Stimulated recall interviews for describing pragmatic epistemology

Christopher W. Shubert[†] and Dawn C. Meredith[*]

The University of New Hampshire, Durham, New Hampshire 03824, USA

Students' epistemologies affect how and what they learn: do they believe physics is a list of equations, or a coherent and sensible description of the physical world? In order to study these epistemologies as part of curricular assessment, we adopt the resources framework, which posits that students have many productive epistemological resources that can be brought to bear as they learn physics. In previous studies, these epistemologies have been either inferred from behavior in learning contexts or probed through surveys or interviews outside of the learning context. We argue that stimulated recall interviews provide a contextually and interpretively valid method to access students' epistemologies that complement existing methods. We develop a stimulated recall interview methodology to assess a curricular intervention and find evidence that epistemological resources aptly describe student epistemologies.

I. INTRODUCTION

As instructors, we have seen students who have worked in ways we would not endorse: using equations without understanding what they meant, completing problems without considering whether or not the answer makes sense, or reading and re-reading the text with little comprehension. We suspect that these students' ideas about knowledge and knowing (i.e., their epistemologies) will not allow them to be successful because they are not engaged in the difficult cognitive work required for deep learning. Hammer, Elby, and colleagues have used detailed case studies to provide strong evidence that indeed students' epistemologies can affect learning negatively and should thus be attended to and explicitly addressed [1–4].

Our interest in student epistemologies arose as part of a larger project to reform our introductory physics course for life scientists [5]. Knowing that many students come to this course with unproductive epistemologies, we developed new labs based on Modeling Instruction [6,7], a curriculum with strong epistemological content. As part of the assessment of these reforms, we wanted to learn what epistemological ideas students used during the lab activities, if these ideas were productive, and if they matured and deepened over the two semesters of the course.

However, we soon realized that gaining access to student epistemologies posed significant methodological challenges. In previous studies, these epistemologies are either inferred from behavior in learning contexts or probed through surveys or interviews outside of learning contexts [3,4]. In this study we want to focus on something different: students' first-hand accounts of their ideas about learning and knowledge as they engage in learning activities. This might seem impossible in practice; how can students engage in authentic learning and reflect deeply on that learning simultaneously without changing what we hope to measure? In this paper, we argue that we can move closer to accessing this rich, privileged data using a new methodology based on stimulated recall interviews (SRI) [8,9].

We begin this article by articulating some important distinctions within epistemology, finally focusing on a notion of pragmatic epistemology. In this same section, we present resource theory as the framework for our study. We review epistemological research within physics education research (PER), and examine the claims to reliability and validity, making a case that there is more work to be done in this area. We introduce stimulated recall interviews and justify claims about the validity and reliability of this method. In the next section, we discuss Modeling Instruction as the locus of this study. Next, we give details about the design and implementation of our methodology which leverages SRI. Because this is a methodology paper, we give extensive details in these sections to facilitate refinement and use of the proposed methodology. Finally, we present the initial data gathered using this methodology as a proof of concept and draw conclusions about the students' use of epistemology in our labs.

II. THEORETICAL PERSPECTIVES

A. Personal scientific epistemology

While the field of epistemology is quite broad, in this section we reduce the scope of epistemology in ways that are appropriate for our study of students working within our

[*]dawn.meredith@unh.edu

[†]Present address: FiveStars Loyalty Inc., San Francisco, 94103, USA.

classrooms on activities that promote the learning of authentic physical science.

Epistemology, broadly defined, is the study of the nature of knowledge and knowledge construction. The focus of epistemology is summarized in the questions of Sandoval [10] and Duschl and Osborne [11]: What exactly do we know? How do we know what we know? And why do we believe it?

We begin our study by narrowing our focus to *scientific* epistemology. While there is no standard definition, a useful starting place is the epistemological themes that Sandoval [10] forwards: scientific knowledge is constructed, there are diversity of scientific methods (e.g., observations, controlled experiments), there are different forms of scientific knowledge (e.g., laws, theories, hypotheses), and some scientific claims are more tentative than others.

Then we focus on *personal* scientific epistemology, or epistemology held by individuals [12]. Investigating personal scientific epistemologies may show that an individual's understanding of the forms of scientific knowledge and the nature of the scientific enterprise are not what the scientific community would consider an appropriate or productive scientific epistemology [13–15]. For example, a student might not see a need for data to be replicable to be believed because they have not had the experience of designing a complicated experiment with the goal of minimizing uncertainties. Last, we focus on authentic school science explorations in the tradition of Modeling Instruction instead of professional science. These enterprises are very different. For example, Modeling Instruction is carried out in a linear and consistent way throughout the semester (see Table II for details); professional science is carried out more flexibly in order to respond to new insights, opportunities, or challenges.

Our work builds heavily on the work of Sandoval [10], who studied "practical epistemologies [defined] as the epistemological ideas that students apply to their own scientific knowledge building through inquiry." Sandoval engaged in this line of research to understand the relationship between inquiry instruction and students' epistemologies. He argued that the first step in such a study is to focus on practical epistemology and not formal scientific epistemologies. To carry out such a study, Sandoval suggested the line of research which we have taken up here: "a fruitful approach would be for researchers to develop interview protocols designed to elicit students' articulation of their reasoning behind epistemologically salient decisions."

Our research focus differs from Sandoval's in two important aspects. First, we broaden the scope of our study to include ideas about learning in addition to the traditional epistemological ideas about knowledge and knowing. These may seem like very similar constructs, but (at least in some cases) can be separated. For example, a naïve idea about the form of knowledge is that physics is a collection of unrelated and undisputed facts while a naïve idea about learning is that all of physics can be learned through memorization alone.

Sandoval [10,16] and Hofer and Pintrich [17] argue that learning and knowing are different but related activities and that keeping this distinction clear is essential to deepening our understanding of both personal epistemology and ideas about learning. However, Elby [18] cautions that we should let the data drive such distinctions rather than *a priori* forcing distinctions, as ideas about learning and knowing may be "inseparably entangled." He urges us to not yet converge on a definition of personal epistemology. We adopt the view of Elby, and allow for the possibility that ideas about learning and ideas about knowledge cannot always be separated.

From these explicit distinctions of epistemology, intended to clarify what we are studying in this project, we define "pragmatic epistemology" as students' ideas about knowledge, knowing, and learning in the context of their own practice of school science. In the next section, after some essential background, we will add one more qualification (concerning resource theory) to our definition of pragmatic epistemology, which will take us one step further from Sandoval's practical epistemology.

B. Resource theory and the form of epistemology

We adopt Hammer's resources framework [19] for our study. This framework is built on Minsky's computational model of the mind [20] and the "knowledge in pieces" theoretical perspective of diSessa [21]. A resource itself is a bit of knowledge that can be applied in many contexts. Epistemological resources allow students to reflect on and guide their own learning and knowing [2]. For example, the epistemological resource [2,19] "knowledge as propagated stuff" is useful when finding out what your partner wants from the grocery store. However, this same resource is not useful to guide a student's entire approach to learning physics. This is a hallmark of resources: they are neither right nor wrong in themselves, but productively or unproductively applied in different contexts. Analogously, phenomenological primitives [21] are resources that help students reason about physical mechanisms. The common example "closer means stronger" is built from many common experiences with heat, sound, and light and can be used to guide student thinking about novel situations.

