe our methods in detail.

*Interview sampling.* Interviews were originally meant to help me understand how researchers make sense of their work – especially work similar to that which was being done at SPU (e.g., micro-analyses/case studies using video). At the time, we thought of those interviews as helping me to understand "qualitative PER." More specifically, the first three or four interviews were conducted with close collaborators of SPU – researchers whom we knew to be doing the kind of research I wanted to understand. As I conducted these interviews and we made a collective decision to pursue our analysis of them as research more formally, we made an effort to branch our sample to (1) additional researchers doing "qualitative PER" but who were not close collaborators of SPU and (2) researchers we perceived to be doing other kinds of research – specifically, to researchers we (then) considered to be doing "quantitative" and "mixed methods" research.[d] We limited interview candidates to researchers who had been out of graduate school for at least three years, because we expected that such interviewees would have a developed research trajectory on which to reflect. We solicited participation in interviews by email and interviewed all eighteen of the physics education researchers that responded affirmatively. As we have reiterated, because an original goal of this project was to better understand what we then thought of as "qualitative PER" – and what we now call case-oriented research – researchers that we perceived to be conducting this kind of research made up a larger fraction of the interview subjects.

[d] See https://arxiv.org/abs/1307.4135 for a broader characterization of what emerged from all of these interviews; here we focus on those interviews that depicted case- and recurrence-oriented research.

The discussion in the previous paragraph was framed in terms of "qualitative," "quantitative," and "mixed methods" PER, but this manuscript uses the terms *case-oriented* and *recurrence-oriented PER.* One may wonder – and many have asked – why we have chosen the latter (instead of the former) terms to describe our work. The reason is this: we ultimately decided that what was most meaningful to us in making sense of our interviews was not as much about methods; it was more about *assumptions* – assumptions about how the social world is structured, what is "real" therein, and how we come to know those real things. "Qualitative" and "quantitative" are often used to refer to *methods*, the processes by which researchers obtain and analyze data. A paradigm, in contrast, is a set of *assumptions* (which may bear out in research aims and methods). The same method may serve different paradigms[e], and a single paradigm may use multiple methods. Because we wish to draw attention to *paradigms,* for the reasons we state in the introduction, we selected names that highlight what we consider to be central to the two paradigms we discuss: case- and recurrence-oriented PER.

*Content of interviews.* Each interview lasted between forty-five minutes and one hour and was conducted either in person, by remote video, or on the phone. The interviews were loosely scripted. Major topics included: the kinds of questions each interviewee is interested in answering, the process by which each interviewee tries to answer these questions, the kinds of claims each interviewee seeks to make, what counts as evidence for these claims, and the criteria each uses to evaluate their research. These questions were informed by my original interest in understanding *how* researchers conceptualize the work they do, broadly speaking. Each interview was recorded, content-logged, and summarized.[19] The summaries were sent to individual interviewees and revised on the basis of their feedback.

*Interpretive framework and analysis of interviews.* The claims in this paper grew out of interactions between the content of our interviews and literature on research methodologies in the social sciences, including education research. Each one – the literature and the interviews – helped us to understand the other. In writing this paper, we take specific perspectives or assumptions that are expressed within the methodological literature – e.g., that case studies are meant to broaden audience perspective,[20-22] or that the recurrence of a result across independent observers lends credibility to the truth of the result[23] – and illustrate what these assumptions look like

[e] For example, interviews may be used for recurrence- or case-oriented research. For the former, for example, one may conduct many interviews with a representative sample of university physics students for the purpose of broadly understanding how they interpret a set of questions about forces. For the latter, for example, one may analyze a single interview (or series of short interviews) in depth for the purpose of understanding how a student's framing[18] of a particular question shifts over the course of a conversation, and in response to what.

in PER, and how they are plausibly tied to particular practices within our field. Further, we *foreground* those assumptions from the methodological literature that were evidenced in our interviews; these are the assumptions that comprise the recurrence- and case-oriented paradigms we describe. In short, the literature helped us to articulate the assumptions within each paradigm, and the interviews helped us to demarcate which aspects of the methodological literature were relevant and/or applicable to research within PER, as described by our interviewees.

We identify this manuscript as case-oriented research. We show how (methodological) theory manifests in the concrete details of researchers' talk about their own work,[24] and we make a theoretical claim that research assumptions are tied to practices of selection, generalization, and causal-claims-making within one field (PER). Further, our effort as we conducted our interviews and analysis was to understand how the meaning that researchers make of what they are doing shapes the substance of their research, drawing on the assumptions of case-oriented research we articulate in Section 3.

In the remainder of the paper, we use examples of published physics or science education research to illustrate our characterizations of recurrence-oriented research, case-oriented research, and practices therein. Examples were selected because they clearly embody either the recurrence-oriented or case-oriented research paradigm or a way of combining them; appear in a journal recognized by the PER community as a primary site for publishing research; and are authored by physics education researchers whose work is recognized as shaping community standards. These papers were selected after our analysis of interviews; they were chosen to (and did) validate and extend our original claims.

Prior to submission of our manuscript, we conducted extensive member checks,[13,14,25] offering interview participants and authors of the published examples we use opportunities to provide us feedback on drafts of this paper, and we revised the manuscript on the basis of their responses. Interviewees resonated with the content of our descriptions of the paradigms, although some objected to being labeled as committed to a single paradigm, or to having their research reduced to a single label. As such, and to indicate that researchers may simultaneously or sequentially participate in different research paradigms, we focus on *research* commitments, rather than *researcher* commitments, throughout the paper. For example, a researcher may believe that (a) social action is shaped by the meaning that students are making of their local environment (assumption within the case-oriented paradigm), while also believing that (b) across many students, social actions can be understood in terms of probabilistic patterns or relationships (assumption within the recurrence-oriented paradigm). However, in our interviews, some researchers expressed strong personal commitments

to the premises, values, and assumptions reflected in our characterizations of a specific paradigm, suggesting that it is possible for researchers to primarily identify with one research paradigm at a given time.

*Limitations.* As with most case-oriented research, our claims draw on small *N* – in this case, interviews with a small number of physics education researchers – which limits their generalizability to the *population* of physics education researchers *writ large*. However, we have been clear that we are not aiming to characterize PER comprehensively, and we do not think our work does this. We mean to *illustrate* how certain assumptions (organized into paradigms) show up within PER and may plausibly link to practices that I originally found difficult to translate/understand. These are theoretical claims that may have broad applicability, but they do not require large numbers to substantiate.

Further, we acknowledge (in fact, appreciate) that our work is *one* perspective about how to characterize or describe research within PER. As above, our characterization is not comprehensive, and it is likely oversimplified – we have focused on a particular aspect of research that was depicted within our interviews.

Relatedly, the content of our interviews was shaped by my interest in understanding how researchers make sense of selection, generalization, and causal-claims-generation. When research assumptions and/or these three practices – or ways in to these topics – came up in the natural course of interviews, I focused on and followed up on this. Likewise, in my content logging and attempts to understand interviewees' points-of-view, this was a (sometimes unconscious, sometimes more explicit) focus of my interest and attention. This narrows the scope of what researchers discussed during the interviews, and thus what we can infer, in a broad sense, about their work, from these interviews.

## 3. Recurrence- and case-oriented physics education research paradigms

In this section, we briefly introduce the two research paradigms that we will focus on in the remainder of the paper: recurrence- and case-oriented PER. In particular, we articulate the assumptions about "knowledge, our social world, our ability to know that world, and our reasons for knowing it"[15] that we see comprising (at least in part) these two paradigms. We introduce recurrence-oriented research first because it is the perspective from which I was trying to understand case-oriented research. In Sections

4-6, we will tie our answers to our questions about selection, generalization, and theories of cause to these paradigms.

### 3.1 Recurrence-oriented physics education research

Recurrence-oriented research is predicated on the assumptions that (1) human behavior is guided by predictable relationships between variables and that (2) real phenomena are reproducible. Quotes from interviews with physics education researchers and published research by Pollock and Finkelstein[26] inform and illustrate our characterization of recurrence-oriented PER. In "Sustaining educational reforms in introductory physics," Pollock and Finkelstein build on an earlier study[27] exploring inter-institutional hand-off of *Tutorials in Introductory Physics*,[28] a PER-based curriculum. In that earlier study, Pollock and Finkelstein showed that the *Tutorials* could be successfully implemented in a context other than the one in which it was originally developed. In the later study that we reference here,[26] the authors explore intra-institutional hand-off of *Tutorials* – i.e., hand-off within the University of Colorado at Boulder (CU-Boulder). They demonstrate both that (1) student learning gains (as measured by $\langle g \rangle$, average normalized gain[29]) are consistently large when *Tutorials* are used in recitation sections at CU-Boulder, and that (2) there is some variation in $\langle g \rangle$ that may be attributable to curricular choice and/or faculty familiarity with PER.

#### 3.1.1 Human behavior is guided by predictable relationships between variables.

Recurrence-oriented research often models human behavior as governed by predictable relationships between variables. Such relationships are understood to be most accurately reflected at the level of populations, which consist of all members of the group of interest. Since obtaining data from an entire population is usually not possible, relationships are apprehended using probabilistic and statistical tools[24,30] that allow researchers to evaluate the possibility of making incorrect population-level inferences on the basis of data collected from a sample.[31-33] This kind of research is modeled after that in the natural sciences, in which nature's uniformity allows a mechanical, chemical, or biological understanding of causation.[21,24,32,34,35] This uniformity does not imply linearity; rather, it implies that variation in human behavior follows a trend.[36,37] Thus, any non-uniformities will likely average out to zero if one considers an entire population (or a representative sample of that population).

In recurrence-oriented PER, both the human and physical worlds are modeled as governed by lawful relationships, with the caveats that (1) there are many more variables to consider in social interactions than in physical ones and (2) the interactions

between these variables are much more complex. One researcher described his research as follows:

> "I'm trying to figure out the underlying dynamics [of physics learning and teaching], and yes, I think the basic approach is similar to the standard physics research approach, which is to try to understand the various factors involved in the system, to try to control a reasonable number of them and to vary certain others to look at the outcomes, with an aim to understanding the underlying dynamics."

Pollock and Finkelstein[26] likewise search for a relationship between variables to explain semester-by-semester differences in introductory physics students' average normalized gain. For example, they highlight the effect of the variable 'instructor' on $\langle g \rangle$, binning instructors according to their familiarity with PER. The authors relate "faculty background" to student learning, writing,

> "…we observe that in instances when PER faculty are involved in instruction, in either the lead or secondary role, students post the highest learning gains. When PER-informed faculty…are involved in instruction, students post higher learning gains than when only traditional faculty are involved."

### 3.1.2 Real phenomena are reproducible.

The probabilistic and statistical tools used in recurrence-oriented research embed specific assumptions about what it means for a claim to be true. In particular, recurrence-oriented research represents human behavior in terms of observable phenomena and predictable relationships that exist "independent of [the scientists'] personal values and sociopolitical beliefs."[38] To ensure that observations and inferences truly reflect these phenomena and relationships – and not biases that result from an unrepresentative sample or from the personal values of the researcher – researchers conducting recurrence-oriented PER privilege phenomena that recur over and over, independent of observer and context.[35] Cook[23] ties *recurrence* to *truth*, saying that even though "observations are never theory-neutral, many of them have stubbornly re-occurred whatever the researcher's predilections" and thus have "such a high degree of facticity that they can be confidently treated as though they were true." One of our interviewees stated:

> "In general, I say, 'That's a very interesting result. Now do it again and see what happens.' And if you get it a third time and if it's similar to what you observed the first two times, then you can begin to believe that you're onto something. But if you do it a second and a third time and it is very different than what happened the first time,…then you have to be very skeptical and say that there's a good chance that this was just a random fluctuation type of thing."

