Factorial measurement of epistemological theories of development

Dr. Todd M. Fernandez, Georgia Institute of Technology

Todd is the Director of Learning Innovation and a Senior Lecturer in the Wallace H. Coulter Department of Biomedical Engineering at Georgia Institute of Technology. His research interests are engineering faculty and students beliefs about knowledge and education with a special focus on how those beliefs interact with engineering education as a culture.

Alexandra Werth, Cornell University

Alexandra Werth is an assistant professor at the Meinig School of Biomedical Engineering, specializing in Engineering Education Research (EER). She focuses on developing evidence-based teaching methodologies to foster authentic learning environments and works to develop novel educational assessment instruments.

Factorial measurement of epistemological theories of development

Abstract:

This paper explores the challenges and opportunities in measuring personal epistemology and epistemic cognition (PE&EC) with a special focus on the unique challenges of engineering education. It is structured in two parts: (1) a retrospective evaluation of current PE&EC measurement instruments and (2) a novel theorized approach to measurement and evidence of validity.

Our evaluation of existing instruments highlights the complexity of validity claims and the challenges in using PE&EC to support broad insights. Our discussion focuses on three critical observations in PE&EC research: (1) the inherent difficulties in measurement at the intersection of attitudes, beliefs, and reasoning; (2) ongoing debates about the nature of validity in PE&EC research, particularly the relationship between psychometrics, measurement, and validity; and (3) the role of the theoretical underpinnings of epistemological beliefs in the practical interpretability of instruments. In particular, we focus on the over-reliance on factor analysis and assumptions of latent structures in standard PE&EC assessments to advocate for a greater focus on aligning measurement with utility and individuals rather than theories.

Moving forward, we propose a novel framework for PE&EC measurement that enables flexibility for diverse research contexts, populations, and designs. Our framework is grounded in factorial survey design, regression techniques, and modern approaches to validity. We argue that this approach will ease existing barriers to quantitative PE&EC research while enabling the reuse and potential unification of existing instruments. Particular to engineering education, the focus on flexible and reusable instruments can enable longitudinal study of key questions about PE&EC - such as why engineering students progress less and later on existing theories.

Introduction

Imagine a student in a first-year engineering design course. Initially, the student might hold a belief that answers are "correct" or "incorrect" with a single solution—an epistemic perspective literature labels *certain knowledge [1]*. This belief might stem from high school experiences, where success hinged on finding the one definitive answer. In the context of their college-level engineering design course, this perspective may clash with engineering work, which often involves navigating ambiguity and competing priorities [2]. Paradoxically, depending on the assignments and assessments in the course, the student's belief may still align with the way they are asked to approach problems, reinforcing their existing perspective [3].

As the engineering student progresses through their classes, internships, and collaborations, they may begin to recognize that many engineering problems have multiple viable solutions. However, their evolving understanding might be challenged by their experience in courses, like thermodynamics, where problems are often framed to have deterministic answers and definitive correctness [4]. These contrasting experiences could lead the student to develop complex and multifaceted epistemic beliefs, such as distinguishing between engineering work and engineering education, or assigning a specific role to authority figures in shaping their understanding of "correct" approaches to engineering [5], [6]. In the context of engineering education, this issue takes on particular relevance. Research suggests that engineering education often fosters less *epistemic growth* compared to other disciplines, particularly the liberal arts [7], [8]. This disparity underscores the need to effectively measure, study, and address epistemological development within the context of engineering education. As engineers, we believe that any effort to solve this problem starts with good data from credible measurement tools.

Tracking shifts in epistemic beliefs and their growth— reveals complexity in characterizing these changes meaningfully and accurately. Is the focus on the collection of different beliefs, the structural relationship between those beliefs, or the student's ability to situate and adapt these beliefs to different contexts? Understanding how students navigate and internalize these beliefs sheds light on the nature of epistemic beliefs, personal epistemology, and epistemic cognition in post-secondary education. However, it also sheds light on the nature of the experience that is shaping them—which is difficult to disentangle [9]. Unsurprisingly, developing instruments to measure epistemic change has proven to be challenging [1], [9], [10], [11]. Epistemological beliefs shape how individuals approach thinking, learning, and even the fundamental nature of disciplines themselves [2], [5], [8], [12]. Shifts in epistemological perspective are seen as a marker of continued cognitive development into adulthood [1], [13].

Measuring beliefs is difficult for a host of reasons. As researchers strive to capture these constructs through surveys, interviews, and other methods, the understanding of psychometrics and instrument development—particularly for terms like "structure," "constructs," and "factors"—become critical[1]. For instance, PE&EC measurements can be challenging due to variations in how beliefs manifest in different contexts, the internal complexity of their structure, and their constant evolution and refinement. Even the terminology within this research domain is subject to debate, with ongoing discussions about whether we are studying personal epistemology or epistemic cognition and whether these concepts represent distinct phenomena [14], [15], [16]. For purposes of this paper, we use the umbrella term Personal Epistemology & Epistemic Cognition (i.e., PE&EC) to encompass the theoretical basis of the field—but that is not universally agreed upon.

It may come as little surprise that instruments designed to measure PE&EC, such as those that might help understand our first-year engineering student, behave inconsistently [17]. This inconsistency does not mean these measures are inherently flawed, poorly developed, nor unimportant. Rather, the topic itself is complex enough that developing strong instruments to measure it is a challenge. Much of the conversation around measuring PE&EC focuses on the psychometric properties of these instruments (c.f., [10], [16], [18], [19], [20]), which reflects the norms of how instrument quality is evaluated in this area [21]. However, the overreliance on psychometrics in instrument development is an ongoing concern across the social sciences [21].

[1] For clarity, we have defined key terms as we use them—specifically "structure," "constructs," and "factors"—in Table 1.

One major concern is the conflation of latent variables and theoretical constructs [22]. As Uher [21] and others [23], [24] observe, focusing first on whether the data fits an expected set of dimensions—before considering other ways of interpreting it—can create problems. . Most notably, it sets up a hierarchy for interpreting data: if the expected structure appears, further analysis and interpretation continues; if it does not, further analysis often stops. This can mislead researchers about what their data actually means and risks overlooking valuable information when the data does not match the expected structure [21], [22].

When it comes to measuring PE&EC, a common assumption is that any measurement of change must keep the same structure over time. . Keeping the same structure makes it easier to tie changes to a theoretical construct. However, research shows that this approach oversimplifies complex and evolving constructs like PE&EC [15], [25], [26].

Table 1. Definitions of a few useful terms from psychometrics and instrument development that will be used frequently to discuss the measurement of PE&EC in this paper.