The resources perspective informs pedagogy and curriculum development by urging educators to consider what resources students bring to the classroom, and how those resources might help them think and reason productively about the topic at hand [1,22]. The resources framework also explains variability and context dependence of student responses, as different resources are activated under different circumstances. For example, a student in a class for their major might behave as though knowledge is constructed, but in a class that they find lacking personal

relevance, they may behave as though knowledge is purely transmitted.

The resources framework is only one of several different frameworks that describes the possible forms of a personal epistemology. In addition to resources, Louca *et al.* [3] discuss two other possible forms for describing personal epistemology: beliefs (which are "comparatively stable, robust, cognitive structures corresponding to articulate, declarative knowledge"), and developmental stages (analogous to Piagetian stages). One significant difference between these different forms of epistemology is stability: beliefs and stages are characterized as stable structures, while resources are context dependent and variable. For example, a teacher can prime students with a more productive epistemology with a single statement: "start with what you know" [23]. These differences in epistemological form lead to significant differences in research methodology and teaching practice [3].

At this point, we fully define "pragmatic epistemology" as the resource-based description of student ideas about knowledge, knowing, and learning as they practice authentic scientific inquiry in an academic setting. This differs from Sandoval's "practical epistemology" in two ways: including ideas about learning and being resource based (as opposed to beliefs based). Respecting Sandoval's call for clarity of definitions [16], a new name is called for.

We end this section with one final useful theoretical construct: epistemological frames. PER has adapted frames from sociolinguistics and discourse analysis [24]. Frames are an answer to the question, "what's going on here?" and help individuals decide what to pay attention to and what ideas and experiences to draw on. For example, two possible frames in a physics classroom are "we are here just to get a grade" and "we are here to improve our understanding of physics." Epistemological frames are important within the resources framework because of the variability of resources. Frames help researchers understand why resources are primed in one instance and not in another.

III. CURRENT METHODOLOGIES TO PROBE EPISTEMOLOGY IN PER

Across the PER epistemology research community there are a few major research paradigms: epistemological beliefs determined through surveys, epistemological frames defined through observation, and epistemological resources identified through interviews and observation. A few key ideas for comparing these paradigms are validity, reliability, and methodological implications. In this paper we will focus on validity and reliability to motivate our development of a different methodology and touch on methodological implications throughout our discussion.

In qualitative research, there are many possible approaches for assessing validity and reliability [25]. For the purposes of our research, validity comes in two major flavors: contextual (does the data come from authentic classroom learning activities or research interventions?) and interpretive (would the participant agree with the interpretation or is the interpretation the researcher's construction?). Reliability for this analysis also comes in two flavors: methodological (given the same students and methodology, could other researchers obtain the same data?) and analytical (given the same student data, would another researcher provide a similar interpretation?).

A. Surveys and beliefs

Within PER there are at least three significant surveys that probe epistemological beliefs: Maryland Physics Expectations Survey (MPEX) [14], Views About Science Survey (VASS) [26], and Colorado Learning Attitudes about Science Survey (C-LASS) [13]. As surveys, each of these instruments has weak claims of contextual validity as they do not assess students in their actual classroom activities; however, the interpretive validity, methodological reliability, and analytical reliability claims are strong. The clearest example for these claims is the C-LASS. To address interpretive validity, the development of the C-LASS involved iterative interviews. As a survey, methodological reliability is built into the absence of direct researcher involvement in data collection, and the analytical reliability is controlled in a similar fashion with a provided statistical analysis package. While this survey clearly addresses three of the issues quite well, the final issue—contextual validity—remains significant, as studies have shown that there is often a difference between actions and self-reports of actions [3].

B. Observations and frames

Epistemological frames, defined in the last section, are applied to videotaped student discussions through observational protocols which focus on behavioral clusters. Because these protocols are applied to actual classroom activities, they are contextually valid; however, the missing interaction between the researchers and participants means that the frames identified by the researchers may not agree with what the students believe they are doing. A further concern is that by being based entirely on observation of behavioral clusters, framing research may not penetrate the sphere of personal epistemology. In fact, an open question regarding frames is whether or not the participants' personal epistemologies align with the group epistemological frame, although it is clear that their outward behavior indicates that this is true [27]. In terms of reliability, frames research is highly methodologically and analytically reliable [28].

C. Observations, interviews, and resources

Epistemological resources round out the field of PER epistemological research constructs, and are the focus of

this research. The work of Hammer, Elby, and colleagues argues that epistemology influences learning [4,29] and that student epistemologies take the form of resources (as opposed to beliefs or developmental stages) [30]. These claims are backed by data that comes from videos of class work and interviews. The use of classroom videos gives strong claims to contextual validity. Interviews [31] that focused on topics tied to the students' classroom experience also have strong claims to contextual validity.

However, this work does not uniformly have the interpretive validity that we seek. In Hammer's early work [31], the final interview for each student probed students' epistemologies directly, using (where possible) students' earlier comments about class activities to frame the questions. This direct questioning based on artifacts allows for claims of interpretive and contextual validity and so has some essential overlap with the interview method we describe in the next section. In later work, the researchers make the conscious choice to not ask students directly about their epistemologies for two reasons [3]. First, resources are often tacit and unspoken [14,32–34]. Second, there is a documented gulf between what individuals, students, and teachers report as their personal scientific epistemology and how they actually work with scientific knowledge [35]. We will justify in Sec. IV A how we deal with these concerns.

Another part of this work suggests likely epistemological resources [2] analogous to diSessa's phenomenological primitives [21]. Examples include "knowledge as propagated stuff," "knowledge as free creation," and "knowledge as fabricated stuff." To back up these claims, they take examples from everyday experience. For example, they quote a plausible conversation with a child: [2] "How do you know your doll's name is Ann?" "I made it up!" From student reflections in an inquiry class with a strong epistemological focus, they also find evidence of *e* resources such as "shopping for ideas," "reconciliation," and "looking for consistency" [2]. Because of the different data sources, the contextual validity for these claims is mixed.

IV. STIMULATED RECALL INTERVIEW PROTOCOLS AND RESOURCES

Epistemological resources as defined thus far are a wonderful theoretical tool. We propose to extend this work in three key ways. First, we access the rich, privileged data that students themselves can provide on how they approach knowledge in our classroom activities. This was suggested by Sandoval [10] but to date has not been probed due to concerns outlined above. Second, we give more detail and contextual validity to the existing list of *e* resources. The detail we seek includes the following: What are the ways in which knowledge is transmitted? What underlies a doubting stance; is it a lack of trust or justification? Last, we address the reliability of epistemological resource work,

which has not yet been addressed. The stimulated recall interview is at the core of all of our work.

We describe our process in four different sections. In this section, we describe the stimulated recall interview protocol and argue that it makes significant progress in the areas of contextual and interpretive validity and reliability. In Sec. V we discuss the pedagogy which was the locus of our study and therefore informed our methodology. In Sec. VI, we discuss the methodology of this study, which combines the SRI method, the resources framework, the method of data analysis, and the research questions to give a consistent and coherent approach to the problem. Last, in Sec. VII we give details of the implementation of the methodology to show how the resources framework and our research question informed our decisions. We believe that this level of detail is required in order to allow for use and refinement of this new methodology.