One of the central questions of Pollock and Finkelstein's (2008) study concerns the reproducibility of large conceptual gains when *Tutorials* are implemented across and within institutions. The first two graphs in the paper show the statistical

indistinguishability of results from (1) implementations of *Tutorials* at (a) the University of Washington (*Tutorials* development site) and (b) CU-Boulder (*Tutorials* implementation site) and (2) the first and second implementations of *Tutorials* at CU-Boulder. These graphs communicate that the gains achieved both at UW and during the first implementation at CU-Boulder do not represent random fluctuations or irreproducible, extenuating circumstances; they represent a *real* curricular effect.

### 3.2. Case-oriented physics education research

Premises of case-oriented research include that (1) social actions are guided by the meanings that people are making of their local environments and that (2) reality is subjectively constructed. Quotes from interviews with physics education researchers and Berland and Hammer's case study, "Framing for Scientific Argumentation,"[39] inform and illustrate our characterizations of case-oriented PER. Berland and Hammer's manuscript presents three episodes from a sixth-grade science class. The authors focus on the social and epistemic expectations constructed by the students and teacher. In the first episode, the "idea sharing discussion," Mr. S (the teacher) nominates students to contribute ideas, often acknowledging or revoicing these ideas without evaluating them. In the second "argumentative" discussion, which takes place later in the unit, students engage with one another's ideas and try "to persuade each other to accept their claims." The "discordant" discussion takes place immediately after the "argumentative" one, when Mr. S "move[s] to resume his role as epistemic and social authority." Unlike in the first two episodes, in the third, "instabilities" emerge from the competing expectations of students and teacher: some participants' expectations were more consistent with the framing of the idea-sharing discussion, whereas others' expectations were more consistent with the framing of the argumentative discussion. The authors connect student and teacher framings in these three episodes to the literature on argumentation and suggest that certain frames are better aligned with productive argumentation practices.

#### 3.2.1 Social actions are guided by locally constructed meanings.

Case-oriented PER is guided by the assumption that people construct locally meaningful interpretations of their environments;[14,24,30,32,40] that people take action on the basis of these interpretations (i.e., these interpretations are causal);[24,41] and that culture can organize interactions and promote shared meanings among groups of people that regularly interact.[24,42] The meanings that participants make of their experiences are dynamic and exist only in the context of local interactions, evolving as they continually (1) make sense of (and shape) their contexts and (2) respond to other participants who are simultaneously making sense of (and shaping) the context.[34,43,44] Case-oriented

PER understands social phenomena in terms of "what people mean and intend by what they say and do" and how these meanings are connected to and shaped by "historical, cultural, institutional, and immediate situational contexts."[38] Researchers choose events for analysis that highlight the social mediation of meaning and/or that reveal local patterns that organize interaction. For example, one interviewee described her ongoing analysis of interviews with middle school students:

> "There were many students who throughout the course of the interview would sometimes use chemistry vocabulary words. And then at other times, they would switch, where they would start drawing on their everyday experiences…And what I have been thinking about that is that that is an epistemological issue, so that actually, students aren't quite sure how to engage in these interviews that we use so regularly as researchers. And that they're sort of trying on a number of different ways that they could engage in the interview."

In this quote, the researcher attributes students' participation and knowledge-on-display to the meaning that they are making of their local context.

Berland and Hammer's[39] analysis also embeds this orientation toward social action. The authors document take-ups and dismissals of ideas that are linked to the meaning students and teacher are making of their shared activity. For example, in the "argumentative" discussion, students "frequently addressed one another directly and responded to each other's arguments," rather than directing their contributions to their teacher, Mr. S. The authors infer stable epistemic and social expectations throughout the discussion: students expect to assess ideas on the basis of evidence and reasoning and to hold ideas in opposition to one another (so that one idea prevails). They also expect to select ideas for further discussion and to control turn-taking themselves. Thus, when Mr. S intervenes to quiet the influx of student contributions, they ignore him, and he acquiesces. His bid for a shift in social expectations – toward himself as moderator of the discussion – is not taken up *because* it is inconsistent with the meanings the students are making of the discussion.

### 3.2.2 Reality is subjectively constructed.

The assumption (discussed immediately above) that social action is guided by meanings that are locally constructed and dynamically evolving is tied to the assumption that reality is subjectively constructed – that these locally constructed meanings are themselves what is *real* to participants, and thus what matters for a research account. Thus, case-oriented research tends to *trust* accounts that attend to the details of context and highlight multiple layers of meaning. Researchers conducting case-oriented research tend to foreground interactional complexity:

> "There was just so much going on in [*Tutorials*] that I had been unaware of. My former thing about pre- and post-testing was just missing so much of the amazing stuff that was *really* happening in the tutorial…Ever since, my research has been organized around trying to appreciate the complexity of what is happening in a fascinating classroom."

Berland and Hammer's study highlights the importance of participants' social and epistemic expectations to the conversational dynamics and to the productivity of students' argumentation practices. In the authors' analysis, the relevant context in which the discussion takes place is the *meanings* that students and teachers are making of "what is going on" in their shared space. These meanings are participants' *real* experience of the context and thus what guide the unfolding dynamic of the discussion. And these meanings are complex – tied to social and epistemic expectations and to verbal and nonverbal signals by which participants communicate their expectations to one another.

## 4. Selection practices: How do researchers engaging in case-oriented PER select episodes for analysis? How do they respond to the inherent subjectivity in selection?

In this section and the next two, we will give our answers to questions about specific research practices in case-oriented PER, and we will show how these practices draw on the assumptions articulated in Section 3. As we go, we will answer parallel questions for recurrence-oriented PER, illustrating the perspective from which I was coming from in trying to participate in and engage with case-oriented PER.

We will start by answering the question, "How do researchers engaging in case-oriented PER select episodes for analysis?" Our broad-brush answer is that they select episodes by *locating cases of theories in context*. For example, when looking at classroom video, a researcher conducting case-oriented PER may ask themselves, "What is this phenomenon a case of?" (rather than, "What are the important patterns or relationships that recur?" as a researcher conducting recurrence-oriented PER might ask). We will show that this difference is entangled with the different premises and purposes of case- and recurrence-oriented research: recurrence-oriented research that makes *population-level* claims requires representative or reproducible data, whereas case-oriented research that makes *theoretical* claims requires cases of relevant theory.

Further, episode selection in case- and recurrence-oriented PER is guided by assumptions about what is real and/or true. In particular, recurrence-oriented research is predicated on the assumption that real phenomena are reproducible, so researchers taking up this paradigm foreground events that recur and/or meet other statistical

standards. Whereas case-oriented research is predicated on the assumption that reality is subjectively constructed, so researchers taking up this paradigm foreground events that highlight the complexity of social interaction and/or the social mediation of meaning.

Answering this question helped me figure out how to *frame* the selection of episodes for analysis in case-oriented PER. My early experience around questions of selection was that, when asked how they select episodes for analysis, researchers conducting case-oriented PER often said that they pick something that is "interesting" or "striking" from their data. This did not make sense to me (and may feel unscientific to others), as a researcher more familiar with the recurrence-oriented PER paradigm. For example, one researcher described the beginning of the case-oriented research process as follows:

> "For me, the process starts when somebody with good professional vision sees something that wows them. The next step is to begin to identify what it is about that that is so impressive. Because I think that a lot of the time as teachers and as researchers, we have sort of a gut reaction, like, 'Look at that! What is that? That is so amazing!' You know, it just gives you chills …"

I now see that this researcher's sense that an episode is "amazing" is structured in ways that are not immediately obvious – that in fact selection involves extensive (sometimes implicit) theoretical knowledge, and that there are parallels between the ways in which phenomena are selected for analysis in case- and recurrence-oriented PER. I hope that our answers in this section make these practices – and the knowledge and assumptions that inform them – more visible.

We also answer the question, "How do researchers engaging in case-oriented PER respond to the inherent subjectivity in selection?" For me, this was more a question of legitimacy than of process or framing. It mattered to me as someone who wanted to both *do* case-oriented PER and *share* my work with others who, like my former self, might be less familiar with this kind of research. Our answer to this question is, again, tied to the premises and assumptions of the paradigm and appears in Section 4.3.

**4.1 Researchers foreground complex participant meanings.**

Here we will start to answer the question, "How do researchers engaging in case-oriented PER select episodes for analysis?" In particular, we will discuss what gets foregrounded in case- and recurrence-oriented research, and how this foregrounding is tied to the paradigmatic assumptions articulated in Section 3.

As we say earlier, case-oriented research is predicated on the assumptions that (1) social action is shaped by locally constructed meanings and that (2) reality is subjectively constructed. Thus, researchers participating in case-oriented PER often

foreground (or tend to select) events that highlight the social mediation of meaning and embed nuance and complexity.

For example, Berland and Hammer[39] are explicit about highlighting the social mediation of meaning: they regularly refer to the verbal and non-verbal messages that participants communicate to one another as they reinforce or bid to shift their shared (or unshared) frames. The "discordant" discussion especially stuck out to Berland and Hammer *because of* its instability and the unsharedness of participant framings. Specifically, Mr. S bid to enforce his epistemic and social authority, shifting away from the stable frame in which students were distributing that authority among themselves. Although some students take up his bid, others resist, and this tension between frames manifested as classroom discord that was evident to the authors. Another researcher described her attention to complex classroom events:

> "[A]ll of my research [is] organized around trying to appreciate the complexity of what is happening in a fascinating classroom. To really try to put my attention on the incredibly cool stuff that could get missed because it's not part of the assessment and/or the instructor isn't there for it."

Although we have separated researchers' attention to participant meaning and their attention to complexity, we suspect that the two are actually entangled in the selection process – that researchers likely choose instances of social mediation of meaning that are also complex.

This is in contrast with recurrence-oriented research, which is predicated on the assumptions that (1) human behavior is guided by lawful (albeit probabilistic) relationships and that (2) real phenomena are reproducible. Whereas selection in case-oriented research flows out of naturalistic observation of complex social phenomena, selection in recurrence-oriented research is often embedded in experimental design: researchers select variables to test and choose statistical tools or make repeated measurements to ensure representativeness and reproducibility. For example, Pollock and Finkelstein[26] express their interest in understanding whether the positive effects of PER-based curricula (especially *Tutorials in Introductory Physics*) can be successfully reproduced (a) across sites (from development site to new institution) and (b) across faculty within a single institution. Because they want to know if the results are reproducible, they (a) make the same measurements as did the curriculum developers and (b) use the same standard conceptual instruments within their institution.

These recurrence-oriented research premises may also direct researchers' attention toward "clean" data, as opposed to data that is messy or complex. We interpret "clean" to describe experimental data that carefully controls for all but the variables of interest and that meets standards of reproducibility or representativeness, either by repeating

the experiment/measurement or by ensuring the sample is random or representative. For example, one researcher said:

> "What I liked about that paper was that it was a very clean study…[I]t was a random handing out of papers. And yet we found a statistically significant difference between how students answered those two questions based on which version they got. The value that I think they have is that they are clean tests of hypothetical models for what's happening with students."

These preferences may help explain why events that are particularly interesting (e.g., complex/messy) to a researcher engaging in case-oriented PER may be dissatisfying or uninteresting (e.g., disorderly) to a researcher engaging in recurrence-oriented PER.

### 4.2 Researchers select cases of theory.

In the previous sub-section, we started to answer the question, "How do researchers engaging in case-oriented PER select episodes for analysis?" by describing what researchers foreground. Here, we will talk more about process – the *mechanics* of selection practices in case-oriented PER.