Team	Definition
Structure	The relationship of multiple variables—especially **items** in a survey. **Structure** can be preordained (e.g., from theory) or emergent (e.g., from observation). It can be strong (i.e., easy to explain or explain much of the variance in data) or it can be weak. Often **structure** is used as a proxy for a latent **construct**, a variable or concept that cannot be observed directly.
Constructs	When well defined, **constructs** typically reflect the link between theoretical development and empirical research. They define the thing that a survey is designed to measure. The **structure** of data from a multi survey is one measure of how well that survey captures a **construct**.
Dimensions and factors	**Dimensions and factors** are similar to **constructs**, albeit one step more granular. **Constructs** can have dimensions, discrete subcomponents that represent building blocks of a theories implementation into empirical research. **Factors**, conversely, most precisely refer to building blocks of the **structures** of survey data.
Items	Survey **items** are the individual questions or statements that make up a survey or questionnaire. Each **item** is designed to elicit a specific piece of information from respondents, typically aligned with the broader **constructs** or **dimensions** the survey aims to measure.
Patterns of variance	**Patterns of variance** across items in a survey can be analyzed to perform dimension reduction—converting all **items** in a survey into a small set of **factors** each characterized by a group of **items**. The set of mathematical tools for evaluating these properties are known as **psychometrics**—most commonly for our purposes, **factor analysis**. On an instrument developer's best day, the psychometric properties of the items that make a survey show that the ***factors explain much of the variance between items*** and the ***content of the items that make up the factors align with the dimensions of the construct they set out to measure***.

Recent observational research on PE&EC challenges the assumption that epistemic beliefs have a stable, unchanging underlying structure. Such finding conflicts with traditional expectations for stable psychometric properties (e.g., [27], [18]). Instead, the results suggest that capturing both structure and structural variation is important to accurately representing growth and constructs of interest.

Despite this, much of the existing research developing or applying quantitative measures of PE&EC makes methodological choices that assume strong measurement requires fitting data into a clear, pre-existing theoretical structure [9], [17], [28]. In practice, this means researchers often prioritize alignment with abstract theory, even when real-world contexts suggest a more complex reality [21], [22], [23].

This raises a critical question for PE&EC research: Are we trying to measure the theory, or how the construct actually functions in context? The challenge in answering that question from existing research is notable, because the answer fundamentally shapes how we approach measurement [29], how people engage their epistemic beliefs in action [12], [30], and what makes our interpretations valid [14], [21, p. 2], [31].

The remainder of this paper presents a conceptual approach to measuring PE&EC without relying on a priori factor analysis and its assumptions about latent structure and theory. It is organized into two parts. First, we review and evaluate existing measures, showing how normative psychometric practices can complicate PE&EC measurement. Second, we describe a set of tools from measurement research that better address challenges like contextual measurement and assumptions about structure, illustrated with examples from a pilot instrument we are developing.

Retrospective evaluation of existing measures

Measurements designed to assess these beliefs often reflect the tensions between theoretical frameworks and practical applications [9], [23]. In our review we highlight three general areas where such measurements have struggled to achieve; practical utility, consistent performance, and a strong claim of valid measurement.

Theoretical Underpinnings and Measurement Challenges

One of the foundational tools in this field, Schommer's **Epistemological Questionnaire (EQ)**, sought to evaluate personal epistemological beliefs across five dimensions (e.g., certain knowledge) [32]. While groundbreaking, the EQ has been critiqued for persistent psychometric instability when used by others. Factor analyses frequently produced inconsistent dimensions from those in the field, low reliability coefficients, and explained only a small proportion of sample variance (20–35%). Such behavior prompted questions about whether the instrument accurately captured the complexity of epistemological beliefs or merely reflected oversimplified constructs [9], [33].

One challenge of the EQ stems from its reliance on fixed, independent dimensions, which may not fully reflect the dynamic and interrelated nature of epistemic beliefs. For those less familiar with the parlance of measurement development, fixed and independent dimensions drive an approach to score calculation that sums or averages items as groups. Such scores are practically useful for score calculation, but problematic if the way responses to items vary is not consistent with the expected dimensions. For instance, results show that a belief in "omniscient authority" is often intertwined with views on "certain knowledge," making it difficult to isolate these constructs meaningfully despite those being independent dimensions for scoring the EQ. Schraw [33] and Watson [9] both critique this approach to scoring, noting that such instruments attempt to disaggregate what is, in reality, a holistic and fluid system of beliefs [9], [33]. That is, a separation that is practically useful for quantifying epistemic beliefs is at odds with actual thought.

Subsequent instruments have typically attempted to improve psychometric performance while retaining the fixed independent dimensions approach. The **Epistemic Beliefs Inventory (EBI)**, a streamlined version of the EQ, sought to improve reliability by reducing the number of items and refining the dimensions. Despite these efforts, the EBI continued to face issues with low reliability coefficients (typically ranging from .50 to .65) and a limited ability to capture the nuanced interplay of beliefs [34]. Similarly, Hofer's **Epistemological Beliefs Questionnaire (EBQ)** introduces two overarching dimensions: the nature of knowing (e.g., certainty and simplicity of knowledge) and the process of knowing (e.g., the justification of knowledge) [5]. While these additions aligned more closely with contemporary theories, the EBQ still struggled with inconsistent factor structures and limited variance explanation [1], [33].

Repeated issues with factor structure and little explanation of variance underscore the challenge of creating generalized measures that are theoretically robust, practically applicable, and consistently demonstrate the properties researchers expect to support the validity of inferences. A recurring critique, highlighted by Watson [9] and others, is the tendency of quantitative approaches to oversimplify the multidimensional and context-dependent nature of epistemological beliefs [9], [26], [33], [35], [36]. For example, individuals' views on the certainty of knowledge may shift depending on the discipline, topic, or specific learning environment [8].

Similarly, progression towards integrating different beliefs, rather than treating them as independent, can be a sign of epistemic growth [27]. Instruments like the EQ, EBI, and EBQ do not account for such variability, leading to results that are less generalizable and potentially misleading. Context dependency in particular has driven a divide in research and theory development. That is whether to measure *personal epistemology* (i.e., an individual construct) or *epistemic cognition* (i.e., as applied by an individual to a specific context) [19], [30].

Moreover, epistemological beliefs are closely tied to other cognitive and affective domains, such as motivation, metacognition, and conceptual change. Hofer and Pintrich argue that beliefs about knowledge are inseparable from beliefs about learning and instruction, further complicating measurement efforts [25]. This interplay makes it challenging to isolate epistemic constructs without inadvertently conflating them with related dimensions - especially in education [9]. As a result, some argue for increased reliance on qualitative methods for analysis of PE&EC [9], [33], [37], [38], [39], [40], [41], [42] while others have focused on developing domain-specific and situated instruments.