A. Validity of SRI

Accessing students' ideas about knowledge, knowing, and learning (their pragmatic epistemology) necessitates gaining insight about their thinking that is not readily apparent through direct observation. This transition from behavioral observation to cognitive observation is undertaken in countless studies where the fundamental approach is always the same: to get participants to verbalize their thought processes. Two major approaches that facilitate this transition are process tracing or think aloud protocols (TAP) and stimulated recall interviews.

Think aloud protocols are interviews in which participants are asked to verbalize their thought process in parallel with the target activity. Stimulated recall interviews are a two-step process: researchers record the target activity first, then develop an interview protocol that allows the participant to reflect on their primary experience with the benefit of the original recording. In this section, we discuss the pros and cons of these methods for our purposes, and justify our development of a methodology based on the SRI method.

An alternative approach to achieve interpretive validity is "member checking," which involves bringing the analysis back to the participants for verification that the data they provided are not misinterpreted [36]. The advantage of SRI and TAP is that the initial interpretation is informed by the participants' reflection. The participants are able to reflect on their experience and interact with the researcher through direct conversation rather than the researcher making final inferences from a observer's perspective alone.

There are well-articulated issues when accessing thought processes through verbalization that are common to both TAP and SRI. The fundamental issue is the problem of verbalizing tacit knowledge or automatic mental processes [8,9,37]. When a researcher asks a participant to explain a decision or thought process that they engage in without explicit thinking, they may generate an explanation on the

spot. The discussion in the research community of these mental processes is ongoing. In this research we assume that the explanation of tacit knowledge and automatic processes is a "good faith effort" whereby participants are giving what insight they can under the conditions of the interview. Even under such an assumption there is the possibility that participants are defaulting to *a priori* theories they have about how they think, how they would like themselves to think, or how they believe the researchers would like them to think [38]. It is argued that such theory confirmation would be sensitive to the complexity of the tacit thought that the participant is explaining; however, such a distinction is beyond the scope of our research, as we accept that even theory confirmation would still provide data relevant to the epistemology of our participants.

In the face of these real concerns about what researchers are accessing in SRI, we present in Table I two non-sequential clips of an SRI interview to argue that this student, Julia, is likely accessing memories and not creating stories on the spot. She is reflecting on a lab about stress and strain where increasingly large masses are hung from an extensible cord. We note that Julia's responses are very detailed, making references to confusions, expectations, mathematical representations, and experimental results. In addition, she uses everyday language and is frank in her confusions, giving some evidence that she is comfortable in the interview and unafraid to speak honestly. Last, she also makes a distinction during the clip, starting at 40:44.9, about what she knew at the time of the lab and what she learned in the week following the lab. This indicates to us that her recollections of the sequence of events in the week before this interview are somewhat detailed and accurate.

In choosing between TAP and SRI, the contextual validity of SRI is a major deciding factor. In TAP, participants are more often than not working with simulations of the activity that the researchers are interested in rather than undertaking the activity in a natural setting [39]. With SRI, the participants go about the authentic target experience (in our case a physics lab) as they would any other time, except with researchers making audiovisual passively. Two concerns arise from the fact that participants in SRI interact with recordings of their first-hand experience from a third person perspective. The first is that the audiovisual recordings are not in fact recordings of their experience, but of a unique observer perspective. This raises the question of whether participants are reflecting on their first-hand experience or on the experience of viewing the recordings. The second concern is the participants' physiological response to the alien act of self-observation. Some participants in SRI report anxiety and distraction while watching themselves, which brings into question their focus and ability to reflect accurately on the first-hand experience that is the target of the SRI. Although we noticed distraction in interviews, students were still able to engage with the interview questions after acclimating to the foreign experience.

Despite these concerns, one can argue that SRI provides access to otherwise inaccessible or "privileged" data in two distinct ways. First, in a comparison of free recall versus stimulated recall of personal experience, it has been shown that stimulated recall significantly improves the degree and volume of experience that can be recalled [40]. This direct result is essential to the bolstering of SRI as a contextually valid methodology. To take full advantage of improved recall with SRI, it is important that the interview take place as soon as possible after the original experience [9,41]. Second, even with the direct access of first-person audiovisual recordings or physiological recordings, the

TABLE I.   Recall and reflection in a SRI.

Interview time	Transcript
0:06:26.3	CWS: Okay, so there, you and your partner are just? feeling out how to apply a free body diagram to the actual experiment, so can you? walk me through what you were thinking in that sequence?
0:06:37.9	Julia: Um, at that point I still wasn't very good with forces, so I was just trying to like, think of everything. They told us to put everything in the equation, the Fnet equals m a. So I was trying to think of where all the different parts would be included, and I was getting hung up on the fact that weight equals m a- er, m g. And I knew that gravity was acceleration, so… [shrug] I was kind of hung up on that?
0:39:53.9	CWS: Alright. Um… And um, in the third clip there that we watched, the—I think it's that the- the data started to kind of really, um, change in direction, and you commented that you expected it to kind of go straight throughout. So, um, do you remember kind of what you were thinking about through that process of y'know you see the actual data trend different than what you expected, and what kind of a reconciliation or anything like that you go through in that?
0:40:44.9	Julia: Um, I was just thinking that if it's the same throughout that the stretch would be the same. So it would be like a linear stretch if you add weight it's going to stretch a little more. If you add the same amount of weight it would make it stretch that much more. But um, from what we've talked about in class since then it makes more sense that it would be curved and not just linear. Um, but at the time we hadn't dealt with materials yet so I was just trying to get a feel for how it would react and if I added a lot of weight like would it reach a point where it only stretches a little each time I add more weight.

researcher misses the essential context of the interpretive framework and personal history of the participant; some of this context and framework is reclaimed through SRI. Combining the exclusivity of access with enhanced recall, SRI stands out as the appropriate method with which to design our research methodology.

B. Reliability of SRI

Reliability is the claim that an independent researcher using the same methodology will produce similar results. There are two places in the SRI protocol in which researcher choices are most in question: (i) the selection of recall artifacts and questions to use in the SRI, and (ii) the coding of the SRI transcript. As with other methods, interrater reliability is used [36] to give a measure of the reproducibility of the method. We discuss our implementation of interrater reliability in Sec. VII.

C. Overview of SRI

We conclude this section by taking a wider view, to examine how SRI fits into the field of studying personal scientific epistemologies, and how it compares to other methods discussed in Sec. III.

SRI allows us to gain detailed first person accounts tied directly to participants' actions that are captured with some degree of faithfulness by artifacts. This makes it good for exploratory studies, where researchers do not yet know what epistemological ideas are at play and would profit from the detail and privileged information that SRI gives. However, because of the detail produced and time required, it is not a useful method in a more mature research program or where large N data is needed.

For example, SRI could have been used in the creation of questions and distractors for the C-LASS assessment of students' beliefs, to gain access to students' perspectives of how they learn and what they know and give claims of contextual and interpretive validity. However, the end goal of the C-LASS is an assessment that can be given by many researchers and instructors to many students, for which purpose a survey is far more appropriate.

SRI also would not have been necessary in the identification of epistemological frames of Scherr and Hammer [27] because these frames were very reliably identified through observation alone. However, there were details of student thinking that were not available through observation, and they could only guess at the students' meaning. For example, Hannah says "I hate the word intuitively," and they note "We may only speculate about the reasons for her distaste" [27] (page 169). In this study observations and SRI would likely give complimentary information: observations accessed the frames that the students themselves would perhaps not be aware of, and SRI would have given insight into student reasoning independent of the identified frame.