Case-oriented PER takes as a premise that the universal properties of teaching and learning are revealed in the details of specific cases. Researchers refine, extend, and refute theories by connecting theory to specific cases, identifying what the case under study is a *case of*.[24,45,46] Researchers conducting case-oriented PER bring their theoretical knowledge to bear (consciously and unconsciously) as they collect and analyze data: they look for (or tend to see) cases of theory. Wylie[47] (quoted in Freeman et al.[48]) calls this "ladening data with theory." One selection criterion is the extent to which the case is likely to contribute to the development or refinement of theory.[49] Often the articulation of which theory or theories a particular case instantiates is not automatic; the researcher has unconsciously used a certain theoretical lens in their selection, and part of the research process is articulating what that lens is. For example, one researcher we interviewed said:

> "I feel like for me, the process starts when somebody with pretty good researcher or teacher eyes sees something that wows them [and then goes on to] look for other things in the video that maybe seem similar to you, so that you can maybe start to say, "Listen, I don't know what it is, but I feel like these things all go together. I think these are all about the same thing. *What thing are these all about?"* To help you articulate your sense of what matters about the episode."

By "what matters," this researcher agreed that she meant the episode's theoretical significance, or what theory the episode is a *case of.*

Berland and Hammer's[39] manuscript illustrates the foregrounding of theoretical cases. The authors selected the "argumentative" discussion because it instantiates productive argumentation practices, and the "discordant" discussion stuck out to them because of its relative instability with respect to frames that embed particular social and epistemic expectations. They deliberately searched for a third episode that was more representative with respect to frame stability and argumentation, choosing the "idea sharing" discussion. The authors write:

> "This study grew out of our shared interest in the contrasting dynamics between the argumentative discussion and the subsequent discord: How could we account for the stability of the former and the instability of the latter [with respect to framing]?...We realized [through our analysis…] that we were looking at two idiosyncratic episodes from this class: Although the students and teacher gave the sense that they knew what they were doing in it, the argumentative discussion was unlike anything the first author had observed in this class…, and the discordant episode was unusual in its discord. To get a sense of how things went 'normally,' we examined earlier class discussions and picked two seemingly typical sessions to study through the theoretical lens of framing."

In short, what was 'striking' or 'interesting' to these researchers – what stuck out to them when they watched classroom video – were *cases of* argumentation and framing.

This contrasts with recurrence-oriented research, which prioritizes recurrence and reproducibility. Thus, researchers look for recurrent patterns and relationships in their data. For example, one researcher said:

> "We are trying to look at these patterns of results and see whether little variations that seem to be out of the norm in various ways are representative of a significant phenomena [sic] that has important ramifications – or has absolutely no relationship to anything significant, and if you do 'em again 99 times, it'll never happen again."

Pollock and Finkelstein[26] notice the "sizable variation of success" in *Tutorials* implementation at CU-Boulder. What was salient to the authors about this variation was its association with the variable of 'course instructor': they observed that students of faculty more familiar with PER outperform students of less familiar faculty. Discussing a table that presents normalized gains on the Force and Motion Conceptual Evaluation[50] for ten different courses, they write, "As seen in Table I, the most significant variations among semesters are associated with the backgrounds of the instructors." Throughout their manuscript, the authors refer to "familiarity/background" as a variable that relates to student performance.

To be clear, though we think the selection criteria we have highlighted thus far in Section 4 are primary within case- and recurrence-oriented PER, we do not mean to imply that these are the *only* criteria that influence selection. The researchers we interviewed mentioned additional criteria that we will not discuss in detail here (e.g.,

one researcher said that she not only foregrounds prevalence but also how "foundational" an idea is in deciding what patterns to report and respond to in her recurrence-oriented research). In fact, though in some cases tacit, we suspect that recurrence plays a role in case-oriented selection and theory plays a role in recurrence-oriented selection. That is, a researcher conducting case-oriented research may be more likely to foreground or select a case for further study if it represents something they have seen a number of times before. Similarly, a researcher conducting recurrence-oriented research may be more likely to notice those patterns that are in some way theoretically significant, drawing on their (explicit or implicit) theoretical understandings. In both cases, respectively, it may not be recurrence or theory-instantiation that is *convincing* or *primary,* but these may play a role in the selection.

### 4.3 Researchers respond to subjectivity by making bias visible.

In this sub-section, we begin to answer our second question about selection, "How do researchers engaging in case-oriented PER respond to the inherent subjectivity in selection?" This question is less about *how* episode selection takes place in case-oriented PER or *what* influences researchers' vision. Rather, it addresses concerns that episode selection by its very nature – selecting cases to illustrate or refine theory – is subjective. This concern may be tied both to a view (common in science culture[51,52]) of subjectivity as something to be guarded against, and also to associated standards and practices for minimizing the subjectivity of research results, which case-oriented research often does not meet. In case-oriented research, however, subjectivity is understood to be inevitable, and associated standards and practices focus on explicit acknowledgment of various sources of bias or influence. In what follows, we argue that the difference between selection practices in case-oriented and recurrence-oriented PER is not in the degree of subjectivity but in the researcher's response to it.

Both selection and invention occur at multiple stages of the case-oriented and recurrence-oriented research processes described by our interviewees. For example, when engaging in case-oriented research that seeks to construct narratives of particular classroom events, researchers make selections as they: choose when and where to record video (which entails selecting relevant populations or content); capture video (which involves pointing the camera in a particular direction[53]); select an episode (which involves choosing a portion of the video corpus to analyze in detail); and formulate claims (which involves highlighting particular parts of the selected episode as evidence). Invention happens in this kind of research as researchers build connections between case and theory (in order to articulate and refine what a particular episode is a *case of*), as well as when they categorize and interpret observations to formulate claims. When engaging in recurrence-oriented research that, for example,

aims to characterize student conceptual understanding and develop research-based curriculum, researchers make selections as they: choose relevant content on the basis of which to write a question (which involves selecting a topic and, in some cases, deciding what are the learning goals for that topic); formulate conceptual questions (which entails choosing some feature of the selected content on which to focus); administer these conceptual questions (which involves selecting relevant populations and settings); formulate interpretive categories (which involves selective attention to the data); and categorize student responses. Invention takes place when, for example, researchers engaging in such recurrence-oriented PER design hypothetical scenarios for conceptual tasks and when they formulate interpretive categories (and in doing so, discover meaning in student responses).

The literature on research methodologies responds to the prevalence and necessity of invention and creativity by addressing the role of the researcher, arguing that it is not their task to ensure that another person would discover the same meaning in the same data.[35] Rather, it is the researcher's task to make explicit why they made the choices they did and why those choices were reasonable ones to make. Marton,[54] discussing phenomenographic research, says:

> "… someone usually asks: Would another researcher working independently arrive at the same set of categories if he or she were studying the same data? On the surface, this appears to be a reasonable question. After all, research results are supposed to be replicable. However, two issues are buried in the question. One concerns the process of discovery: Would other researchers find the same conceptions or categories if they were doing the study for the first time? (Analogously, we might ask, Would two botanists discover the same plants and species if they independently explored the same island?) The other issue concerns whether a conception or category can be found or recognized by others once it has been described to them by the original researcher. The point I want to make is that replicability in the second sense is reasonable to expect, but in the first sense it is not. The original finding of the categories of description is a form of discovery, and discoveries do not have to be replicable. On the other hand, once the categories have been found, it must be possible to reach a high degree of intersubjective agreement concerning their presence or absence if other researchers are to be able to use them. Structurally, the distinction I draw here is similar to that between inventing an experiment and carrying it out. Nobody would require different researchers independently to invent the same experiment. Once it has been invented, however, it should be carried out with similar results even in different places by different researchers."

We suspect that most physics education researchers would acknowledge that their selection and analysis practices involve the kind of subjectivity we describe above. However, the two different paradigms call for very different responses to this research reality.[38] In case-oriented research, the assumption is that it is impossible to eliminate the influence of the researcher on the research, and so this paradigm calls on researchers to seek to articulate the theories and perspectives that affect (or bias) their selection and analysis:[48,55]

> "The researcher brings ways of thinking about learning, about physics, about learning physics, and about how people do (or should) behave and think. A good qualitative researcher acknowledges the many subjectivities or tacit theories that, even though he may not be aware of, guide his actions (where to point the video camera, what he notices going on in class, what type of data he decides to collect)."[13]

In recurrence-oriented research, in contrast, there is an effort to eliminate "selection bias"[40] by privileging those phenomena that recur over and over, independent of observer and context:

> "Even if observations are never theory-neutral, many of them have stubbornly re-occurred whatever the researcher's predilections…So, even if there are no 'facts' we can independently know to be certain, there are still many propositions with such a high-degree of facticity that they can be confidently treated as though they were true."[23]

These very different approaches entail (and are entailed by) quite different perspectives about what constitutes rigorous work:

> "If you believe, for instance, that good scientists should be objective in the sense of producing knowledge that is epistemologically independent of their personal values and sociopolitical beliefs, then you are likely to privilege as rigorous those methods that demonstrate agreement (replication or reproducibility) among independent observers. If, instead, you believe knowledge is unavoidably shaped by the preconceptions of the knowers (and that independent agreement may simply be a manifestation of a shared bias among the members of a research community), then you are likely to privilege as rigorous those methods that illuminate the nature of the bias and the social, cultural, and political factors that shaped it."[38]

These differences in perspective are connected to the different assumptions about social phenomena/human behavior taken within case-oriented and recurrence-oriented research. In particular, in case-oriented research, the important causal influences on human action and behavior are the context-sensitive meanings made by participants, so it makes sense that one would respond to the subjective influence of the researcher by seeking to make visible the "historical, cultural, institutional, and immediate situational" [38] influences on the meanings one is making of one's own research. In contrast, in recurrence-oriented research, the assumption is that there are predictable, causal relationships that govern human behavior, which suggests that these relationships can exist completely apart from our understanding or study of them. In this case, probability and statistics or repeated observations are seen as providing a means to assess or minimize the likelihood that the phenomenon has been mischaracterized or misunderstood.[17,32]

Concerns about the subjectivity of case-oriented research are likely tied not only to bias in the selection of cases but also to questions of generalizability – concerns that a single case is insufficient to make generalizable claims, or that claims made on the basis of a

single case are naturally subjective. These concerns may be grounded in the perspective that "generalities" are those patterns and relationships that recur across cases and that research results are trustworthy (i.e., represent real results that can inform predictions) when they recur over and over. From such a perspective, the analysis of a single case may be seen as subjective in the sense of generating results that may be biased or random. In Section 5, we will argue that case-oriented research takes a different perspective on generalizability and on what is trustworthy and useful, one that is intimately linked to the study of cases.

## 5. Generalization practices: How do researchers engaging in case-oriented PER generalize from single cases? What purposes do such generalizations serve?

In this section, we begin to answer two questions about generalization practices: "How do researchers engaging in case-oriented PER generalize from single cases?" and "What purposes do such generalizations serve?" As before, we connect our answers to the paradigmatic assumptions we articulated in Section 3.

Case-oriented research takes up a perspective in which the universal is manifested in the particular details of specific cases, rather than (or in addition to) in the recurring patterns that emerge across cases.[24,46] With this perspective, it is not only possible but *necessary* that researchers look to cases to find universals (i.e., to make generalizable claims). This perspective is clarified in the following excerpt from the literature on research methodologies:

> "Mainstream positivist [recurrence-oriented] research on teaching searches for general characteristics of the analytically generalized teacher. From an interpretive [case-oriented] point of view, however, effective teaching is seen not as a set of generalized attributes of a teacher or of students. Rather, effective teaching is seen as occurring in the particular and concrete circumstances of the practice of a specific teacher with a specific set of students 'this year,' 'this day,' and 'this moment'…This is not to say that interpretive [qualitative] research is not interested in the discovery of universals, but it takes a different route to their discovery, given the assumptions about the state of nature in social life that interpretive researchers make. The search is not for *abstract universals* arrived at by statistical generalization from a sample to a population, but for *concrete universals*, arrived at by studying a specific case in great detail and then comparing it with other cases studied in great detail. The assumption is that when we see a particular instance of a teacher teaching, some aspects of what occurs are absolutely generic, that is, they apply cross-culturally and across human history to all teaching situations...These [universal] properties are manifested in the concrete, not in the abstract."[24]

Erickson's explanation that "some aspects of what occurs [in a particular instance of a teacher teaching] are absolutely generic" corresponds to the idea that, in a given moment, the teacher instantiates a theory (or several) – that some of their actions are a *case of* a more general theory. Thus, case-oriented PER uses the word "general" to signify representative of or embodying a theory. In this paradigm, generalizations are made by connecting case to theory. In contrast, in recurrence-oriented PER, the word "general" typically means representative of a population or phenomenon; generalizations are made by identifying patterns and relationships that emerge across cases.