Qualitative and Mixed-Methods Approaches

Qualitative methods, including interviews, vignettes, and concept maps, offer richer insights into epistemological and ontological beliefs[9]. Qualitative approaches to studying epistemological belief had a long history in the field. Many of the core foundational theories that others now seek measure quantitatively began as qualitative research studies [25]. In engineering education, the rich nature and resultant value of qualitative approaches is seen in research on both students and faculty [8], [43]. Some efforts have been made to make qualitative approaches more efficient.

For example, the **Views on the Nature of Science (VNOS)** employ open-ended prompts to elicit nuanced conceptions of the nature of science[37]. These approaches align with constructivist theories, emphasizing the dynamic and context-sensitive nature of knowledge. However, their reliance on labor-intensive coding and follow-up interviews limits scalability, particularly in large-scale studies as currently employed.

Concept maps [38], [39], reflections [27], and storyboards [27], [40], [41], [42] have also been explored as tools, capturing the relationships among beliefs and their development over time. These approaches can allow for deeper exploration of tacit knowledge but still require substantial effort in analysis, making them less feasible for widespread application. An advantage of deeper exploration is the ability to interrogate how individuals structure individual beliefs [18], [27]. Structural interrogation often shows individual deviations from those suggested in theory [8], [23], [27]. For example, Kang [27] highlights how pre-service science teachers can hold theoretically incompatible beliefs about the nature of science and the methods of learning science. That is, accepting science as a process of inquiry but retaining a belief that learning science is about collecting facts. While theoretically aberrant, individuals are able to rationalize this mix beliefs through unique combinations and conceptions of their world - e.g., to engage in the process of science, one must have a sufficient foundation of scientific facts. All to say, the ability to show how the *structure* of individual beliefs deviates from the structure of theory shows the clear benefit of qualitative techniques over quantitative approaches that rely on a predefined structure.

Evolution of Discipline-based Quantitative Instruments

The development of discipline-specific epistemological assessments represents a significant evolution in understanding how students engage with and perceive knowledge within STEM fields [44], [45]. Such measurements meet a clear challenge with general instruments—understanding how epistemological beliefs are situated by individuals in disciplinary contexts. This section outlines key developments in the field, focusing on instruments with names such as VASS, MPEX, CLASS, E-CLASS, and EBAE. Such instruments have evolved as a response to specific needs and illustrate the gradual refinement of methodologies, validation efforts, and applications to understand how students engage with knowledge. Notably, most focus on STEM disciplines[2]. Each aims to provide targeted insights into students' beliefs and attitudes, with a primary goal being to enable educators to refine instructional practices and enhance learning outcomes. However, they still highlight challenges in achieving robust validity, scalability, and actionable applications.

The **Views About Sciences Survey (VASS)** is one early effort to systematically assess student beliefs about science. The **VASS**, aimed to capture students' epistemological and cognitive perspectives[39]. VASS distinguishes between "expert" views aligned with scientific realism and "folk" views tied to naive realism. It evaluates beliefs across six dimensions: structure, methodology, validity, learnability, reflective thinking, and personal relevance, which they view as stable between expert and folk views.The survey items present participants with pairs of opposing statements on a Likert-like scale (Figure 1). Validation work included expert reviews, iterative item development, and large scale pilot testing [40]. Statistical alignment with measures such as the Force Concept Inventory (FCI) [46] reinforced VASS' reliability, though its unique scoring format presented challenges for traditional reliability analyses [40].

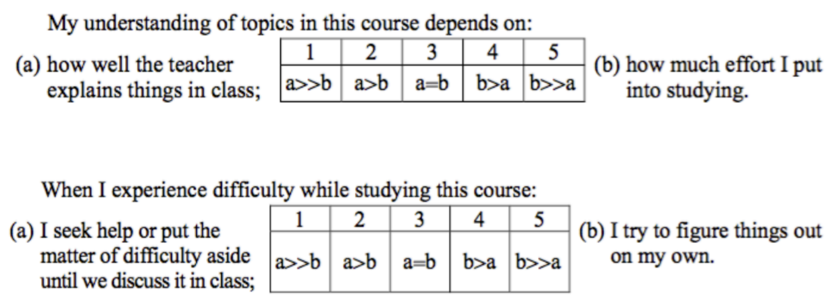

Figure 1. Example of item from the **The Views About Sciences Survey (VASS).**

The **Maryland Physics Expectations Survey (MPEX)** was developed around the same as the VASS and focuses on students' attitudes toward learning physics specifically [36], [47]. MPEX probes six dimensions, including independence, coherence, concepts, reality link, math link, and effort, to assess whether students adopt expert-like approaches to physics learning. For example, the items in Figure 2 probe the coherence dimension. A student with expert-like beliefs regarding coherence would view physics as a connected and consistent framework, while a student with less-expert-like beliefs might see physics as a collection of unrelated facts or "pieces." The first item in Figure 2 situates this belief within the specific context of the course, while the second item takes a broader perspective, asking about physics more abstractly.

Early applications of the MPEX revealed a troubling trend: students often became less expert-like after one semester of physics instruction, especially in dimensions like effort and coherence. Validation efforts included expert benchmarks and student interviews to ensure items reflected intended constructs. However, MPEX's focus on introductory courses limited its applicability, and its reliance on self-reported data raised questions about the alignment between students' attitudes and behaviors.

[2] This section is not all-encompassing. While physics education, in particular, has driven many advancements in epistemological assessment, their tools and methodologies may not fully capture the nuances of epistemic beliefs research in other fields, such as biology, chemistry, social sciences, and the humanities.

A significant problem in this course is being able to memorize all the information I need to know.

Strongly Disagree 1 2 3 4 5 Strongly Agree

Knowledge in physics consists of many pieces of information each of which applies primarily to a specific situation.

Strongly Disagree 1 2 3 4 5 Strongly Agree

Figure 2. Example of items from the **The Maryland Physics Expectations Survey (MPEX).**

The **Colorado Learning Attitudes about Science Survey (CLASS)** is another instrument designed to measure students' beliefs about physics and the process of learning physics. Developed at the University of Colorado Boulder, CLASS targets gaps in other instruments by including empirically derived categories and refined, contextually clear statements[48]. CLASS uses a 5-point Likert scale to gauge students' agreement with 42 statements (Figure 3). These statements assess students' attitudes across eight categories, including real-world connections, personal interest, conceptual understanding, and problem-solving confidence. Unlike previous instruments that relied on *a priori* categories, CLASS is the first of the STEM epistemology surveys to employ factor analysis to identify coherent groupings of statements based on actual student responses rather than theory or prior instruments [48].