In conclusion, SRI is a valuable new tool in the study of personal epistemologies that complements existing methods by adding interpretive validity and detail, although at the cost of significant time.

V. MODELING INFORMED INSTRUCTION

Before describing the methodology, it is essential to give details of the pedagogy studied, as these details will inform many steps in the SRI protocol. In choosing Modeling Instruction as the starting point for the reform of the course, we commit to the underlying pedagogical and epistemological structure of Modeling Instruction. A central aspect of Modeling Instruction is teaching a set of core scientific models (e.g., constant velocity, constant force) that students create and apply by following the modeling cycle [7] as shown in Table II. The modeling cycle is broken down into two stages, model development and model deployment, each with their own individual phases. In our adaptation of Modeling Instruction, we use the majority of our course's laboratory meetings for all student-centered model development activities (some models are developed in interactive lecture demonstrations), and we use the remaining lecture and laboratory meetings for model deployment activities.

In the model development activities students engage in the first stage of the modeling cycle by creating a model of a natural phenomenon through an empirical investigation devised with their critical input. The model development activities written for this course are referred to as Modeling Informed Instruction (MII) and follow the adapted structure outlined in Table II; this structure will be referred to later in the interview development process and analysis. Within the activities, the lab guides deliberately state the purpose of each section and the prompts carry epistemological messages; three examples are shown in Table III. The rest of this research is concerned with enacted epistemology in these activities, specifically how to access the privileged data of how students engage in the design, undertaking, and

TABLE II. A comparison of the steps in the modeling cycle [7] to the steps in our Modeling Informed Instruction.

Modeling instruction	MII
Qualitative ideas	Prior concepts
Identify variables	Preliminary model
Plan experiment	Relationships and plan experiment
Data acquisition	Execution and data collection
Analysis of experiment	Construct representations of data
Presentation of experimental results	(no explicit parallel)
Presentation of models and peer evaluation	Presentation of models and peer evaluation

TABLE III. Examples of epistemological messages in MII.

Prior concepts and models	Purpose: The concepts and models that are developed throughout the course are built upon each week. Reviewing your scientific understanding from previous activities will help you keep everything coherent and consistent.
Qualitative exploration	Purpose: This section gives you primary experience with the phenomenon you will model in this activity. Take the time to observe carefully, what you notice now will play an important role in how you develop your model.
Connect to prior models	Looking at the phenomenon so far you might see how static situations like the hanging cord are just equilibrium applications of the Newton's second law model.

analysis of experiments to create scientific models of natural phenomena.

VI. METHODOLOGY

We are now in a position to state the goal of our methodological work more precisely: to design a methodology that gains access to students' privileged information about their own pragmatic epistemologies in the classroom. The methodology presented in this section, which combines the resources framework and SRI, is our answer to the issues of contextual validity, methodological reliability, and interpretive validity. It is our hope that if this methodology can be well enough articulated and disseminated, then claims for methodological reliability can be strengthened and claims of analytical reliability can be proposed.

A. Methodological influences

Our research methodology design gives us access to data that addresses our specific research goals. The design has three major influences: modeling theories of science, pragmatic epistemology, and grounded theory.

The data used to describe students' pragmatic epistemologies come from student engagement in the MII activities discussed earlier. The design choices made in creating MII, especially the underlying philosophy of modeling theories of science, influenced the results of this study as they are a primary source of epistemological messages transmitted to the students. In the terminology of qualitative inquiry, the locus of this study must be put in terms of the deep integration of the research within the overarching course reform project, the MII activities, and the interview process [42]. Identifying the locus in this way is an essential aspect of developing the credibility of the results, because the data upon which the results are based is a product of this complex network of influences.

The end result of this research is the identification of pragmatic epistemological resources based on privileged knowledge of student personal epistemology. With the explicit locus identified, we can move on to discuss the focus of this study: the engagement and approach that students take to the MII activities as modeled through the theoretical lens of pragmatic epistemology. This lens is the second major influence to the methodology of this study; it

defines the way in which we attend to the data as it is coded. Since the coding of the data is the substrate upon which the analysis of student epistemology takes place, the effect of this theoretical lens is central to the claims we make about the identification of epistemological resources. A key point to reiterate here is that the description of pragmatic epistemology is not intended to reflect the actual epistemological structures in the mind that give rise to behavior, but to identify effective structures, those that we can describe within our framework to explain student learning behavior as determined by the structures of the mind. We do argue later that these structures have reasonable claims to validity due to our methodological choices.

Finally, the overall approach to developing and identifying pragmatic epistemological resources from the SRIs is most directly influenced by grounded theory [42,43]. There are several essential aspects of the grounded theory tradition, and this research does not authentically engage with all; however, the tradition of grounded theory acts here as a concrete reference to clarify our methodological choices. The key aspects of grounded theory to be discussed are as follows: an atheoretical approach to data analysis (one should not have a preconceived theory that they wish to apply to their data, a grounded theory comes out of exploratory data analysis—our approach has the expected structure of resources, but the resources themselves and their organization is left to be discovered), parallel data gathering, and analysis (one should begin analyzing data as soon as it exists and continue collecting data concurrently as the theory develops), intentional data selection (one should choose to take new data when and where it will help complete the developing theory), two-stage coding through constant comparison (one should first "open code" their data, and then refine those codes into "focused codes" by comparing the open codes across the body of data). Each key aspect of grounded theory that plays a role in the methodological design and is touched upon in the next section.

VII. IMPLEMENTING STIMULATED RECALL INTERVIEW PROTOCOL

The process of designing a stimulated recall interview methodology based on our research goals and questions is

TABLE IV.   Steps in our methodology.

Step in process	Goals
Group and individual selection	Identify students willing to opt into both classroom taping and quick followup interview.
Record natural classroom videos (NCV)	This is an opt-in process; researcher plays a minimal role.
Describe NCV	Description of student actions to gain overview and facilitate clip selection process for SRI.
Select groups for SRI	Choose lab groups to invite to SRI based on richness of discussion and availability of students for SRI.
Select artifacts for SRI	Select clips that highlight key epistemological decisions.
Conduct SRIs	Ask students to reflect on class activities with NCV clips and worksheets to stimulate memory.
Transcribe SRIs	Transcription focused on coherent phrases.
Code all SRI clips	Use key words to code SRI transcripts. Use grounded theory approach to develop and prune key words.
Generate quantitative analysis	Using key words developed in the last step, find correlations between different codes.

explained in this section and outlined in Table IV. We present a walk through of a complete data acquisition and analysis sequence, with a detailed discussion of crucial choices made and the effect of these choices on the research as a whole. We provide details of the interrater reliability measure for one interview. Finally, we summarize the claim that our methodology uniquely provides a reliable and valid approach to epistemological resource identification.

A. Implementing a stimulated recall interview protocol

In this section we lay out the entire implementation process, from selecting video taping groups during MII activities to applying and analyzing pragmatic epistemological codes. Along the way, we discuss the concerns that arose, how they were addressed, and the effects of the resulting choices in a comprehensive manner.