I think the question of generalization is particularly significant to understanding case-oriented research: for many researchers, including myself when I started at SPU, generalization[f] is the *purpose* of research. Unarticulated paradigmatic differences in the meaning of generalization or in assumptions about where universals are located, how to locate such universals, and how research generalizations might serve the broader community may promote misinterpretations of case-oriented research as ungeneralizable or without purpose. At least this was the case for me – not so much that I thought case-oriented research *was* without purpose, but definitely that I did not know (and/or could not articulate) what it was. Hence our second question, "What purposes do generalizations from single cases serve?," answered in Sections 5.2 and 5.3.

### 5.1 Researchers separate the universal from the particular by identifying what a given case is a *case of*.

Drawing on the premise articulated above – *that* the general is manifested in the particular – researchers engaging in case-oriented PER seek to

> "uncover the different layers of universality and particularity that are confronted in the specific case at hand – what is broadly universal, what generalizes to other similar situations, what is unique to the given instance."[24]

In other words, researchers first discern what the case under study is a *case of*, and then they build connections between case and theory (i.e., the universal). We sense that this is what one of our interviewees means when she says:

> "I feel like for me, the process starts when somebody with pretty good researcher or teacher eyes sees something that wows them [and then goes on to] look for other things in the video that maybe

[f] I spend most of this section articulating differences I see between the meaning of generalization in case- and in recurrence-oriented research. However, I do think these paradigms share an emphasis on generalization as producing insights that are applicable in situations beyond the one(s) at hand.

> seem similar to you, so that you can maybe start to say, "Listen, I don't know what it is, but I feel like these things all go together. I think these are all about the same thing. *What thing are these all about?"* To help you articulate your sense of what matters about the episode." (quoted above)

These connections between case and theory reveal ways in which the universal (theory) is manifested in the concrete (case), providing opportunities for clarification of the theory itself and, in some cases, pointing researchers to new theoretical territory (e.g., when the theories that are brought to bear cannot account for the details of the case).[33,56] Without these case-theory connections, selection and analysis of episodes does not produce generalizations; there are a nearly limitless number of possible – even true – claims that can be made about any given episode, but a limited number have theoretical significance, or "matter," in the words of our interviewee.

Eisenhart[49] offers an illustration not only of the kind of theory that case-oriented research seeks to discover but also of how connections (i) between cases or (ii) between case and existing theory clarify the theory itself. She excerpts Becker's[57] summary of in-depth studies of men's prison culture and efforts to generalize these results to a women's prison. In her description, studies of men's prisons demonstrated that inmates develop an elaborate prison culture, including a market for scarce material goods and personal services and a code of conduct emphasizing the protection of information. Researchers attributed these inventions to the deprivations of prison life. Other researchers, with this theory in mind, studied a women's prison, but found no comparable code of conduct and a very different structure of social relationships. Rather than invalidating the original theory that the deprivation of prison life led to the creation of a prison culture, the new study extended the theory, recognizing that prison deprives women of different things than men (e.g., familial protection rather than autonomy) and that "[t]heir culture responds to that difference."

Knowledge of how researchers conducting case studies discern the universal from the particulars of a given case gives us insight as we evaluate and seek to understand research in our own field. For example, this knowledge supports us in making explicit the process by which Berland and Hammer[39] build connections between case and theory. These authors articulate what their three episodes are *cases of,* saying:

> "The empirical case we present shows multiple stabilities in the students' and teacher's understandings of what is taking place during argumentative and more traditional class discussions, with dynamics at the levels both of individuals and of the class as a whole. The theoretical case we present is that these phenomena of student, teacher, and class dynamics connects to prior work on *frames* and *framing* [citations from framing literature]…"

Throughout their analysis, they connect the dynamics of their cases to the literature on framing, offering their readers a particular theoretical lens through which to view the

discussion. They also point out the ways in which the students instantiate (or do not instantiate) scientific argumentation practices in each discussion. They use their cases to inform theory by showing how specific framings do or do not align with productive argumentation practices. Consistent with case-oriented research premises, this connection is forged through *context*: the authors point to literature that suggests that certain contexts – or certain ways that students "experience their context" – support and call for scientific argumentation practices, and they show that the meanings Mr. S' students are making of their classroom discussions support and/or call for such practices. Berland and Hammer also connect case to case, drawing parallels and distinctions between the framings, dynamics, and argumentation practices in the three episodes.

This means of discovering universals contrasts with that in recurrence-oriented PER, in which the task of the researcher is to uncover patterns and relationships that are representative (and thus predictive) of some phenomenon or population. In practice, this is often difficult, since it requires random sampling (i.e., sampling that ensures that each member of a population has an equal chance of being selected) and, in some cases, random assignment (i.e., each member of a sample has an equal chance of being assigned to the control and experimental groups).[58] When these requirements cannot be met, researchers look for patterns and relationships

> "us[ing] a different generalization model, one that emphasizes how consistently a causal relationship reproduces across multiple sources of heterogeneity…The operative question is: Can the same causal relationship be observed across different laboratories, time periods, regions of the country, and ways of operationalizing the cause and effect?"[23]

The model that this quote describes maps onto several of Pollock and Finkelstein's[26] claims. For example, as discussed earlier, the authors seek to discern whether the effectiveness of *Tutorials* is reproducible across institutions and intra-institutional implementations. This is particularly relevant to claims of representativeness: neither the original implementation of *Tutorials* (by the curriculum developers) nor the original implementation at CU-Boulder (the authors' institution) represent typical implementation conditions. To address questions about representativeness, the authors reproduced the measurements across and within institutions. The authors found that the gains posted (1) by CU-Boulder and (2) by non-PER faculty at CU-Boulder were consistently large, which suggests that the *Tutorials* may be *generally* effective.

### 5.2 Single cases illustrate theories and broaden awareness.

Even if one understands *that* case-oriented research subscribes to a different model for generalization and understands *how* researchers go about discerning the universal from

the particular, one may still be left with the question of how such research is *useful* if it is derived from a single case or context. This is basis for our second question, "What purposes do generalizations from single cases serve?" Our answer to this question is that just as case-oriented and recurrence-oriented research are predicated on different (1) perspectives on where universals manifest and (2) stances toward cause in the social world/human behavior, so also do these paradigms differently frame the usefulness of research results. Case-oriented PER produces research results that are useful in an *awareness* sense, rather than in a *representative* sense. For example, one researcher whom we interviewed stated his goal as "expanding the perspective" of his readers:

> "My goal is to say at the end of the day that I've expanded the perspective that a reader of my work might have. That at the beginning they wouldn't have thought that this [event] was something of interest, then notice this set of things that happened, the complexity of it, the richness of it, and notice how much we could be paying attention to or are paying attention to when we naturally interact with the world."

In particular, case-oriented research broadens readers' awareness by illustrating what a theory *looks like* in context, which may heighten readers' theoretical vision in their own contexts, or by refining theory, adding to readers' existing theoretical understanding. The notion that (i) providing readers with new theoretical lenses will shape what readers see and how they see it is grounded in the case-oriented assumption that (ii) it is impossible to separate one's theoretical commitments (which affect how one meaning-makes) from one's vision of and for events.

The research literature affirms this purpose of case-oriented PER, calling this kind of research particularly appropriate for enhancing readers' awareness of (and in) situations similar to the one studied.[21,22] The situation need only be similar in that it instantiates the same theory; it may be very different in other ways. Using Eisenhart's illustration above, the theory that "prisoners develop a culture that solves the problems created by the deprivations of prison life" can explain the cultures of both the men's and women's prisons, even though the specific cultures developed in men's and women's prisons look very different in practice. Erickson[59] says that for qualitative research,

> "judgments of external validity [generalizability] lie in the eye (and experience) of the reader. If you as a reader recognize in my descriptions processes you find also at work in settings you know, then you are determining that what I am saying below 'generalizes' beyond the cases I am reporting."

Wehlage[20] calls this "generalization by analogy":

> "The implication of this view is that generalization is more like thinking by analogy than discovering law-like empirical relationships…[O]ne is to use the data from the case study as an example of the kind of thing that happens in situations like that…Despite some difficulties with this concept, there

> is strong appeal in the notion that field studies can provide us with a broader range of (surrogate) experience than we could otherwise have, and that the generalized insights that we can take from this experience will help us to act more intelligently in future contexts."

Berland and Hammer's[39] manuscript reflects the position that theory development is especially suited for broadening the reader's perspective and sharpening awareness. In particular, the authors highlight the "productive resources for argumentation" that students in Mr. S' class bring to bear in the "argumentative" discussion. They suggest that educators recognize and "tap into" these resources by fostering contexts that "students recognize as argumentative." By illustrating what such resources can look like, Berland and Hammer enhance readers' vision for their own students' resources for argumentation, and by proposing a theoretical connection between framing and argumentation, they broaden readers' perspective of how they might foster argumentation (through certain framings) or what may be constraining the engagement of their students in argumentation. Berland and Hammer do not prescribe a specific structure for contexts that promote argumentation; in fact, they warn against focusing on "steps or components of argumentation." Instead, they recommend that instructors support students in achieving results that are meaningful to them.

Striving for usefulness through awareness is more appropriate to case-oriented research than is striving for representativeness, given the case-oriented research premise that social action is guided by *locally* constructed meanings. What the researcher hopes the reader will become aware of is the universal – for example, *that* people coordinate multiple modalities to communicate understanding, not that *specific* (future) participants will coordinate *specific* modalities to communicate a *specific* understanding. In other words, the aim is not to establish that the universal would manifest in a predictable way in another setting. One interviewee said:

> "…if [another researcher is] interpreting my work to be about [predictability], then no wonder [they] *ought* to be disappointed, because that's not it. I think of my research as being much more likely to illustrate the mechanism by which some learning process occurs. It would be about a small number of situations, and it would hopefully make a convincing case that that can be how things work, but not that it is how things would work next time."

Recurrence-oriented research, on the other hand, aims to offer readers results that are useful because they are *representative* and can thus inform predictions about a particular population*:*

> "You'd like to be able to make generalizations about *some* population that is reproducible. It doesn't have to be average, it could be the above-average students. You could say, 'We have studied the above-average students, but we believe that these results are representative of a definable group, more than in the spring semester of my Physics 102 course at X university.' [I]t's of no interest to

> anyone else unless you can show how it relates to their situation. So science works on generalizable results."

Thus, if one wants to know whether a recurrence-oriented study is useful for one's own context, one must first figure out whether one's sample is representative of the same population as the sample in the study. If this is the case, the same parameters and relationships should apply. This concern for representativeness as a critical element of cross-contextual generalizability to other contexts (i.e., external validity) is echoed by the literature on (quantitative) experimental validity.[60-63]

Pollock and Finkelstein's[26] paper reflects the "representative" sense of the usefulness of research. They offer extensive demographic information and implementation specifics – information that others could use to compare their sample to the authors'. They choose $\langle g \rangle$ as their measure of student achievement, a construct that has been shown to be independent of students' pre-test scores[29] (thus normalizing their pre-instructional state). They anticipate that some readers will call into question the *representativeness* of their claim that the *Tutorials* are effective and that these readers may instead think that the effectiveness of the *Tutorials* is attributable to well-prepared faculty. They acknowledge the limitation of their study in explicitly addressing this concern, saying, "Of course, the most compelling study would be to control for faculty member and vary the curricula. However, we have not had opportunity to [do so]." However, they point to the consistent positive effects of the curriculum *across* instructors, marshaling this recurrence as evidence that "the materials themselves…are important."