A significant problem in learning physics is being able to memorize all the information I need to know.

Strongly Disagree 1 2 3 4 5 Strongly Agree

Knowledge in physics consists of many disconnected topics.

Strongly Disagree 1 2 3 4 5 Strongly Agree

Figure 3. Example of items from the **The Colorado Learning Attitudes about Science Survey (CLASS).**

CLASS has gained widespread use for its ability to probe the beliefs and attitudes that influence students' engagement with and understanding of physics [49], [50]. Validation studies for CLASS incorporated interviews, statistical analyses, and reliability testing on over 5,000 student responses across various physics courses [51]. These efforts established strong correlations between favorable attitudes (aligned with expert beliefs) and improved learning outcomes, emphasizing the importance of addressing attitudes in physics education. For instance, CLASS has revealed significant gender-based differences in personal interest and real-world connection scores, which have implications for promoting equity and inclusivity in STEM fields [52]. CLASS has also been adapted for use in other disciplines, including chemistry [46] and biology [47], demonstrating its versatility as an instrument for assessing student beliefs about science. The original physics CLASS has also been translated in many languages, including Arabic, Chinese, Finnish, German, Japanese, Spanish, and Turkish.

However, challenges remain in interpreting CLASS data and using results for instructional design, particularly the complexity of interpreting shifts in beliefs over time. Additionally, Douglas et al. [53] highlighted that many original CLASS items exhibited weak psychometric properties. Through exploratory and confirmatory factor analyses of data from 3,844 introductory physics students, they proposed a revised 15-item CLASS instrument with three distinct factors: (1) Personal Application, measuring how students relate physics to real-world contexts; (2) Problem Solving and Learning, assessing attitudes toward problem-solving approaches in physics; and (3) Effort and Sense Making, capturing students' dedication to understanding physics concepts. They argued that the revised model showed improved internal consistency and structural fit, though further work, especially across demographic groups remained necessary.

The development of CLASS, along with these findings, underscore a pivotal shift in STEM PE&EC survey design, where both content validity—ensuring survey items align with disciplinary goals—and construct validity—testing whether the survey reliably measures underlying theoretical constructs—are both prioritized. In addition, this evolution reflects a shift toward treating validity as a dynamic, iterative process supported by both theoretical and empirical evidence, moving beyond permanent claims of validity/non validity of instruments to ensure robust and meaningful results [31].

The **Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS)** measures undergraduate students' epistemologies, expectations, and attitudes about experimental physics, with a focus on laboratory courses [54], [55]. E-CLASS examines how students perceive experimental physics both from their personal perspective and as they believe professional physicists view the field. The survey also addresses how laboratory practices are valued and experienced in their courses. Developed in response to calls for reform in STEM laboratory curricula, E-CLASS is designed to provide instructors insights into the alignment of instructional practices with the skills and attitudes necessary for scientific research [54], [56], [57].

E-CLASS is structured around 30 core items, each presented in a paired format to assess both personal and professional perspectives (Figure 4). Additional post-instruction items evaluate students' perceptions of how their course emphasized various experimental practices in terms of grading and engagement. E-CLASS is administered online removing logistical barriers [58], [59], such as the need for in-class time and manual scoring, while offering automated reports for instructors that help with interpretation by contextualizing student responses through visualizations and national benchmarks. This structured reporting enables actionable feedback to guide course improvements, which is a challenge with other instruments. These features empower instructors to make informed, data-driven changes to laboratory instruction and support broader research initiatives. The survey has been extensively used in first-year through advanced-level physics labs across numerous institutions, making it a valuable tool for both formative course evaluation and broader research on laboratory education.

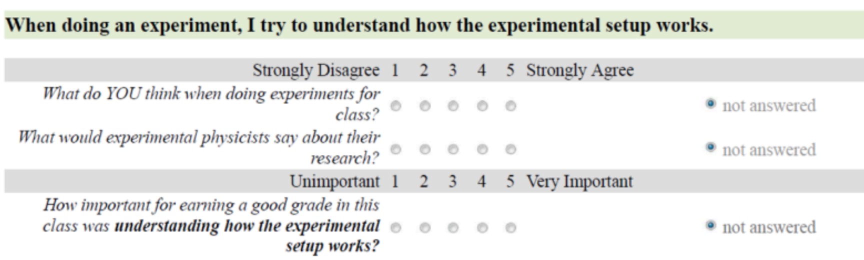

Figure 4. Example item from the **The Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS).**

Efforts to establish the validity and reliability of the E-CLASS including content and face validity efforts through iterative faculty reviews and student think-aloud interviews [48], [49], [50], [53]. Those studies provided evidence about the clarity of item wording, alignment with expert learning goals, and the survey's ability to measure distinct aspects of students' attitudes. For example, during think-aloud interviews with students, the research team found that students often asked if the interviewer wanted them to report their actual beliefs or the right answer. This drove the paired response question format where students are explicitly asked about their view and the view of an experimental physicist. Further statistical analyses demonstrated acceptable levels of reliability (e.g., internal consistency and test-retest reliability) and validity (e.g., concurrent and convergent validity). Factor analysis of the E-CLASS revealed that item responses do not organize into distinct factors. The lack of factor structure suggests that the items address a diverse range of potentially overlapping learning goals rather than aligning neatly with specific constructs. As a result, E-CLASS items are analyzed as standalone constructs, reflecting common learning objectives in physics labs. This flexibility allows instructors to prioritize or emphasize specific items based on their unique course goals and desired learning outcomes. While this adaptability may benefit instructors engaged in scholarly teaching, the frequent use of an overall score to study changes in attitudes may be less appropriate given the instrument's design. Additionally, although the E-CLASS provides robust tools for assessing student attitudes in experimental physics, its utility in other STEM disciplines or non-laboratory contexts remains unvalidated, limiting its broader applicability.

The **Epistemological Beliefs Assessment for Engineering** (EBAE) was developed to evaluate first-year engineering students' beliefs about the nature of engineering knowledge and knowing, focusing on four dimensions: certainty of knowledge, simplicity of knowledge, source of knowing, and justification for knowing [24]. Adapted from the EBAPS [60] and rooted in the theoretical framework of Hofer and Pintrich (1997) [25], the EBAE contextualizes general epistemological constructs specifically for engineering providing a tailored approach to understanding how students perceive and engage with engineering as a discipline.