1. Obtaining natural classroom video

We want to understand student pragmatic epistemologies, that is, the epistemologies that they invoke while working on classroom activities; therefore, we must videotape students as they work in the classroom. The recording of groups in their usual classroom activities is opt-out; however, the SRI falls under a separate opt-in decision which creates a self-selection bias in this study. Students who do not opt-in for interviews are still videotaped in class, and their interaction with the other group members can be discussed in the interviews.

Individuals are assigned to groups at random, and these groups are assigned to tables within the classroom at random; however, all videotaping occurs at tables located in the back of the classroom so that the video camera and microphone setup are as unobtrusive as possible within the classroom. We do not believe this affects our results.

The final selection protocol is to video on only one day per week. We choose one group from each of three to four lab sections during that day to observe, generating three or four natural classroom videos (NCVs) per week. This

protocol is put into place for practical reasons based on the size of the research team and the amount of data looked at during a single week. This protocol allows for the sampling of a large variety of individuals and groups without having a constant presence in the classroom.

2. Describing natural classroom video

Once the natural classroom videos are recorded we import them into Transana, the video transcription and analysis software we use for all video work in this study. We then watch the video and write a noninterpretive description with time stamps at least every 2 min within Transana. Noninterpretive in this context means that the researcher's impression of student activity is not included. We describe students as "reacting," as opposed to "surprised" or "confused" by their observations. The purpose of this description is to provide a baseline account of the activity in the video for the forthcoming clip selection process.

3. Selecting groups for reflective interviews

Three or four NCVs are described per week to get a sense of the richness of the conversation for each group. In an effort to keep the timeliness of the SRI method in check, only one group per week is fully prepared for the SRI. This in turn leads to one to three individual interviews, depending on how many group members agree to be interviewed. The time between the NCV and the interview is ideally kept to approximately one week.

The selection of groups to interview introduces some biases. The choice of which group NCV to prepare for interviews depends on two major factors: how many group members are willing to participate in the individual reflective interviews, and how rich the student interactions are in the NCV. The number of group members willing to participate in interviews is again an instance of self-selection bias; however, the ability to access multiple individual student perspectives on common clips of group activity is decidedly unique and a strength of this

methodology as it may shine new light on the concept of epistemological frames [27]. The richness of NCV depends on factors such as the amount of talking that group members do, the physical engagement of the group members in the activity, and the general energy level of the group members. These factors are important because the goal of this methodology is to access student personal epistemologies. Although quiet groups would give valid data on pragmatic epistemology, the use of video data in this methodology is such that more expressive groups yield more easily accessible and illustrative data for its descriptive goal.

4. Selecting video clips for interviews

In order to create SRIs, the clips for the interview need to be selected. The selection of NCV clips for a SRI is a process based on two goals: the investigation of the connection between the MII activity design and student pragmatic epistemology, and the identification of student pragmatic epistemological resources.

In order to identify the connection between MII activity design and student pragmatic epistemology, the selection of NCV clips needs to be based on the MII design. For example, every MII model development activity includes a section on variable identification, intended to establish an open and constructive approach to experimental design; therefore, we seek video clips that encompass some of these discussions to uncover student approaches to this section in particular. The sections are shown in Table II.

The NCV is given two layers of codes. The first layer of codes for the group activity is based on where the group is in the MII activity itself. A second layer codes for common but peripheral factors that affect student engagement. For example, TA interactions, interactions with other groups, and off-topic discussions are all given codes. These peripheral codes are similar to epistemological framing behavioral codes; however, because the major focus of this methodology is on the MII design, the focus in these instances is on the divergence of behavior away from the activity as opposed to following epistemological frame coding guidelines explicitly [28]. This approach leads to a relatively consistent set of NCV clips across the interviews. The common selections from the MII activity structure are as follows: initial exploration of setup, variable identification, experiment planning, data taking, and data analysis. Focusing the clips in this way is a key point of hybridization between grounded theory and our research goals. Directed data selection is traditionally done based on what is being found in the data [42,43], whereas here we base the data selection on the MII activity context. However, the selection of certain clips based on epistemological salience, such as group discussions or spoken reflections, opens up this research to the type of insight that grounded theory is designed to uncover [42].

5. Preparing stimulated recall interview protocols

Once we identify clips as described above, we use Transana to create the interview protocol. For each clip, we jot down the key reasons for selection and formulate one or more questions to ask participants during the interview. These reasons and questions are essential for the interrater reliability (Sec. VII B) and are limited in kind. For example, clip selections frequently focus on students generating explanations. The interview questions for these clips probe the backing for the explanation or the role of a partner's explanation in the interviewees thinking.

In constructing these questions our focus is on an epistemological aspect of the clip. For instance, if the group latches on to an idea that one of the participants brings forward, we might ask the participant that came up with the idea where that idea came from. This question balances the epistemological target and the good practice of open ended interview questions. In circumstances where multiple group members are interviewed individually, we may prepare different questions for the same clips based on the different individuals' roles in the interaction. For example, in a clip where we ask one participant where their idea comes from, we might ask their partners, in separate interviews, how they engage with or evaluate their peer's idea.

Initially, we structure the interview in the same order as the activity itself: starting with the exploration, and ending with the data analysis and discussion. However, interviews are key researcher-participant interactions with the goals of establishing rapport and working toward a good interview experience for both the participant and interviewer. In light of these goals, some interviews jump around to clips related to ideas that the participant brings up. We use this technique to try to build depth of understanding on specific topics when the opportunity arises [44].

6. Performing stimulated recall interviews

We perform the individual interviews in a group study room with a television and speakers, which allows both the participant and interviewer to watch the video clips comfortably. A large table gives the participant room to look at both a printed copy of the lab activity and their group's work. The interviews are approximately 1 h in length and follow the protocol described above. Before the interview begins, the researcher reiterates the informed consent guidelines.

The researcher introduces each clip by describing from where in the activity the clip was pulled and by giving a general overview of what will be seen. After the researcher and participant watch the clip, the researcher asks one of the prepared questions. Depending on how the participant responds, the researcher either asks a follow-up question, asks another prepared question, or iterates the process with a new clip. As mentioned in the last section, the prepared

TABLE V.   The nested structure of analytical clipping and coding in SRI analysis.

NCV sequence clip (MII-Design Codes)				
SRI question response clip (Question Intent Codes)			SRI question response clip (Question Intent Codes)	
Question	Response clip (Response and Epistemological Interpretation Codes)		Question	Response clip (Response and Epistemological Interpretation Codes)

order of the interview may deviate depending on the participant responses.

7. Transcribing SRIs

Transcribing the interviews is performed within Transana. Each interview is transcribed with the focus on narrative responses. No extra efforts are taken to attend to paraverbal and nonverbal cues. We make this choice because the focus of this methodology is on informing the interpretation of the activity in the NCV. However, when a student's nonverbal communication is essential to the narrative, we note it. For example, when a participant says "like this" and makes a hand motion, that hand motion is recorded. By transcribing for narrative, insights from details of the interview may be missed. These insights may be valuable, but are seated in the context of the interview and not the NCV, where this analysis is focused.

8. Coding SRIs for pragmatic epistemological resource application

Once we transcribe the interviews into Transana, the focus shifts to analyzing and interpreting the impact of the interviews to understand student pragmatic epistemology. Transana requires that we base the analysis on applying a keyword coding to selections of the interview transcript correlated to the interview video. Each NCV clip along with the questions and responses that follow are selected into a large *sequence clip*. For example, a sequence clip in the interview with Julia on the extensible cord was the group considering how to draw a free body diagram for that situation. For each NCV sequence clip there can be one or more questions and responses; each question and response based on the NCV clip is selected into a smaller *question-response clip*. Finally, each participant response is selected into a *response clip* by itself. This provides a nested analytical structure for applying key words to these clips and allows a different set of codes for each kind of clip, as shown in Table V.