### 5.3 Single cases address certain types of research goals and questions.

Part of what makes case-oriented research appropriate for the articulation of specific kinds of generalizations is its alignment with particular research goals, as reported by our interview participants. For example, case-oriented PER seeks to broaden audience perspective by illustrating, building, and refining theories. Researchers clarify participants' points of view, reveal and challenge implicit assumptions, demonstrate possibility, develop mechanisms that explain certain teaching and learning phenomena, and coordinate multiple modalities to better understand thinking and learning. Recurrence-oriented PER, on the other hand, seeks to help readers plan and predict instruction by identifying recurring teaching and learning phenomena, such as conceptual difficulties that students may encounter when learning concept *x*; and instructional causes and effects, such as variables that influence learning gains and misconception-like patterns in student responses. We can logically connect each of these aims to a researcher's choice to study single cases or representative populations. For example, if one seeks to reveal and challenge implicit assumptions, one need only

deeply study a *single* (detailed) instance that contradicts a standard assumption. (The goal in doing so is to refine the assumption, not simply to demonstrate the contradiction.) Likewise, to demonstrate possibility – e.g., to show that a type of interaction is possible or that a type of learning *can* happen in a science classroom – only requires a single instance. On the other hand, large *N* is necessary if one wants to make claims about instructional effectiveness for a *representative* student or population. Similarly, if one seeks to identify prevalent student ideas or get a broad sense of the set of ideas that introductory physics students may have about a particular concept, one must ask questions of large numbers of students.

Tables 1 and 2 summarize the claims made by eight additional papers that we identified as case- and recurrence-oriented PER. In each case, we provide the evidence that authors cite for their claim(s) and the significance of the analysis as a whole.

**Table 1.** Claims and evidence from examples of case-oriented physics education research

| Author and title of paper | Major claim(s) | Evidence for claim(s) | Significance: Informs theory about? |
|---|---|---|---|
| Harlow, "Structures and Improvisation for Inquiry-Based Science Instruction: A Teacher's Adaptation of a Model of Magnetism Activity"[64] | (1) Ms. Carter (elementary school science teacher) revises a research-based "Models of Magnetism" PET[65] activity to test the models for a magnetized nail that her students propose (rather than the "charge separation" model that the curriculum anticipates). The "*differences* [in the enacted curriculum] at the event level were necessary to preserve the *similarities* at the scientific practice level."<br>(2) Ms. Carter drew on knowledge of "the nature of science, children's learning, and science content" as she transformed the activity. | (1) Ms. Carter asks her students to conduct experiments that (i) challenge their proposed models for a magnetized nail and that (ii) are different from the experiments proposed by the "Models of Magnetism" unit.<br>(2) Ms. Carter suggested her students test their models, deviated from the experiment planned by the curriculum, and proposed a model that challenged her students' models. | (1) Role of adaptive instruction in inquiry-based learning<br>(2) Pedagogical content knowledge entailed by scientific inquiry in the classroom |
| Gupta, Elby, and Conlin, "How substance-based ontologies for gravity can be productive: A case study"[66] | "Despite the seeming unproductiveness of a substance ontology of force, we argue that thinking of gravity as 'stufflike' contributed to learners' conceptual process in learning about gravity, forces, and motion, progress that would likely have been less transformative for the participants had the instructors steered learners away from this 'misontology.'" | (1) Lynn's and Daniel's use of substance-like reasoning about gravity fed into the correct "Newtonian compensation" argument for why more and less massive objects fall at the same rate.<br>(2) Lynn cites "figuring it [the answer to this question] out myself" as a transformative experience. | Role of misontologies in learning |
| Lising and Elby, "The impact of epistemology on learning: A case study from introductory physics"[67] | Jan (a student enrolled in an introductory physics course at the University of Maryland) experiences an "epistemological barrier" between everyday and formal reasoning that often "keeps [her] from looking for | Jan chooses (consciously or unconsciously) (i) not to use knowledge in the classroom context that she did use in interview contexts and (ii) not to *reconcile* formal and informal knowledge in either context. | Role of epistemology in learning |

| | | | |
|---|---|---|---|
| | connections between ideas from the different sides." | | |
| Richards, "The Role of Affect in Sustaining Teachers' Attention and Responsiveness to Student Thinking"[68] | The affective experiences of Ms. L (fifth grade science teacher) and Ms. R (sixth grade science teacher) promoted and sustained their in-the-moment attention and responsiveness to student thinking. | Ms. L and Ms. R shifted their attention toward student thinking immediately following displays of curiosity/excitement and concern/frustration, respectively, and their responses to student thinking were plausibly linked to these affective experiences. | Role of teacher affect in teacher attention and responsiveness |

The claims illustrated by these four examples of case-oriented PER are interpretations of what is happening in a specific instance that draw on and have the potential to inform theory. Evidence for the claims are sequences of local events. Harlow's case shows the central role that adaptive instruction plays in elementary students' authentic scientific inquiry; Gupta et al. challenge the notion that misontologies are always unproductive for learning; Lising and Elby propose a mechanism that links epistemology and learning; and Richards bridges the literatures on affect and teacher attention/responsiveness to show that the former may play a role in stabilizing the latter. In each paper, the authors limit their claim(s) to the details of their individual cases; however, in providing a narrative and interpretations of their local events, these researchers show the audience what it can look like for a specific theory to "show up," in context, seeking to expand readers' vision for and understanding of their own contexts.

**Table 2.** Claims and evidence from examples of recurrence-oriented physics education research

| **Author and title of paper** | **Major claim(s)** | **Evidence for claim(s)** | **Significance: Informs predictions about?** |
|---|---|---|---|
| Brewe, Traxler, de la Garza, and Kramer, "Extending positive CLASS results across multiple instructors and multiple classes of Modeling Instruction"[69] | "Modeling Instruction curriculum and pedagogy support the development of more favorable attitudes toward learning physics, independent of instructor." | Recurrence of attitudinal gains (as measured by the CLASS) across Modeling Instruction implementations | Factors that may contribute to positive attitudinal shifts |
| Hazari, Potvin, Lock, Lung, Sonnert, and Sadler, "Factors that affect the physical science career interest of female students: Testing five | "[D]iscussions about women's underrepresentation [in science] have a significant positive effect" on the physical science career interest of female students; having "a single-sex physics class," "female physics teacher," and/or "female scientist guest speakers in science class" does | (1) Females who reported experiencing discussions of women's underrepresentation in their high school physics course were significantly more likely to express interest in pursuing a career in the physical sciences than were females who did not experience such discussions. (Both groups of | Classroom conditions that affect physical science career interest in females |

| | | | |
|---|---|---|---|
| common hypotheses"[70] | not, nor does "discussing the work of female scientists in physics class." | females were statistically equivalent on the confounding variables.)<br>(2) Females who reported (i) having a single-sex physics class, (ii) having a female physics teacher, (iii) having female scientist guest speakers in science class, or (iv) discussing the work of female scientists in physics class were statistically indistinguishable from females who did not, in terms of their expression of interest in pursuing a career in the physical sciences. (In each case, both groups of females were statistically equivalent on the confounding variables.) | |
| Koenig, Endorf, and Braun, "Effectiveness of different tutorial recitation teaching methods and its implications for TA training"[71] | *Tutorials*[28] recitation sections that use cooperative group work and Socratic instructor-student dialogue are more effective than those that use traditional lecture, individual group work, or cooperative group work with no Socratic dialogue. | (1) Introductory physics students who participated in (a) a recitation section that incorporated cooperative groups and Socratic dialogue between TAs and students performed significantly better on a conceptual post-test than did students who participated in recitation sections that incorporated (b) traditional lecture, (c) individual group *Tutorials* work, or (d) cooperative group *Tutorials* work (with no Socratic dialogue).<br>(2) Fewer students experiencing style (a) "continued to use the same incorrect reasoning" on the post-test as on the pre-test, and more students in style (a) used the work-kinetic energy and impulse-momentum theorems on the post-test, compared to styles (b), (c), and (d). | Necessary components of physics teaching assistant training |
| Mikula and Heckler, "Student Difficulties With Trigonometric Vector Components Persist in Multiple Student Populations"[72] | (1) "[S]tudent difficulties with trigonometric vector components are persistent and pervasive...and [students' trigonometric vector component skills] are far below the requisite near-perfect accuracy needed for such fundamental skills."<br>(2) Both (i) percentages of correct answers and (ii) percentages of certain errors in student reasoning depend on (a) the angle configuration and (b) the component of the vector requested. | (1) When asked to write an expression for the components of a vector after relevant course instruction, introductory physics students often made sign errors, interchanged sine and cosine, and "answer[ed] based on incorrectly drawn triangles and incorrectly placed angles."<br>(2) Changing the angle configuration (while keeping the component requested the same) changes the (i) percentages of correct answers and (ii) percentages of certain errors in student reasoning. Changing the component requested (while keeping the angle configuration the same) changes the (i) percentages of correct answers and (ii) percentages of certain errors in student reasoning. | Difficulties students may have with vector components and factors that may affect difficulties |

In contrast to the claims listed in Table 1, claims made by the examples of recurrence-oriented research in Table 2 are posed in terms of population-level patterns and relationships that have the potential to inform instructional predictions. Evidence for the claims includes the recurrence of research results and statistically significant

differences between groups that are equivalent on all relevant measures but one. Brewe et al. demonstrate the effectiveness of a research-based curriculum in shifting students' attitudes about science; Hazari et al. determine what affects females' physical science career interest; Koenig et al. discern which of four recitation styles is most effective at improving student performance on written conceptual questions; and Mikula and Heckler report patterns and dependencies in student understanding of vector components. In drawing from random samples and/or privileging recurrence and reproducibility, these authors make claims about *populations*, seeking to inform predictions about the effectiveness of certain kinds of instruction for certain groups of students, or about what instructors may expect from these students.

## 6. Theories of cause: How do researchers engaging in case-oriented PER make causal claims from single cases? What purposes do such claims serve?

In this section, we answer our third pair of questions, "How do researchers engaging in case-oriented PER make causal claims from single cases?" and "What purposes do such claims serve?" Our answer to the first is that in case-oriented PER, cause is inferred from process – i.e., from a visualizable sequence of events that plausibly links local causes and effects – rather than from controlled experiments, as in recurrence-oriented PER. This perspective can be connected to the case-oriented research premises that action is shaped by the meaning-perspectives of participants and that reality is subjectively constructed. Meaning-making, and thus *cause* in the case-oriented research paradigm, is a process that is revealed in the details of specific cases as they unfold.

This question and its answer was particularly significant to me for the following reason: even after I had accepted that universals can be discerned from the particulars of cases, I still assumed that causal claims should be generated via controlled experimentation. However, this assumption was embedded within a theory of cause more characteristic of recurrence-oriented research. Without understanding the distinct theory of cause associated with the case-oriented PER paradigm, I was confused by the process by which causal claims are generated in case-oriented research. It was only when I understood case-oriented research assumptions and how they bore out in this different theory of cause that I could meaningfully engage in and with case-oriented PER.

We answer the second question – "What purposes do claims generated from single cases serve?" – in a way parallel to our answers about generalization from Section V.