In a pilot study involving 43 first-year engineering students, factor analysis was used to determine 13 items from the original 22, which aligned with the four dimensions of epistemological beliefs. For example, items included *"Students usually understand engineering better when they present their solutions to their classmates and teachers"* and *"Most engineering principles are set in stone and cannot be argued or changed,"* which students were asked to score on a 100-point Likert type scale with ten-unit intervals. The results indicated that first-year engineering students held slightly sophisticated epistemological beliefs overall, with their most advanced views related to the simplicity of engineering knowledge [24]. Students generally recognized that engineering knowledge is interconnected and context-dependent rather than a mere accumulation of facts. However, their beliefs about the certainty of engineering knowledge were less developed, reflecting a tendency to see knowledge as fixed rather than fluid.

Despite these promising findings, the study faced limitations, including a small sample size and low participation rate (27%), which restricted its generalizability. Future research was recommended to validate the instrument with larger and more diverse populations and to explore factors influencing the development of sophisticated epistemological beliefs in engineering students[17]. However, as far as we are able to identify, little work has continued to develop or use this instrument. Other efforts at developing similar engineering specific instruments (e.g., work by Faber et al.[51]) take the position that the EBAE remains too general, failing to account for differences in students' epistemic beliefs across the multiple disciplines that make up engineering[28]. This critique reflects a broader trend in PE&EC measurement development: the move toward increasingly specific theoretical constructs and highly contextualized instruments. For example, to address the limitations of the EBAE, Faber et al. first identified an more specific set of characteristics within PE&EC theory and applied them to a more specific context (i.e., engineering coursework) with the aim of garnering psychometric evidence that supports using their modified instrument. Notably, they went beyond traditional quantitative methods by including qualitative items designed to enable a deeper exploration of students' beliefs and to enhance item-level validity[28]. However, this work, understandably, adheres to the widely accepted "gold-standard" validation practices, which privilege normative psychometric methods[21], [29], [31], [61]. These practices often emphasize item-level validity over scale validity and rely on established assumptions about the nature of PE&EC [11], [62]. While this approach has clear strengths in ensuring methodological robustness, it may also limit the scope of inquiry by focusing narrowly on fitting pre-defined theoretical structures.

We propose that an alternative hypothesis warrants consideration: while psychometric validation is undeniably valuable, it should not be the sole criterion for evaluating measurement instruments. A broader approach that reassesses underlying assumptions about PE&EC and integrates complementary validation strategies—such as mixed-methods frameworks or multilevel modelling—may provide richer insights into the dynamic and multidimensional nature of epistemological beliefs. By moving beyond traditional practices, we can better capture the complexities of how students engage with and develop these beliefs in real-world contexts.

Discussion about current instruments

The evaluation of current instruments for measuring PE&EC highlights persistent challenges at the intersection of attitudes, beliefs, and reasoning. Three critical issues arise in this context: (1) Measurement complexities, (2) debates around validity, and (3) theoretical and practical (mis)alignment.

Assessing PE&EC often requires navigating the overlapping and interacting domains of attitudes, beliefs, and reasoning processes. While instruments like CLASS and E-CLASS attempt to address these complexities through nuanced and context-sensitive design, they still face limitations in fully capturing the dynamic interplay of these constructs. The reliance on fixed-item formats and latent structures in many instruments often simplifies the inherently interconnected nature of epistemological beliefs, leading to an incomplete understanding of learners' cognitive frameworks.

The question of validity remains central to PE&EC research, particularly concerning the balance between psychometric underpinnings and practical applicability. Instruments such as the MPEX and VASS emphasize statistical validity through techniques like factor analysis, yet they often overlook the interpretive validity required to make findings actionable for educators and learners. That is, they apply analysis methods that presume and can only show growth on stable and independent dimensions. They cannot show change of structure, reconfiguration, internal groups, or other ways of beliefs changing. In fact, such changes are likely to manifest in data as evidence of instrument dysfunction. This gap suggests a need for instruments that prioritize meaningful insights over rigid psychometric structures, emphasizing usability in real-world educational contexts.

A recurring critique of PE&EC instruments is their misalignment between theoretical underpinnings and practical interpretability. For example, while tools like EBAPS and the EBAE are rooted in sophisticated epistemological frameworks, their practical applications are constrained by narrow disciplinary scopes or limited generalizability. This disconnect underscores the need for instruments that bridge theoretical constructs with actionable insights, enabling educators to translate findings into improved instructional strategies.

Additionally, the over-reliance on factor analysis and assumptions of latent structures in many assessments has led to debates about the appropriateness of these methods for such complex constructs. As Watson (2020) argues, traditional psychometric approaches often fail to account for the fluidity and context-dependence of epistemic beliefs, potentially misrepresenting the nuanced interrelationships within these systems. Future efforts must prioritize aligning measurement tools with the practical needs of educators and learners, ensuring that results are not only theoretically valid but also practically interpretable and useful.

Ongoing Needs and Future Directions

Despite advancements in discipline-based instruments, the very specificity that has improved their function also limits the ability to draw generalizable conclusions about epistemology more broadly. While tailored tools, like CLASS, provide rich insights within their respective domains, they require significant adaptation to be applicable in broader or interdisciplinary settings. The increasing specificity of these instruments enhances measurement accuracy but comes at the cost of scalability and transferability. In general, such approaches should be prioritize what we know about the nature of PE&EC, its growth, and its instability over adherence to stable theory.

A promising direction lies in alternative strategies that build on the strengths of existing measures while addressing their limitations. For instance, shifting from absolute to relative scoring frameworks, as seen in E-CLASS, allows for more meaningful comparisons and contextualized interpretations. Relative metrics offer a "ground truth" for learners, making results more interpretable and actionable. This usability-focused approach, such as online administration and auto-scoring, can empower educators and students to engage more effectively with the findings, fostering a deeper understanding of epistemic beliefs. Similarly, analysis methods that evaluate explained variance and variance relatively (i.e., theorized structure concurrently with context) can understand whether low factor performance reflects a problem or an insight. This direction does not require entirely new items or instruments, but can take advantage of alternative approaches to survey instruments and data analysis.

To overcome barriers to broader application, future approaches should emphasize adaptability and contextual relevance over alignment with theoretically grounded structure. This might involve developing modular tools that can be tailored to specific disciplines or educational levels while retaining a core framework that supports cross-disciplinary insights. Additionally, integrating qualitative methods, such as interviews or concept maps, or even video games[63], [64] with quantitative metrics can provide a richer, more nuanced understanding of epistemic beliefs, bridging the gap between theoretical complexity and practical utility. Emerging generative AI (genAI) technologies offer the potential to scale these traditionally resource-intensive qualitative approaches. For example, AI-powered natural language processing tools can assist in coding and analyzing open-ended responses from interviews or concept maps, significantly reducing the time and effort required for manual analysis []. These systems can identify patterns, themes, and relationships within qualitative data with high accuracy, enabling researchers to process larger datasets without sacrificing depth or nuance[65] .