9. Analyzing pragmatic epistemological codes

The analysis of the code applications comes in two flavors: qualitative code reduction and quantitative code correlation. These two analyses each target one of the research goals of this study. The qualitative code reduction

gives rise to a set of Epistemological-Interpretation Codes, which is a limited catalog of pragmatic epistemological resources. The quantitative code correlation between MII-Design Codes and Epistemological-Interpretation Codes is the basis of evidence-based claims about the pragmatic epistemologies of students during MII model development activities. Such correlation could point towards the effectiveness of inquiry based instruction at developing an inquiry oriented epistemology in students. For example, in the "preliminary model" stage of MII, we find that "knowledge construction by physical observation" is the most common e resource used. Since this is also an appropriate resource, we have some evidence that the MII design is working as intended.

Qualitative code reduction follows the principle of constant comparison from grounded theory mentioned earlier. The number of unique codes applied to the interviews grows during the initial interview analyses, then the number levels off as the keywords defined saturate the space of responses. Finally, the number is reduced by elimination of redundant codes, which requires careful reflection and examination of current codes.

Quantitative code correlation is done via a simple algorithm implemented in a script. The algorithm operates on a "Clip Keyword Data Export" from Transana. The algorithm identifies every keyword that we apply concurrently through the nested structure of the clips. This allows us to associate the MII-Design Code of the NCV sequence clip with all Epistemological-Interpretation Codes for the nested response clips. The result of this algorithm is a symmetric correlation table with a row and column for each keyword defined in the system. From this table the frequency of a single key word is found on the diagonal, and the off-diagonal values give the number of times an "MII-Design Code" and an "Epistemological Interpretation Codes" appear in nested clips. Other off-diagonal values also give correlations between two Epistemological-Interpretation Codes in the same response clip, but we did not use this data.

B. Interrater reliability of the stimulated recall interview protocol

In this section, we describe the design and implementation of an interrater reliability measure for the creation of stimulated recall interview protocols. As noted earlier, this

reliability measure is designed to explore the question of whether independent researchers, with a common set of guidelines, can select similar clips of NCV for similar reasons, and propose similar questions to be asked in an interview. This measure is an important first step to show that this methodology could be replicated in studies performed by larger research groups. The second issue of reliability (interview coding for the production of pragmatic epistemological resources) is not addressed in this study. This means that even though we will show that preparing SRIs to access the privileged information described above is a reliable process, the analysis of the interviews themselves is currently dependent on the individual researcher. The codes that are applied to the interviews are defined, but there is no independent coding as of this writing.

1. Defining interrater reliability

As it is practiced throughout the PER community, interrater reliability (IRR) measures whether or not a research practice can be replicated among individual researchers, usually within the same research group [28]. The standard approach to IRR has five steps: (i) a research practice is defined on an initial set of data by one or more researchers; (ii) this definition is communicated to researchers that have not worked on the data; (iii) two or more researchers implement the practice independently on a new set of data; (iv) the initial implementations are compared; and finally, (v) the discrepancies in the implementations are discussed, allowing the researchers to come to an agreement if there is initial disagreement and a final measure of IRR is calculated [27].

The key concerns for implementing this IRR algorithm are steps (ii) and (v). In step (ii), the researchers that developed the practice must communicate the practice in an efficient and effective manner; this is sometimes done using a data set for training, where researchers may enter a master-apprentice power relationship. In step (v), the discussion of initial disagreements and the transition to agreement is a process that has the potential to degrade the measure if one researcher consistently yields to the other because of an underlying power dynamic.

2. The SRI IRR measure

In this study, the IRR measure covers two major steps of the research practice: the selection of NCV clips and the preparation of questions to be asked concerning those clips.

Step (i) of the IRR the development of the research practice is thoroughly presented above.

Step (ii) of the IRR requires describing the research practice to the independent practitioner. The key aspects of the protocol described are identical to those laid out in the section of this paper covering clip selection.

Step (iii) of the IRR requires each researcher to produce the following: the start and end times of their NCV clip selections, their reasons for selecting that section of NCV, and one or more epistemologically oriented questions to be asked of the participants based on the NCV clip. The first two pieces from this step are shown in Table VI.

Step (iv) of the IRR requires that the selections, reasonings, and questions are compared independently to compute initial IRR values. These three comparisons are made as follows: Do the clip selections overlap with any selections made by the other researcher? If so, are the reasons for selection based on the same key aspects of the clip? If so, do the questions that are proposed share a common purpose?

Step (v) of the IRR process is a joint venture between both researchers to identify the perceived gaps between their decisions, to discuss their stances, and to determine if the perceived gaps are true discrepancies within their practice. This process is a lengthy iterative discussion that takes on each discrepancy anew and requires clear articulation of not just the researchers' choices, but why they

TABLE VI. Interrater reliability reasons for clip selection after step three, before discussion.

Clip Selection Time	Reasoning CWS	Reasoning DCM
16:30.5 19:10.6	Exploring equipment; discussing resistor vs capacitor; do wire colors matter; does order of circuit matter	Focus on color of wire as possible issue
20:41.4 20:57.5	Failed current probing; reaction to TA giving information	(Clip not selected by DCM)
24:01.0 24:56.0	Discuss dead battery; equipment troubleshooting; confirming what told or variation; lead to TA help	(20:58 through 24:06 +) interaction with equipment and computer; confusion; playing around; trying to make sense
27:20 27:55	(Clip not selected by CWS)	Sarah checks leads verifies TA information
28:33.1 29:34.7	Charge model; confusion and skipping;	(28:51 29:30) Reversal of signs from worksheet; skip it (charge model)
30:16.3 31:35.6	ID Variables prompt; use of equipment and handout to determine experiment?; "modeling setup"	(29:30-34:44) interaction with sheet; ID variables; use of blanks

made their choices. For instance, we find that separating out the intent of each question (i.e., the idea which the researchers hope to uncover through the question) from the wording of the question gives better insight into whether the researchers are in agreement.

3. Results of the IRR

The above IRR is performed on two NCV episodes. The first episode is treated as a training data set, where we only analyze a portion of the two-hour video. In this process, the researchers find that epistemological notions such as "students appear confused" could be too interpretive to be reliable between researchers and reiterates the need to focus on concrete actions and cues from the activity guide.

After the training session, the researchers choose a NCV clip for IRR for which CWS has already performed interviews with all three group members. This choice allows DCM to specifically identify which participant she would like to ask a question of, or if she would like to pose multiple questions to different individuals, comparable to CWS's process of preparing for three independent interviews.

The results of steps (iv) and (v) are shown in Table VII. The "Selection Reason Matches" metric in step four is calculated as the ratio of the number of clips for which the reason selected matched to the number of clips for which the selected NCV time overlapped. In step (v), however, the total number of overlapping clips increases by 5 based on discussion between the researchers about their time selections. The "Question Intent Matches" metrics are the ratio of the number of clips where the indicated question intent matched compared to the number of clips where the selection reason matched. For example, in a clip where a group is determining what function will give them the best fit for their data, both DCM and CWS posit questions that target how the students evaluate the fit options and make a decision as a group; the question intent code focuses on the decision making process of the group. Increases from step (iv) to step (v) are again due to discussion.