### 6.1 Researchers infer cause from a visualizable sequence of events that plausibly links local causes and effects

Our answer to the first question, "How do researchers engaging in case-oriented PER make causal claims from single cases?," is grounded in the case-oriented PER premise that social action is guided by locally-constructed participant meanings. That is, in case-oriented research, the *causes* of social action are participants' locally meaningful interpretations of their environments.[24,41] These meaning-perspectives are not only context-dependent but also dynamic, evolving as participants continually make sense of (and shape) their contexts and respond to other participants who are simultaneously making sense of (and shaping) the context.[25,34,44] Participant meanings often include non-physical entities, such as mental processes, that cannot be converted into variables without misrepresenting their true nature.[31,73] This perspective is echoed by the following quote:

> "When an historian asks 'Why did Brutus stab Caesar?' he means 'What did Brutus think, which made him decide to stab Caesar?' The cause of the event, for him, means the thought in the mind of the person by whose agency the event came about; and this is not something other than the event, it is the inside of the event itself."[74,g]

Because these meanings are inextricably linked to the local context, causal mechanisms do not necessarily produce regularities.[25] What is of interest in a causal account is how a *specific* event evolves and what mechanisms shape it[25,73] – those "processes that resulted in a specific outcome in a particular context."[75]

These assumptions are connected to the choice to infer causality from "visualizable sequences of events, each event flowing into the next" (Maxwell,[25] quoting Weiss[76]), as in Figure 1. Smith[77] traces this to Gould,[78] who writes:

> "…explanation takes the form of a narrative: E, the phenomenon to be explained, arose because D came before, preceded by C, B, and A. If any of those earlier stages had not occurred, or had transpired in a different way, then E would not exist (or would be present in a substantially altered form from E, requiring a different (but equally credible) explanation."

Thus, in this perspective, causality is "observed" when a sequence of events is connected by a plausible and compelling explanation (rooted in evidence from, for example, a video episode) for why one event follows the other. Because this theory deals with *local* causality – those events and processes that lead to specific outcomes –

[g] Case-oriented research that sought to understand why Brutus stabbed Caesar might connect this case to theory about, for example, some of the elements of violent political struggles or possible motives for murder.

it lends itself well to case studies or other methods that use small numbers of individuals.[25]

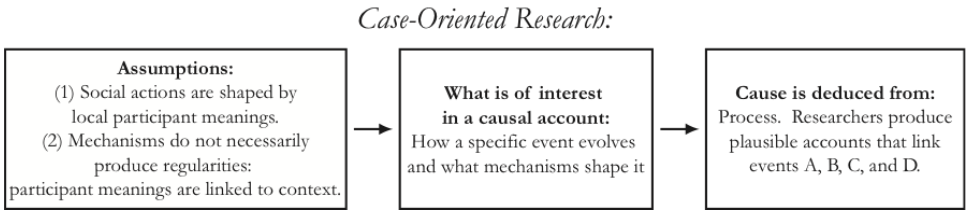


**Figure 1:** Simplified relationship between assumptions, focal points of causal accounts, and means by which cause is inferred in case-oriented research. The relationship between these three is more complex than this; each of the entities in the boxes may inform the other, and the arrows we have drawn can go the other way (e.g., from what is of interest to the assumptions taken up).

For example, one researcher we interviewed described a causal claim that she inferred from process:

> "I think the thing I said was that Energy Theater *promotes,* and I think that…matters, because it's not a claim about frequency, it's a claim about process…I want to make a causal relationship between the Energy Theater representation and this disambiguating matter and energy. So I'm trying to say things like "because they were moving around material bodies in order to figure out what happened with the energy in the light bulb…see how…the ropes…assisted them in thinking about such-and-such. And had it been a bar chart, like, see how that really could only have happened because of the way Energy Theater is, right." So…because my claim is a *promotes,*…if I can show like a plausible causal relationship, even in one good case, if I can do it well enough that you…the reader are like, "Wow, yeah, it is the ropes, isn't it?," then I don't have to show you ten examples of it being the ropes, because you get it, you believe that it's because of the ropes and you see easily that it would be because of the ropes."

When this researcher says "see how the ropes assisted them in thinking about such-and-such," she refers to a sequence of events in which teachers' interactions with ropes (a part of the Energy Theater representation) play a role in the disambiguation of matter and energy.

Berland and Hammer[39] make several causal claims, all deduced from process. In particular, they explain the dynamics of each episode in terms of framings – which embed particular social and epistemic expectations – that are co-constructed and dynamically negotiated by participants. For example, the authors attribute the discord in the third episode to inconsistent expectations among students and teacher. Figure 2 outlines their argument: In the midst of the argumentative discussion, in which students stably assume social and epistemic authority (and the teacher reinforced this framing),

Mr. S bids to "move on" and walks to the front of the room, signaling a framing shift and a bid for others to recognize *him* as social and epistemic authority. Several students take up his bid, requesting answers from Mr. S and cheering when he offered them. Others, however, maintain the argumentative discussion frame, sustaining their social and epistemic authority by talking out of turn and challenging or questioning Mr. S' claims, which "le[a]d[s] to instabilities". As they do so, Mr. S dismisses their challenges, pressing them to accept his answer, and disciplines students who talk out of turn, reinforcing his framing of himself as social and epistemic authority. The authors summarize by suggesting "that the tension that emerged in the discordant discussion *resulted from* the combination of more traditional school framings and those that align with scientific argumentation" (emphasis added).

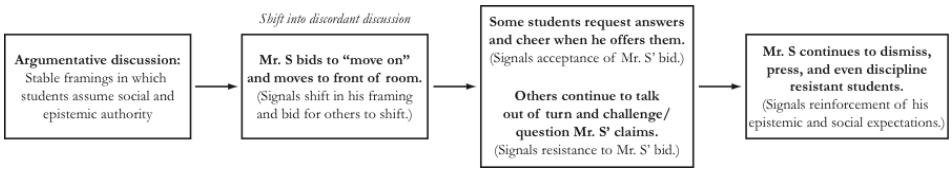


**Figure 2:** Outline of Berland and Hammer's[39] causal claim, which connects actions to evolving framings

In short, the authors attribute the discord that originally captured their attention to the *meanings* that participants make of their participation and authority in this discussion. These meanings evolve dynamically, shifting or stabilizing (or even strengthening) in response to verbal and nonverbal signals communicated between participants (e.g., Mr. S' dismissal and discipline of particular students was a response to their resistance of his original bid, and they were perceived as resistant in the context of Mr. S' shifting his framing of the discussion).

In contrast, in recurrence-oriented research, the *causes* of human behavior are (population-level) relationships between variables. These relationships are thought to exist "independent of [scientists'] personal values and sociopolitical beliefs."[38] (See also Ref. 40.) What is of interest in a causal account are the context-independent relationships that can predict population-level behavior. These assumptions are connected to the choice to infer cause from controlled experiments, as in Figure 3.

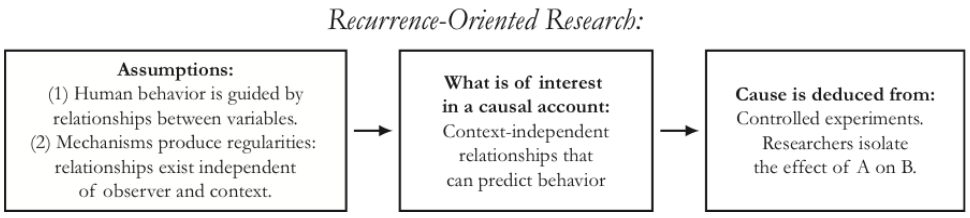


**Figure 3.** Simplified relationship between assumptions, focal points of causal accounts, and means by which cause is inferred in recurrence-oriented research. As in Figure 1, the relationship between these three is more complex than this; each of the entities in the boxes may inform the other, and the arrows we have drawn can go the other way (e.g., from what is of interest to the assumptions taken up).

Inferring population-level causal relationships not only requires random sampling (to ensure that the sample is representative of the population at large); it also requires that treatment and control groups are statistically equivalent on all measures but the variable being tested (or that the effect of confounding variables is measurable via some other means).[37,58] The latter ensures that the difference(s) between the treatment and control group can be attributed to the treatment, and not to some other variable that has not been taken into account.[79] Researchers use experimental techniques – such as random assignment to treatment and control groups – and statistical methods – such as multi-variate regression analysis – to meet this requirement.[23,36,60]

This orientation toward cause as inferred from controlled experiments was reflected in our interviews with physics education researchers and in our example of published recurrence-oriented PER. For example, one researcher described his work as

> "trying to study in detail why students are answering the way that they do. I'm very agnostic about that, and I try to take an extremely empirical approach…I try…as much as possible…to collect data…that people would regard as reliable and reproducible."

He goes on to say that typically, his research involves both reporting patterns in student responses to questions about topic *X* and trying to understand the variables that might affect students' responses to these questions. He pursues the latter via controlled experiments. For example, to understand which variables affect student responses to questions about two-dimensional 'time of flight', this researcher and his colleagues first devise questions that illustrate trajectories with different characteristics (e.g., different heights, ranges, areas under the curve). (In other words, they choose relevant variables.) They ask each question in multiple contexts to ensure that students interpret the question the way it was meant to be interpreted, eliminating questions that are idiosyncratic. Drawing from the remaining questions, they show students two trajectories at a time, each with different characteristics, and ask them to choose the

one that has the longest time of flight. The characteristics (inputs) are varied independently and simultaneously, and researchers watch to see how the patterns in student responses (output) change. He describes the results as follows:

> "Students tend to say that if something goes really, really far, it takes really, really long, independent of whether it goes high or not…[T]hose things come into play when they're making a decision, because you can see that when you change these things, their answer changes."

Pollock and Finkelstein's[26] manuscript serves as an example of published research that embodies this approach to causal questions and processes, using a *post hoc* experimental design. The authors make two (primary) causal claims: (1) that student learning gains are impacted by the curriculum, and (2) that student learning gains rely on faculty background. These claims are deduced from the covariation of the normalized gain $\langle g \rangle$ on a standardized assessment with the recitation curriculum ("all courses with tutorial experiences lead to learning gains higher than all classes that have traditional recitation sections"), and with faculty background (there is a "sizable variation of success among these implementations"). They posit reasons for the "potential effect of faculty", such as that PER faculty may be more familiar with the development of the reformed curriculum and may therefore implement it in a way that is better aligned with curricular goals. Pollock and Finkelstein's manuscript further instantiates this theory of cause by claiming that (emphasis added) "the *most compelling [evidence* that curriculum choice matters] would be to *control for faculty member and vary the curricula.* However, we have not had opportunity to [do so]." They go on to cite the recurrence of positive gains across instructors as evidence for this relationship in the absence of controlled experiments.

### 6.2 Researchers make generalizations by connecting local mechanisms to theory.

Researchers engaging in case-oriented PER infer causal claims from visualizable sequences of events, assuming a theory of cause in which social actions shape local participant meanings and in which mechanisms do not necessarily produce regularities. Researchers make *generalizable* claims from single cases by connecting local mechanisms to theory.[13,22,25] Maxwell[25] states that research efforts are "most productive if they are informed by, and contribute to, a detailed theory…of the causal process being investigated." For example, Berland and Hammer's[39] interpretation of the sequence of events in Mr. S' classroom is informed by theory on framing and theory on argumentation, and the relationship between these theories that emerges from their analysis – that certain framings may be more productive for argumentation – contributes to both theoretical spaces.

This contrasts with the practices of recurrence-oriented PER, in which researchers generalize causal relationships by replicating experiments or by observing the recurrence of causal relationships across contexts.[23,79,80] For example, Pollock and Finkelstein's manuscript[26] reproduces a curricular effect, showing that the *Tutorials* are generally effective (i.e., beyond non-standard educational contexts).

### 6.3 Process-oriented theory of cause addresses different research questions than experimentally-oriented theory of cause

Just as answering to different perspectives on generalization means that case- and recurrence-oriented research generate different kinds of claims, answering to different theories of cause means that these two paradigms generate different kinds of *causal* claims. Researchers (like my former self) who question the possibility of making causal claims on the basis of a single case may expect all research to establish population-level relationships that are predictive of other events. In this section, we revisit the additional examples of published case- and recurrence-oriented PER introduced in Section 5.3, highlighting those that make causal claims and indicating the means by which these claims were inferred. Our goal is to provide examples of the kinds of causal claims that case-oriented research can generate, to show how these are tied to the case-oriented theory of cause we have articulated in this section, and to contrast these claims with examples that are associated with the recurrence-oriented theory of cause.