Moreover, genAI tools can facilitate real-time feedback during qualitative assessments, such as automatically generating follow-up questions based on participant responses, enhancing the depth of inquiry [BLINDED]. By leveraging such technologies, researchers can scale qualitative methods to larger populations, integrating the interpretive richness of these approaches with the efficiency and breadth of quantitative tools. This integration not only makes qualitative methods more scalable but also enhances their applicability in diverse educational contexts, opening new avenues for understanding the dynamic and context-dependent nature of epistemic beliefs.

Prospective Strategy

With the challenges of PE&EC measurement in mind, the second half of this paper proposes a fundamentally different approach to instrument development, deployment, and analysis (i.e., a new framework for measurement). Our approach is based on two observations: Prior efforts have laid important groundwork, and addressing core measurement challenges is essential to move the field forward [66].

This framework brings together two established areas of research: factorial survey methods for data collection and multilevel measurement modeling for analysis. We hypothesize that combining these approaches can ease existing barriers to quantitative PE&EC research while enabling broader reuse and potential unification of existing instruments. The differences between current practices and our proposed approach is organized around four core principles: (1) Retaining as much variance in the data throughout analysis, (2) grounding measurement in context, (3) enabling measurement flexibility, and (4) supporting data aggregation, are outlined in detail in Table 2.

Table 2. Alignment of current and proposed approaches to PE&EC instrument design with key principles for addressing challenges in designing quantitative instruments.

Principles	Existing approaches	Proposed framework
Retain as much variance in data throughout analysis	Factor analysis treats item-item variance as error, removing most variance via scoring prior to inferential analysis	Multilevel measurement modeling treats item-item variance as data that should be analyzed alongside, and relative to, other sources of variation. .
Grounding measurement in context	Instruments are contextualized through changes in wording or instructions.	Context is captured as a variable, allowing researchers to directly quantify how specific variations affect responses.

Enabling measurement flexibility	Instruments are fixed sets made up of established items known to work together as a scale and/or subscales.	Contextual dimensions, scenarios, and items/scales can be swapped or modified within defined limitations.
Supporting data aggregation	Traditional psychometric approaches require reassessment and revalidation when applying instruments to new populations.	Factorial surveys by necessity are designed for sparse data analysis techniques, support aggregation even when items or context vary (within limits).

Situating Measurement: Factorial Survey Design

Factorial surveys invert typical latent construct measurement practices. Typical latent construct instruments ask many items about a singular (or no clear) context to achieve construct coverage. In contrast, factorial surveys ask the same question(s) multiple times while presenting the question in different scenarios varied on dimensions of interest [67], [68]. Each scenario is a composite of one 'level' (e.g., male or female) of each dimension of interest. This is an especially useful approach when self-reports are likely to be affected by social desirability bias or other factors [69]. Overall, factorial surveys provide a quasi-experimental approach to studying the nature and scale of situatedness in thinking.

In doing so, factorial surveys capture how judgements, beliefs, or actions change with concrete (i.e., quantifiable) and realistic changes in context [70]. In a single-item factorial survey, what is captured is primarily the effect of contextual variation—responses are relative, illustrating how changing the context of an item changes participants' responses [71]. Analysis then focuses on identifying which dimensions of context affect responses. However, researchers have also demonstrated how small groups of items, organized around an expected latent structure, can be used in factorial survey designs [72]. In these cases, analysis can explore not only how context affects responses but also how context affects the structure of the relationship of the items.

One challenge in factorial surveys is managing survey scope—specifically,determining the number of contexts and contextual dimensions that are most relevant [70]. Because factorial surveys are designed with sparse data in mind, it is rarely necessary nor optimal for participants to respond to all possible scenarios. This has implications for data collection and analysis [70], [71], [73].

Two examples of factorial surveys help illustrate the flexibility and value of this approach to PE&EC. Ludwick & Zeller [73] describe a survey to measure nurses' clinical judgements about patient restraints. They note that prior studies in this area have used singular contexts that oversimplify real world observations and have issues of bias. Each participant responded to two questions about the scale of confusion and about likelihood of using restraints in 4-6 randomly generated scenarios that vary on 7 dimensions such as age, diagnosis, and cognitive state (e.g., The patient is a pleasant, 50-year old White male who has pneumonia. On a busy night shift, you assess the patient, you find that the patient has taken off all clothing). While prior, single scenario research showed that decision might vary amongst individuals, the factorial survey approach highlights how dimensions of contextual variation (e.g., night shift vs. day shift) intersect with individual variance. Similarly, Fernandez & Duval-Couetil [72] create a survey to understand how engineering students' business decision making is affected by context. Their survey uses four scenarios on two binary axes and asks respondents five questions about each scenario. The items are drawn from a prior theoretical framework that proposes two sets of heuristics for business decision making. Prior qualitative research showed that experts consistently picked from one set of heuristics along clear contextual boundaries. The purpose of Fernandez & Duval-Couetil's survey is to test whether (1) students contextualize their heuristics as do experts and (2) whether students structure the selection of heuristics in the same consistent way as experts. That is, they embed a construct with a hypothesized structure into a factorial survey instrument and test how context affects the fit of that structure as well as responses to individual items. While not tested, they note that it is possible to randomly assign items to scenarios as well - increasing construct representation albeit at the expense of statistical power.

Factorial surveys offer a powerful advantage for PE&EC measurement by capturing how specific contextual factors influence beliefs and judgments. Instead of assuming stability across contexts, they allow researchers to directly observe how responses shift across different situations, providing a clearer picture of enacted epistemic beliefs [11], [38], [68]. Their flexible design supports randomization of items and scenarios, making it possible to test open questions about the overlap or distinction between PE&EC theories [9], and they focus on relative changes in data, making validation more transparent and grounded [62], [63]. However, factorial surveys are not without challenges: they cannot specify in advance which contextual dimensions will matter, and balancing survey length, representation, and the number of contextual variables demands careful design decisions [61]. Results are relative rather than absolute, so aggregation across studies is possible, but requires intentional reuse of items, consistent treatment of contextual factors, and shared analysis techniques [61]. Factorial surveys still rely on self-reported data and demand careful validation, but their quasi-experimental structure offers important advantages for establishing internal validity compared to traditional survey methods [67]. Overall, factorial surveys shift the challenge of variability from a liability into a strength, offering a flexible, context-sensitive, and forward-looking approach to study how epistemic beliefs operate in real-world settings.