From these results, we claim that it would be reasonable to aggregate data from interview protocols developed by either of these researchers. In a large-scale qualitative study, practical issues such as this become important; this result sets a precedent for evaluating an SRI for data aggregation,

and is the first step to moving classroom video analysis beyond behavioral coding.

VIII. ANALYSIS

The analysis of the interview data is briefly described in the previous section, and here we expand on the process by which epistemological coding is applied to student responses.

A. Code application to interview data

When we apply codes to the interview data, the interview transcript is already organized as nested clips described in the previous section. The structure of these clips is an important methodological decision point. The interview itself is broken down into question-answer pairs, which are then broken down into the question itself as asked by the interviewer and the participant response. The participant response is always clipped as a single analytic unit; a one word response and a 1 min response are treated as the same size *response clip*. The codes are applied to a single clip; therefore, single responses can have a large number of codes applied to them. Applying many codes to one clip still allows pragmatic epistemological resources to be identified.

In a trial of breaking responses down by sentences or coherent phrases, the volume of analytical clips was overwhelming. Codes are applied to a single clip, which creates difficulty where codes are needed to span clips.

There are two major types of codes that we apply to the interview *response clips*. The first is the *response code*, which is a set of descriptive codes developed from the interview data to identify the coherent phrasing of the responses themselves (see Table VIII for examples). These codes serve a secondary purpose beyond describing the essential elements of the student response. They also allow the researcher to "code everything." This descriptive layer of codes does not necessarily focus on epistemology, but instead gives the researcher a way to analytically notice details of responses without conflating them as epistemological. As *response codes* are descriptive, it is essential to maintain a catalog as the research progresses, and to define each code clearly when it is first applied. These definitions are constantly referred to in order to make decisions about whether or not a new response code is needed, or if a prior code is appropriate.

The second major type of code we apply to the interview response clips is the epistemological interpretation code. These are interpretive codes, reflecting the essential relationship between the researcher and the data. The researcher cannot be excluded from the discussion of these codes. We make a second coding pass of the interview, and in this process we evaluate the statements of the participants to determine whether or not they might indicate something about the participant's pragmatic epistemology.

TABLE VII.   Interrater reliability results.

Metric	Prediscussion (Step 4)	Postdiscussion (Step 5)
Selection reason matches	85% (11/13)	94% (17/18)
Question intent matches (All)	82% (9/11)	82% (14/17)
Question intent matches (Any)	91% (10/11)	96% (16/17)

TABLE VIII.   Response and Epistemological Interpretation Codes applied to a response clip.

Transcript	Response Code	Epistemological Interpretation Code
So, we were trying to figure out what we were going to be measuring	Define-Variables, Deciding through discussion	Knowledge-type-decision
What would be the constant variables, versus like the- what we were changing	Define-Variables, Experiment-Design	Knowledge-type-decision
We knew that we had to find out something about the material, in the end	Goal-Oriented Model	Action-motivation Goal-of-activity, knowledge-type-expectation, knowledge-type-property-of-material
All of our data was going to tell us something	None	Knowledge-source data, knowledge-justification-measurement or data
We knew that the length was changing	Cause-effect-knowledge-from-observation	Knowledge-by-construction-physical-observation, knowledge-type-property-of-material
We knew it was because of the force, and the weight we were adding	Cause-effect-knowledge-from-observation	Knowledge-by-construction-theoretical-concept

The epistemological interpretation codes are first written as open-coding interpretations of the data. These are highly specific statements about the participant response and how it is seen by the researcher as reflecting the participant's epistemology. The researcher's perspective on resources also affects this process, as they keep the codes focused on small productive descriptions. In a similar process to the response coding, these codes are defined each time they are created. As the analysis continues, similarities between responses are identified; the epistemological interpretation codes start to show a structure larger than themselves, and thus begins the discussion of *epistemological aspects* as a higher order organization of pragmatic epistemological resources coming out of this interview data.

B. Epistemological aspects

Once we complete the process of applying both response and epistemological interpretation codes to a few interviews, the explosive growth of codes slows and each subsequent interview sees more code overlap and less code development. The epistemological-interpretation codes at this point have grown out of contextualized responses, are focused explicitly on interpreting student responses in epistemological terms, and avoid overreach by maintaining the role of response codes as describing nonepistemological aspects of student responses.

At this point, we are able to return to our initial goal of assessing the course reforms by identifying student epistemological resources used during lab, determining if these resources are productive, and if they change over the course of the semester. The first step that we can take at this point is to define epistemological aspects—large scale structures of student epistemologies. These aspects, which we describe here, are embedded in the interpretation, and each epistemological interpretation code is rewritten later to identify the associated epistemological aspect.

We do not expect to create every epistemological-interpretation code needed to describe pragmatic epistemology here; however, after a few interviews, the epistemological-interpretation codes have a clear organizational structure that we describe as *epistemological aspects*. These aspects help to integrate our research results with the overarching body of epistemological research. The top levels of these come from our knowledge of epistemological theory, and play a significant role in driving the question-intent codes that describe our interview protocols. The lowest level of these organizational structures are the epistemological-interpretation codes themselves, which we directly interpret from the interview transcripts. The middle layers of the organizational structure connect the two extremes by labeling emergent structure through constant comparison qualitative analysis techniques.

For example, there are several instances where the epistemological-interpretation codes describe an attribute of the knowledge claim, such as the familiarity of an idea. At first, this idea seems to fall outside of the epistemological aspect "stability of knowledge" as it does not explicitly describe a degree of certainty. However, familiarity might define a stage in developing trust, and trust of knowledge underlies stability or certainty. By reflecting on the organization of codes in this manner, we can avoid an explosion of epistemological aspects while making these aspects more robust.

We give here a brief overview of the four epistemological aspects we identify through our research.

1. Source of knowledge

The source of knowledge has two parts: the source object, which is a person or artifact that is the physical source of the knowledge; and the source mechanism, which is the way the knowledge came to the individual. These two subcategories come from theoretical considerations; however, they align with several Epistemological-Interpretation Codes.

2. Utility of knowledge

An important epistemological aspect that is not discussed in other studies is the utility of knowledge. In this study, we find that students have specific purposes or motivations in mind as they work through each activity. Their approaches to prompts in the activity appear to be affected by the ways in which they expect to use their responses or knowledge at a later time. Two subcategories of this aspect are temporal and application. The temporal category answers the question: "Does the student expect to use the ideas they are generating in response to a prompt beyond the prompt itself?" In contrast, the application category answers the question: "How does the student expect to use their ideas? Consistency checking is a key application of ideas that several participants bring up.

3. Stability of knowledge

In the context of this study, resources in the stability epistemological aspect describe a student view of a particular knowledge claim as stable or unstable. We find three major distinctions in the data: knowledge is certain, true or false; knowledge is scoped, it has a distinct range of validity; and knowledge is uncertain, there is limited confidence in the claim. The stability of knowledge must also have some form of justification, and this falls under the same aspect. For example, we see knowledge claims justified by invoking the authority of an individual, an equation, by referring to a mechanism, or explicitly through reference to observations and data.