As we discussed in Section 5.3, the causal claims made by the three manuscripts in Table 3 are about what is happening in a specific instance that draws on and has the potential to inform theory. In each case, the claims propose a mechanistic relationship that accounts for a series of events in a classroom context. For example, Gupta, Elby, and Conlin examine the evolution of one group of teachers' discourse to infer the productive role that a "misontology" plays in their understanding of gravity; Lising and Elby infer a causal relationship from the way in which Jan engages in a series of events, both in the classroom and in interviews; and Richards analyzes sequences of classroom events to show how specific affective experiences are initiated and then sustain teacher attention to student thinking.

**Table 3.** Causal claims made by additional examples of case-oriented PER

| Author and title of paper | Causal claim(s) | Cause was inferred from: |
|---|---|---|
| Gupta, Elby, and Conlin, "How substance-based ontologies for gravity can be productive: A case study"[66] | "…[Lynn's] Galilean reasoning emerged *because of*, not in spite of, the teachers' misontologies of gravity. And this idea of each coin in a roll feeling a certain amount of gravity then *fed into* the Newtonian compensation argument whereby the heavier object feels more gravitational pull but also puts up more resistance to getting moved." (emphases added) | Evolution of classroom discourse |

| | | |
|---|---|---|
| Lising and Elby, "The impact of epistemology on learning: A case study from introductory physics"[67] | Jan's experience of a "epistemological barrier" between everyday and formal reasoning that often "*keeps [her]* from looking for connections between ideas from the different sides." (emphasis added) | Series of classroom and interview events, refutation of alternative explanations |
| Richards, "The Role of Affect in Sustaining Teachers' Attention and Responsiveness to Student Thinking"[68] | The affective experiences of Ms. L and Ms. R *stabilize* their attention and responsiveness to student thinking. | Sequence of classroom events |

In contrast, the claims made by the recurrence-oriented papers listed in Table 4 are not accounts of what is happening in a specific instance but instead are reports of relationships between two variables for a particular population (e.g., female college students or introductory physics students). Koenig et al and Mikula and Heckler each plan and conduct controlled experiments to test relationships between variables; Hazari et al. use multi-variate matching techniques[81,82] to compose control and treatment groups after the fact; and Brewe et al. use the consistency of specific outcomes to propose plausible mechanisms for positive attitudinal shifts. Brewe et al. are careful to qualify that the co-occurrence of certain plausible causes and effects is "not adequate to draw causal conclusions regarding to what specifically the shifts should be attributed" but can "provide insight into the factors that could mechanistically explain" the shifts.

**Table 4.** Causal claims made by additional examples of recurrence-oriented PER

| **Author and title of paper** | **Causal claim(s)** | **Cause was inferred from:** |
|---|---|---|
| Brewe, Traxler, de la Garza, and Kramer, "Extending positive CLASS results across multiple instructors and multiple classes of Modeling Instruction"[69] | Students' positive attitudinal shifts:**<br>(1) Can be "attribute[d]" to Modeling Instruction.<br>(2) Likely do not "ar[i]se from a 'good semester' or any unique expertise of the professor."<br>(3) May be attributable to small class sizes.<br>(4) May be promoted by explicit focus on epistemological resources. | Consistent co-occurrence of variables of interest (or lack thereof) |
| Hazari, Potvin, Lock, Lung, Sonnert, and Sadler, "Factors that affect the physical science career interest of female students: Testing five common hypotheses"[70] | "[D]iscussions about women's underrepresentation [in science] *have a significant positive effect*" on the physical science career interest of female students; having "a single-sex physics class," "female physics teacher," and/or "female scientist guest speakers in science class" does not, nor does "discussing the work of female scientists in physics class." (emphasis added) | Multi-variate matching methods that isolates variables of interest |
| Koenig, Endorf, and Braun, "Effectiveness of different tutorial recitation teaching methods and its implications for TA training"[71] | The "manner in which the *Tutorials*[28] are taught" *affects* student understanding. | Controlled experiment that isolates variables |
| Mikula and Heckler, "Student Difficulties With Trigonometric Vector Components Persist in Multiple Student Populations"[72] | (i) Percentages of correct answers and (ii) percentages of certain errors in student reasoning *depend on* (a) the angle configuration and (b) the component of the vector requested. | Controlled experiment that isolates variables |

When comparing the claims made by case- and recurrence-oriented research, one point of confusion may be that each case-oriented claim can be rephrased in terms of a relationship between variables: ontologies affect learning, epistemology affects learning, and affect affects attention. Readers may wonder how these claims are different than those made by recurrence-oriented research. This brings up an important point: both case- and recurrence-oriented PER are seeking mechanisms that explain teaching and learning phenomena, and the product (claims) of both is often a mechanism. What differs between the two is where researchers expect mechanisms to appear (at the level of the population versus the case), the ways in which they are expected to generalize (to a population versus to theory), and how they expect to use them (to make predictions versus to broaden readers' awareness). When a researcher conducting case-oriented PER reports that "epistemology affects learning," they have likely brought the lens of epistemology to bear on an instance of learning and seen that it can explain how participants are making meaning of their experiences. They expect this lens to be useful in other contexts, but in ways that are intimately tied to these other contexts. The mechanism is not expected to produce regularities. On the other hand, when a researcher conducting recurrence-oriented research reports that "a particular intervention shifts students' attitudes about science," they have likely shown that this mechanism explains regularities in their data, and they expect it to continue to do so, such that readers can predict the effectiveness of the intervention for other students who are members of the same population.

# 7. Discussion

In this paper, we have answered questions about three practices in case-oriented PER:

**Table 5.** Summary of questions and answers

| Practice | Questions | Our answers |
|---|---|---|
| Selection | How do researchers engaging in case-oriented PER select episodes for analysis? | Foreground complex participant meanings<br>Select cases of theory |
| | How do they respond to the inherent subjectivity in selection? | Make bias visible |
| Generalization | How do researchers engaging in case-oriented PER generalize from single cases? | Separate the universal from the particular by identifying what a given case is a *case of* |
| | What purposes do such generalizations serve? | Illustrate theories and broaden awareness<br>Address specific research goals and questions |
| | | |

| | | |
|---|---|---|
| Causal-claims-making | How do researchers engaging in case-oriented PER make causal claims from single cases? | Infer cause from a visualizable sequence of events that plausibly links local causes and effects<br>Connect local mechanisms to theory |
| | What purposes do such claims serve? | Illustrate theories and broaden awareness<br>Address specific research goals and questions |

We showed ways in which these practices are grounded in the case-oriented paradigm itself – i.e., in assumptions about "knowledge, our social world, our ability to know that world, and our reasons for knowing it" that frame and guide a particular orientation toward research, "including what questions to ask, what methods to use, what knowledge claims to strive for, and what defines high-quality work."[15] Namely, to the assumptions that (1) social actions are guided by locally-constructed meanings and (2) reality is subjectively constructed.

We contrasted these practices with those in recurrence-oriented research, which is predicated on the assumptions that (1) human behavior is guided by predictable relationships between variables and (2) real phenomena are reproducible. In recurrence-oriented PER, then, selection involves foregrounding recurring patterns and relationships and/or clean data, and researchers are called on to reduce "selection bias."[40] Generalization is in the service of uncovering patterns and relationships that are *representative*, again foregrounding recurrence, and claims serve instructional planning and predictions. Cause is conceived in terms of population-level relationships between variables, often inferred from controlled experiments.

The answers to these three pairs of questions are what helped me to not only *do* but also *appreciate the rigor of* case-oriented PER. I can see now that I was trying to make sense of these practices from a recurrence-oriented research perspective. If I can use myself as a case, what I hope this paper illustrates is what is possible when one assumes that people do what they do for a reason and try to figure that reason out by listening to them – i.e., that empathy can produce appreciation and understanding. If I got to choose the purpose this paper serves in our community, it would be to contribute to lowering the "empathy wall"[16] and to deeper understanding and appreciation of one another's work.

Understanding these two research paradigms has also helped me to become more responsible in my participation in them both. For example, doing the work that produced this paper helped me to articulate the model of generalizability that I use as I engage in recurrence-oriented research, strengthening the theoretical framework that I lay out in my grants and papers. And, as I have reiterated over and over, answering

these questions *allowed* me to do case-oriented research, as a person who needs explicit framing to thrive.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Nos. 0822342 and 1222732. We gratefully acknowledge the following researchers for participating in interviews with Robertson: Leslie Atkins Elliott (now at Boise State University), Andrew Boudreaux (Western Washington University), Melissa Dancy (University of Colorado), Brian Frank (Middle Tennessee State University), Ayush Gupta (University of Maryland), David Hammer (Tufts University), Danielle Harlow (University of California, Santa Barbara), Charles Henderson (Western Michigan University), Paula Heron (University of Washington), Andrew Heckler (The Ohio State University), Stephen Kanim (New Mexico State University), Sarah (Sam) McKagan (American Association of Physics Teachers), David Meltzer (Arizona State University), Sanjay Rebello (now at Purdue University), Rosemary Russ (University of Wisconsin – Madison), Mel Sabella (Chicago State University), Rachel Scherr (Seattle Pacific University), and Michael Wittmann (University of Maine). We are thankful for feedback on this manuscript from James Day, Paula R. L. Heron, David E. Meltzer, Justin C. Robertson, Eleanor Sayre, Thomas M. Scaife, and Michael C. Wittmann. We also appreciate the thoughtful feedback from members of the Energy Project team, particularly Stamatis Vokos, from three anonymous reviewers, and from anonymous reviewers of previous iterations of this work.

## References

[1] A. D. Robertson, *An Investigation of University Student and K-12 Teacher Reasoning About Key Ideas in the Development of the Particulate Nature of Matter,* Ph.D. thesis, University of Washington, 2011.
[2] A. D. Robertson and P. S. Shaffer, "University student and K-12 teacher reasoning about the basic tenets of kinetic-molecular theory, Part I: Volume of an ideal gas," Am. J. Phys. **81**, 303 (2013).
[3] A. D. Robertson and P. S. Shaffer, ""Combustion always produces carbon dioxide and water": A discussion of university chemistry student use of rules in place of principles," Chem. Educ. Res. Prac. **15**, 763 (2014).
[4] A. D. Robertson and P. S. Shaffer, "University student reasoning about the basic tenets of kinetic-molecular theory, Part II: Pressure of an ideal gas," Am. J. Phys. **84**, 795 (2016).