Data Analysis: Multilevel modeling

Multilevel modeling is a set of statistical techniques used to nest variables across different *levels [74]*. Levels refer to structural hierarchies (e.g., individuals within classrooms within universities), and should not be confused with the "contexts" varied within factorial survey scenarios. Multilevel approaches are beneficial when analyzing variables distributed across such heiarchies [74], [75]. For example, a model predicting standardized test performance might consider variables at the individual level (e.g., gender, socioeconomic status), the class or teacher level (e.g., pedagogical practices), and at the school level (e.g., type of institution, location). At its core, multilevel modeling is a form of regression that partitions variance across multiple variables simultaneously, allowing each level to have its own error structure. This leads to models that are more representative, robust, and interpretable [71].

In our framework for measuring PE&EC, this hierarchical structure is fundamental. Factorial survey data naturally lends itself to multilevel analysis: "Analysis models…should reflect the fact that factorial surveys produce data pertaining to two distinct levels: the individual level, and the [scenario] level" [71]. Table 3 outlines how single-level, two-level, and three-level regression models differ. A single-level model collapses all sources of variance into a single error term. A two-level model separates two sources of error, individual-level variance (idiosyncrasy[71]) from scenario-level variance. This is appropriate when a factorial survey includes only one item. However, single-item designs limit the ability to infer relationships between multiple epistemic beliefs—i.e., latent constructs.

To address this limitation, multilevel measurement modeling extends the standard multilevel approach by adding a third level to account for error and predictors that explain item response variance [76]. In theory, a 'psychometrically perfect' (albeit redundant) scale would have no item-level variance, with all the variance explained by the latent construct. In practice, however, some item-level variance is expected. Multilevel measurement modeling estimates the proportion of variance explained by the latent construct (a variable) and the residual variance among items.. Functionally, the evaluation of latent structure akin to factor analysis techniques becomes embedded in a larger model that can differentiate unexplained among-item variance from other sources (e.g., among-individuals) [75], [77]. The model explained here represents the most basic of a multilevel model.

In practice, predictors can be added at the appropriate level to further apportion relative variance. For example, scenario-level predictors might include contextual dimensions (e.g., classroom vs. internship), individual-level predictors could be demographic variables (e.g., gender), and item-level predictors might track which instrument the item originated from. This flexibility supports a richer analysis of how epistemic beliefs are enacted and structured.

Unlike traditional factor analysis, which establishes structure before modeling, multilevel measurement modeling allows researchers to observe how variance attributed to constructs changes as new predictors are added. For example, theories proposing gender-based epistemologies [9], [73] can be tested by examining how item variance is explained by a shared construct (1) without gender as a predictor, (2) with gender as a second predictor, and (3) with and interaction term of gender and the construct. This shifts the focus from assuming static structure to demonstrating it with data. In summary, multilevel modeling enables researchers to test longstanding questions in PE&EC—such as the overlap of different theories or the situatedness of beliefs—by appropriately leveraging the nested structure of factorial survey design.

Table 3. Alignment of current and proposed approaches to PE&EC instrument design with key principles for address challenges in designing quantitative instruments.

Single level regression Nonhierarchical regression analysis of factorial surveys collapses among-individual and among-scenario variance into one term	Response = Overall_avg ...+ Error
Two level regression Variance among respondents is accounted for and individually estimated separately from variance among scenarios.	**Full equation:** Response_{ij} = Overall_Avg_{00} + $\text{Scenario\_term}_{i0}$ + $\text{Individual\_term}_{0j}$ + Error **Respondent equation (variation across individuals):** Response_{ij} = Avg_Scenario … + Individual_term **Scenario equation (variation across scenarios):** Avg_Response = Overall_Avg … + Scenario_term *i = a specific scenario, j = a specific respondent*
Three level regression Variance between items in a scale is accounted for and estimated separately from variance among individuals or scenarios.	**Full equation:** Response_{ijk} = Overall_Avg_{000} +$\text{Scenario\_term}_{i00}$+$\text{Individual\_term}_{0j0}$+$\text{Item\_term}_{00k}$ + E **Item and scale equation (variation across items):** Response_{ijk} = Avg_Person … + Item_term **Respondent equation (variation across individuals):** Avg_Person = Avg_Scenario … + Individual_term **Scenario equation (variation across scenarios:** Avg_Scenario = Overall_Avg … + Scenario_term *i = a specific scenario, j = a specific respondent, k = a specific item*
Notes: - Ellipses (...) denote where other predictive variables can be applied at each level of the multilevel model - For this paper and audience, we use informal mathematical notation rather than established notation for multilevel modeling (c.f., formulas 1-3 in Peugh [74] or Hox et al. [71])	

While multilevel models offer substantial benefits, they are not the only tools available. Person, rather than item, centered techniques provide a provocative set of alternatives to Hofer's question of 'What develops' in epistemic development [9]. Past work has applied cluster analysis to PE&EC measurement (c.f. [5], [37], [74] for examples and reviews of similar work). Interestingly, these approaches often identify groups of students with similar belief patterns, challenging assumptions about disciplinary differences or dimensional constructs [75]—perhaps suggesting that there are groups of students across disciplines who hold similar beliefs progress in similar ways or suggest a need to understand whether disciplines create or attract those holding similar epistemological beliefs. These results more closely represent the stage models of fundamental theories of development [20] than the dimensionalized approaches pioneered by Schomme r[26]. Techniques, such as latent profile and latent class analysis, exist to extend cluster analysis in ways similar to how multilevel modeling extends simple regression [76].

The primary challenge of multilevel measurement modeling is of sample size and power. DAs the number of contextual dimensions increases, the factorial combination of scenarios grows rapidly, reducing the number of participants who see any given scenario [61], [64]. Researchers must balance the number of dimensions, scenarios, and items with participant sample size and the number of scenarios each participant completes. This introduces complex, nonlinear relationships between study design choices and power [77], [78]. Tools exist to assist with these calculations [79], [80], but researchers may need to estimate parameters due to limited prior work. Careful study design is therefore essential to ensure adequate power while exploring meaningful contextual variation.

Summary

Since Perry's pioneering work in 1970, the only real constant in PE&EC research has been calls for more clarity [13], [14], [15]. Sandoval makes a compelling argument that in the field writ large, clarity is often reduced to construct definition—especially the pursuit of a unitary theory of individual epistemological development. Our approach aligns with Sandoval's description of a more effective approach to clarity in PE&EC research grounded in a path to clarity rather than perfect clarity at the outset of research. Construct definition can be adjusted and refined in factorial surveys over time as well as tested for coherence, stability, and held relative to other variables through multilevel modeling. Clarity, therefore, is a product of data rather than definition.