4. Structure of knowledge

The idea of whether knowledge is isolated or connected to other ideas is significant in our data. We call this the "sophistication" of knowledge. This aspect also contains the various types of knowledge described as epistemic forms, such as facts or processes.

IX. CONCLUSION

This work seeks to describe pragmatic epistemology, a resource-theory based interpretation of student work in classroom laboratory activities. Along the way, we show that SRI allows researchers to access a deeper layer of student data (their recall and reflections on their recorded work) which is the keystone to the arguments of validity and reliability made herein. By developing a methodology based on SRI situated within the MII lab activities focused on describing student pragmatic epistemology, we move resource theory forward down a data driven path. SRI complements existing methods that include surveys, classroom observations, and interviews.

Once the methodology is well studied, we can return to the primary goal of assessing curricular materials and pedagogies from an epistemological perspective with a method that is grounded in the actual process that students undertake in the classroom. For example, we can gain access to the epistemological resources that students are actually using during the labs, investigate which resources seem to be most productive, and if students' epistemologies are becoming more sophisticated over the course of the semester. Alternatively, and perhaps more importantly, the development of a consistent framework of epistemological resources can be undertaken to provide a common language for understanding the development of epistemology within students' scientific education. In either case, there is no doubt that extending our understanding of students' epistemology in scientific education is an important step in promoting the development of the next generation of scientists.

ACKNOWLEDGMENTS

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

[2] D. Hammer and A. Elby, Tapping epistemological resources for learning physics, J. Learn. Sci. **12**, 53 (2003).

[3] L. Louca, A. Elby, D. Hammer, and T. Kagey, Educ. Psychol. **39**, 57 (2004).

[4] L. Lising and A. Elby, Epistemological resources: Applying a new epistemological framework to science instruction, Am. J. Phys. **73**, 372 (2005).

[5] D. C. Meredith and J. A. Bolker, Rounding off the cow: Challenges and successes in an interdisciplinary physics course for life science students, Am. J. Phys. **80,** 913 (2012).

[6] D. Hestenes, Modeling games in the Newtonian World, Am. J. Phys. **60,** 732 (1992).

[7] M. Wells, D. Hestenes, and G. Swackhamer, A modeling method for high school physics instruction, Am. J. Phys. **63,** 606 (1995).

[8] J. Calderhead, Stimulated recall: a method for research on teaching, Br. J. Educ. Psychol. **51,** 211 (1981).

[9] J. Lyle, Stimulated recall: a report on its use in naturalistic research, Br. Educ. Res. J. **29,** 861 (2003).

[10] W. A. Sandoval, Understanding students' practical epistemologies and their influence on learning through inquiry, Sci. Educ. **89,** 634 (2005).

[11] R. A. Duschl and J. Osborne, Supporting and promoting argumentation discourse in science education, Studies Sci. Educ. **38,** 39 (2002).

[12] B. K. Hofer and P. R. Pintrich, *Personal Epistemology: The Psychology of Belief About Knowledge and Knowing* (Lawrence Erlbaum Associates, Mahwah, NJ, 2002).

[13] W. K. Adams, P. Perkins, K. K., M. N. S., Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, Phys. Rev. ST Phys. Educ. Res. **2,** 010101 (2006).

[14] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66,** 212 (1998).

[15] A. Elby and D. Hammer, On the substance of a sophisticated epistemology, Sci. Educ. **85,** 554 (2001).

[16] W. A. Sandoval, In defense of clarity in the study of personal epistemology, J. Learn. Sci. **18,** 150 (2009).

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

[18] A. Elby, Defining Personal Epistemology: A Response to Hofer & Pintrich (1997) and Sandoval (2005), J. Learn. Sci. **18,** 138 (2009).

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

[20] M. L. Minsky, *Society of Mind* (Simon and Schuster, New York, NY, 1986).

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

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

[23] S. A. Rosenberg, D. Hammer, and J. Phelan, Multiple epistemological coherences in an Eighth-grade discussion of the rock cycle, J. Learn. Sci. **15,** 261 (2006).

[24] D. Tannen, *Framing in Discourse* (Oxford University Press, New York, NY, 1993).

[25] J. W. Creswell, *Qualitative Inquiry & Research Design: Choosing Among Five Approaches* (Sage Publications, Thousand Oaks, CA, 2007).

[26] I. Halloun and D. Hestenes, Interpreting VASS dimensions and profiles for physics students, Sci. Educ. **7,** 553 (1998).

[27] R. Scherr and D. Hammer, Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics, Cognit. Instr. **27,** 147 (2009).

[28] R. E. Scherr, Video analysis for insight and coding: Examples from tutorials in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5,** 020106 (2009).

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

[30] D. Hammer and A. Elby, On the form of a personal epistemology, in *Personal Epistemology: The Psychology of Belief About Knowledge and Knowing*, edited by B. K. Hofer and P. R. Pintrich (Lawrence Erlbaum Associates, Mahwah, NJ, 2002), p. 169.

[31] D. Hammer, Epistemological beliefs in introductory physics, Cognit. Instr. **12,** 151 (1994).

[32] M. C. Wittmann, Using resource graphs to represent conceptual change, Phys. Rev. ST Phys. Educ. Res. **2,** 020105 (2006).

[33] R. E. Scherr, Modeling student thinking: An example from special relativity, Am. J. Phys. **75,** 272 (2007).

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

[35] A. A. diSessa, A. Elby, and D. Hammer, J's Epistemological Stance and Strategies, in *Intentional Conceptual Change*, edited by G. M. Sinatra and P. R. Pintrich (Lawrence Erlbaum Associates, Mahwah, 2002), pp. 237–290.

[36] V. K. Otero and D. B. Harlow, Getting started in qualitative Physics Education Research, in *Getting Started in Physics Education Research*, edited by C. Henderson and K. A. Harper (Am. Assoc. Physics Teachers, College Park, MD, 2009), Vol. 1.

[37] A. D. deGroot, *Thought and Choice in Chess* (Mouton, New York, NY, 1965).

[38] R. E. Nisbett and T. D. Wilson, Telling more than we can know: Verbal reports on mental processes, Psychol. Rev. **84,** 231 (1977).

[39] J. J. Chini, A. Carmichael, N. S. Rebello, and S. Puntambekar, Does theTeaching/Learning Interview Provide an Accurate Snapshot of Classroom Learning?, in *Physics Education Research Conference Proceedings*, edited by M. Sabella, C. Singh, and C. Henderson (AIP, Ann Arbor, 2009), pp. 113–116.

[40] E. L. Gaier, Memory Under Conditions of Stimulated Recall, J. Gen. Psychol. **50,** 147 (1954).

[41] J. Brownlee, J.-J Syu, J. Mascadri, C. Cobb-Moore, S. Walker, E. Johansson, G. Boulton-Lewis, and J. Ailwood, Teachers' and children's personal epistemologies for moral education: Case studies in early years elementary education, Teaching and Teacher Education **28,** 440 (2012).

[42] T. H. Schram, *Conceptualizing and Proposing Qualitative Research* (Pearson, Merrill, Prentice Hall, Upper Saddle River, NJ, 2006).

[43] K. Charmaz, *Constructing Grounded Theory* (Sage Publications, Thousand Oaks, CA, 2006).

[44] I. Seidman, *Interviewing as Qualitative Research* (Teachers College Press, New York, NY, 2006).