[5] S. B. McKagan, R. E. Scherr, E. W. Close, and H. G. Close, in *Proceedings of the 2011 Physics Education Research Conference*, edited by N. S. Rebello, P. V. Engelhardt, and C. Singh (AIP Conference Proceedings, Melville, NY, 2012), p. 279.
[6] R. E. Scherr, H. G. Close, E. W. Close, V. J. Flood, S. B. McKagan, A. D. Robertson, L. Seeley, M. C. Wittmann, and S. Vokos, "Negotiating energy dynamics through embodied action in a materially structured environment," Phys. Rev. ST - Phys. Ed. Res. **9**, 020105 (2013).
[7] A. R. Daane, S. B. McKagan, S. Vokos, and R. E. Scherr, "Energy conservation in dissipative processes: Teacher expectations and strategies associated with imperceptible thermal energy," Phys. Rev. ST - Phys. Ed. Res. **11**, 1 (2015).
[8] A. R. Daane, S. Vokos, and R. E. Scherr, "Goals for teacher learning about energy degradation and usefulness," Phys. Rev. ST - Phys. Ed. Res. **10**, 1 (2014).
[9] A. R. Daane, L. Wells, and R. E. Scherr, "Energy Theater," Phys. Teach. **52**, 291 (2014).
[10] R. E. Scherr, H. G. Close, A. R. Daane, L. S. DeWater, B. W. Harrer, A. D. Robertson, L. Seeley, and S. Vokos, "Energy Tracking Diagrams," Phys. Teach. **54**, 96 (2016).
[11] R. E. Scherr, H. G. Close, S. B. McKagan, and S. Vokos, "Representing energy. I. Representing a substance ontology for energy," Phys. Rev. ST - Phys. Ed. Res. **8**, 020114 (2012).
[12] R. E. Scherr, H. G. Close, S. B. McKagan, and S. Vokos, "Representing energy. II. Energy tracking representations," Phys. Rev. ST - Phys. Ed. Res. **8**, 020115 (2012).
[13] V. K. Otero and D. B. Harlow, "Getting Started in Qualitative Physics Education Research," Rev. in PER **2** (2009).
[14] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (SAGE Publications, Inc., Thousand Oaks, CA, 2009), 3rd edn.
[15] J. C. Greene and V. J. Caracelli, "Defining and Describing the Paradigm Issue in Mixed-Method Evaluation," New Dir. Eval. **74**, 5 (1997).
[16] A. R. Hochschild, *Strangers in Their Own Land: Anger and Mourning on the American Right* (The New Press, New York, NY, 2016).
[17] E. W. Eisner and A. Peshkin, in *Qualitative Inquiry in Education: The Continuing Debate*, edited by E. W. Eisner, and A. Peshkin (Teachers College Press, New York, 1990), p. 171.
[18] E. Goffmann, *Frame analysis: an essay on the organization of experience* (Harvard University Press, Cambridge, MA, 1974).
[19] S. J. Derry et al., "Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics," J. Learn. Sci. **19**, 3 (2010).
[20] G. Wehlage, in *The study of schooling*, edited by T. Papkewitz, and B. Tabchmie (Praeger, New York, 1981), p. 211.

[21] R. Donmoyer, in *Qualitative Inquiry in Education: The Continuing Debate*, edited by E. W. Eisner, and A. Peshkin (Teachers College Press, New York, 1990), p. 175.
[22] J. A. Maxwell, "Understanding and Validity in Qualitative Research," Harvard Educ. Rev. **62**, 279 (1992).
[23] T. D. Cook, "Randomized Experiments in Educational Policy Research: A Critical Examination of the Reasons the Educational Evaluation Community has Offered for not Doing Them," Educ. Eval. Policy An. **24**, 175 (2002).
[24] F. Erickson, in *Handbook of Research on Teaching*, edited by M. C. Wittrock (Macmillan, New York, 1986), p. 119.
[25] J. A. Maxwell, "Using Qualitative Methods for Causal Explanation," Field Method. **16**, 243 (2004).
[26] S. J. Pollock and N. D. Finkelstein, "Sustaining educational reforms in introductory physics," Phys. Rev. ST - Phys. Ed. Res. **4**, 1 (2008).
[27] N. D. Finkelstein and S. J. Pollock, "Replicating and understanding successful innovations: Implementing tutorials in introductory physics," Phys. Rev. ST - Phys. Ed. Res. **1** (2005).
[28] L. C. McDermott, P. S. Shaffer, and the PEG at the University of Washington, *Tutorials in Introductory Physics* (Prentice Hall College Division, 2011), Preliminary 2nd edn.
[29] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand student survey of mechanics test data for introductory physics," Am. J. Phys. **66**, 64 (1998).
[30] E. G. Guba and Y. S. Lincoln, in *The Sage Handbook of Qualitative Research*, edited by N. K. Denzin, and Y. S. Lincoln (Sage Publications, Inc., United States of America, 2005), p. 191.
[31] W.-M. Roth, in *Generalizing from Educational Research: Beyond Qualitative and Quantitative Polarization*, edited by K. Ercikan, and W.-M. Roth (Routledge, New York, 2009), p. 235.
[32] J. W. Willis, in *Foundations of Qualitative Research: Interpretive and Critical Approaches* (SAGE Publications, Inc., 2007).
[33] W. A. Firestone, "Alternative Arguments for Generalizing from Data as Applied to Qualitative Research," Educ. Res. **22**, 16 (1993).
[34] E. G. Mishler, "Meaning in Context: Is There Any Other Kind?," Harvard Educ. Rev. **49**, 1 (1979).
[35] J. W. Schofield, in *Qualitative Inquiry in Education: The Continuing Debate*, edited by E. W. Eisner, and A. Peshkin (Teachers College Press, New York, 1990), p. 201.
[36] A. Field, *Discovering Statistics Using SPSS* (SAGE Publications Inc., Thousand Oaks, CA, 2009), 3rd edn.

[37] L. Ding and X. Liu, "Getting Started with Quantitative Methods in Physics Education Research," Rev. in PER **3** (2012).
[38] P. A. Moss, D. C. Phillips, F. D. Erickson, R. E. Floden, P. A. Lather, and B. L. Schneider, "Learning From Our Differences: A Dialogue Across Perspectives on Quality in Education Research," Educ. Res. **38**, 501 (2009).
[39] L. K. Berland and D. Hammer, "Framing for Scientific Argumentation," J. Res. Sci. Teach. **49**, 68 (2012).
[40] W. A. Firestone, "Meaning in Method: The Rhetoric of Quantitative and Qualitative Research," Educ. Res. **16**, 16 (1987).
[41] E. Bredo, in *Handbook of Complementary Methods in Education Research*, edited by J. L. Green et al. (Lawrence Erlbaum Associates, Mahwah, NJ, 2006), p. 3.
[42] K. M. Anderson-Levitt, in *Handbook of Complementary Methods in Education Research*, edited by J. L. Green et al. (Lawrence Erlbaum Associates, Mahwah, NJ, 2006), p. 279.
[43] F. Erickson, "Some Thoughts on 'Proximal' Formative Assessment of Student Learning," Yearb. Nat. Soc. Study Educ. **106**, 186 (2007).
[44] H. S. Becker, in *The Jack-roller*, edited by C. Shaw (University of Chicago Press, Chicago, 1966).
[45] https://arxiv.org/abs/1307.4135
[46] L. S. Shulman, "Those who understand: Knowledge growth in teaching," Educ. Res. **15**, 4 (1986).
[47] A. Wylie, *Thinking from things: Essays in the philosophy of archaeology* (University of California Press, Berkeley, CA, 2002).
[48] M. Freeman, K. de Marrais, J. Preissle, K. Roulston, and E. A. St. Pierre, "Standards of Evidence in Qualitative Research: An Incitement to Discourse," Educ. Res. **36**, 25 (2007).
[49] M. Eisenhart, in *Generalizing from Educational Research: Beyond Qualitative and Quantitative Polarization*, edited by K. Ercikan, and W.-M. Roth (Routledge, New York, 2009), p. 51.
[50] 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).
[51] S. Traweek, *Beamtimes and Lifetimes: The World of High Energy Physicists* (Harvard University Press, Cambridge, MA, 1988).
[52] S. Harding, *Whose science? Whose knowledge? Thinking from women's lives* (Cornell University Press, Ithaca, NY, 1991).
[53] S. J. Derry et al., "Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics," J. Learn. Sci. **19**, 3 (2010).
[54] F. Marton, "Phenomenography - A Research Approach to Investigating Different Understandings of Reality," J. Thought **21**, 28 (1986).

[55] A. Peshkin, "In Search of Subjectivity - One's Own," Educ. Res. **17**, 17 (1988).
[56] R. K. Yin, *Case study research: Design and methods* (Sage, Newbury Park, CA, 1989), 2nd edn.
[57] H. Becker, in *Qualitative inquiry in education*, edited by E. W. Eisner, and A. Peshkin (Teachers College Press, New York, 1990), p. 233.
[58] F. Pyrczak, *Making Sense of Statistics* (Pyrczak Publishing, Glendale, CA, 2006), 4th edn.
[59] F. Erickson, in *Mathematics Teacher Noticing: Seeing Through Teachers' Eyes*, edited by M. G. Sherin, V. R. Jacobs, and R. A. Philipp (Routledge, New York, NY, 2011), p. 17.
[60] G. Schraw, in *1998 Physics Education Research Conference Proceedings*, edited by T. C. Koch, and R. G. FullerLincoln, NE, 1998).
[61] D. M. Meltzer, in *2002 Physics Education Research Conference Proceedings*, edited by S. Franklin, K. Cummings, and J. MarxBoise, ID, 2003).
[62] M. Duneier, in *Workshop on Scientific Foundations of Qualitative Research*, edited by C. C. Ragin, J. Nagel, and P. White (National Science Foundation, 2004), p. 77.
[63] W. M. K. Trochim, *Research Methods Knowledge Base* (New York, NY, 2006). URL: http://www.socialresearchmethods.net/kb/index.php
[64] D. B. Harlow, "Structures and Improvisation for Inquiry-Based Science Instruction: A Teacher's Adaptation of a Model of Magnetism Activity," Sci. Educ. **94**, 142 (2010).
[65] F. Goldberg, S. Robinson, and V. Otero, *Physics and Everyday Thinking* (It's About Time, Armonk, NY, 2006).
[66] A. Gupta, A. Elby, and L. D. Conlin, "How substance-based ontologies for gravity can be productive: A case study," Phys. Rev. ST - Phys. Educ. Res. **10**, 1 (2014).
[67] L. Lising and A. Elby, "The impact of epistemology on learning: A case study from introductory physics," Am. J. Phys. **73**, 372 (2005).
[68] J. Richards, in *2013 Physics Education Research Conference Proceedings*, edited by P. Engelhardt, A. D. Churukian, and D. L. Jones (American Association of Physics Teachers, 2013), p. 301.
[69] E. Brewe, A. Traxler, J. de la Garza, and L. H. Kramer, "Extending positive CLASS results across multiple instructors and multiple classes of Modeling Instruction," Phys. Rev. ST - Phys. Educ. Res. **9**, 1 (2013).
[70] Z. Hazari, G. Potvin, R. M. Lock, F. Lung, G. Sonnert, and P. M. Sadler, "Factors that affect the physical science career interest of female students: Testing five common hypotheses," Phys. Rev. ST - Phys. Educ. Res. **9**, 1 (2013).
[71] 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**, 1 (2007).

[72] B. D. Mikula and A. F. Heckler, in *2013 Physics Education Research Conference Proceedings*, edited by P. Engelhardt, A. D. Churukian, and D. L. Jones (American Association of Physics Teachers, 2013), p. 253.
[73] J. A. Maxwell, "Causal Explanation, Qualitative Research, and Scientific Inquiry in Education," Educ. Res. **33**, 3 (2004).
[74] L. S. Shulman, *Paradigms and Programs* (Macmillan Publishing Company, New York, NY, 1986), Vol. 1, Research in Teaching and Learning.
[75] J. A. Maxwell, "The Importance of Qualitative Research for Causal Explanation in Education," Qual. Inquiry **18**, 655 (2012).
[76] R. S. Weiss, *Learning from strangers: The art and method of qualitative interviewing* (Free Press, New York, 1994).
[77] M. L. Smith, in *Handbook of Complementary Methods in Education Research*, edited by J. L. Green et al. (Lawrence Erlbaum Associates, Mahwah, NJ, 2006), p. 457.
[78] S. J. Gould, *Wonderful life* (Norton, New York, 1989).
[79] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Houghton Mifflin Company, Boston, MA, 2002).
[80] National Research Council, *Scientific Research in Education* (Center for Education, Division of Behavioral and Social Sciences and Education, Washington, D.C., 2002).
[81] J. S. Sekhon, "Multivariate and propensity score matching software with automated balance optimization: The matching package for R," J. Stat. Softw. **42** (2011).
[82] D. B. Rubin, "Bias reduction using mahalanobis-metric matching," Biometrics **36**, 293 (1980).