Generally, factorial surveys and multilevel analysis support a path towards clarity in large part by making the critiques and assumptions of existing instruments into testable. Sandoval's defense of clarity notes three propositions from Andrew Elby to improve PE&EC research: "(a) that ideas about knowing and learning should not be conflated, (b) that psychological definitions of epistemology should align with philosophical ones, and (c) that definitional clarity will aid research in this area."[14, p. 150]. While Elby argues this as a criteria for research, Sandoval conceptualizes it as a project for researchers. Our framework aligns with that project as follows:

(a) The contextual variation of factorial surveys allows the relative comparison of both beliefs about learning and beliefs about knowing in and out of school. Such data analyzed using multilevel measurement modeling can show if/how beliefs about learning and knowing are related while also understanding if/how that relationship is contextually mediated, a sign of epistemic growth, or explained by some individual characteristic.

(b) Multilevel measurement models include predictor variables at the item and scale level. Those variables make it possible to test whether/when different construct definitions and items written to represent them differ in their ability to predict beliefs. Further, our approach de-centers dimensionalization. Doing so removes translational steps (i.e., the dimensionalization of theory) between theory and items. Generally, the small number of items and their situatedness of factorial surveys enables more direct interrogation of many constructs[69]. Doing so also aligns with work in philosophy (e.g., by Gerkken[18]) and in psychology (e.g., by Sandoval[45] and Belenky[78]) that focus on holism and situated enactment as key definitional tools for epistemology.

(c) While there has been value in discipline specific instruments for PE&EC, it remains unclear what can be inferred from one or across many. A single context reduces clarity because the relative effect of that contexts is not measurable. Factorial surveys clearly and directly capture data about relative contextual shifts, which can be modeled, tested for stability, and compared to the individual characteristics that researchers are most interested in. Plainly, the ability to establish what and whether disciplinary context affects answers makes interpretations of an individual's growth, and the generalized or field specific nature of that growth, clearer.

The way in which our approach aligns with the discussion of the future of the field leads to natural plans for future work as well as, we believe, a natural fit for such work in engineering. Our immediate next step is to trial the framework we propose above as a pilot instrument. We intend to use items from existing PE&EC surveys to do so, focusing on establishing the contextualization. We see engineering education as a natural fit for testing this work because of what we know about the nature of engineering faculty, student, and field epistemologies. Students are noted for inconsistent epistemic growth and beliefs [9], [17], [28]. They are likely to experience disciplinary socialization and learning in classes run by faculty whose epistemic beliefs are at odds with the higher stages of theoretical models of development[8]. Finally, the field itself draws on diverse domains of knowledge built on positivist, constructivist, interpretivist, and even critical ways of epistemic beliefs[2]. In summary, the complex landscape of epistemic beliefs in engineering make it an optimal domain in which to stress test a theoretical approach to measuring situated epistemic beliefs.

References

[1] B. K. Hofer, “Personal Epistemology Research: Implications for Learning and Teaching,” *Educ. Psychol. Rev.*, vol. 13, no. 4, pp. 353–383, 2001, [Online]. Available: http://www.jstor.org/stable/23363497

[2] A. D. de Figueiredo, “Toward an Epistemology of Engineering,” *Engineering, The Royal Academy of Engineering …*, Nov. 01, 2008. Accessed: Mar. 16, 2023. [Online]. Available: https://papers.ssrn.com/abstract=1314224

[3] W. C. Newstetter, “Of green monkeys and failed affordances: A case study of a mechanical engineering design course,” *Res. Eng. Des.*, vol. 10, no. 2, pp. 118–128, 1998, [Online]. Available: http://web.mit.edu/jrankin/www/pbl/mechE_newsletter.pdf

[4] M. Chua and L. E. Cagle, “You can change the world, but not this homework assignment: The contradictory rhetoric of engineering agency,” pp. 1–4, Oct. 2017, [Online]. Available: https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=YpEiPJcAAAAJ:dhFuZR0502QC

[5] C. M. Gray and T. M. Fernandez, “When world (view) s collide: Contested epistemologies and ontologies in transdisciplinary education,” *The International journal of engineering education*, vol. 34, no. 2, pp. 574–589, 2018, [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=6874326

[6] T. Fernandez and M. Jacobson, “Engineering Students’ Performance of Prototyping: Process, Purpose, and Perception in the Design Classroom,” *International Journal of Engineering Education*, vol. 38, no. 6, pp. 1761–1778, 2022.

[7] J. Cruz, N. Bruhis, N. Kellam, and S. Jayasuriya, “Students’ implicit epistemologies when working at the intersection of engineering and the arts,” *Int. J. STEM Educ.*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40594-021-00289-w.

[8] D. Montfort, S. Brown, and D. Shinew, “The personal epistemologies of civil engineering faculty,” *J. Eng. Educ.*, vol. 103, no. 3, pp. 388–416, Jul. 2014, doi: 10.1002/jee.20050.

[9] E. Watson, “The slippery business of measuring beliefs: Lessons from a failed attempt at developing an instrument to measure teachers’ epistemic beliefs about physics knowledge,” *The Electronic Journal for Research in Science & Mathematics Education*, vol. 24, no. 2, pp. 119–140, 2020, [Online]. Available: https://ejse.southwestern.edu/article/view/20294

[10] A.-M. Cazan, “Validity of the Discipline Focused Epistemological Beliefs Questionnaire (DFEBQ) on a Romanian Sample,” *Procedia - Social and Behavioral Sciences*, vol. 78, pp. 713–717, May 2013, doi: 10.1016/j.sbspro.2013.04.381.

[11] D. Hammer and A. Elby, “On the form of a personal epistemology,” *Personal epistemology: The psychology of beliefs about knowledge and knowing*, vol. 169190, 2002, [Online]. Available: https://books.google.com/books?hl=en&lr=&id=dUEeYuvj-PsC&oi=fnd&pg=PA169&dq=study+design+personal+epistemology&ots=qk25-s_J1j&sig=cz5BUdjZrElK16pN_ZvS9M-BrfM

[12] C. Faber and L. C. Benson, “Engineering students’ epistemic cognition in the context of problem solving,” *J. Eng. Educ.*, vol. 106, no. 4, pp. 677–709, Oct. 2017, doi: 10.1002/jee.20183.

[13] B. K. Hofer and L. D. Bendixen, “Personal epistemology: Theory, research, and future directions,” in *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues*, vol. 1, Washington: American Psychological Association, 2012, pp. 227–256. doi: 10.1037/13273-009.

[14] W. A. Sandoval, “In Defense of Clarity in the Study of Personal Epistemology,” *Journal of the Learning Sciences*, vol. 18, no. 1, pp. 150–161, Jan. 2009, doi: 10.1080/10508400802581700.

[15] P. R. Pintrich, “Future challenges and directions for theory and research on personal epistemology,” *Personal epistemology: The psychology of beliefs about knowledge and knowing*, pp. 389–414, 2002, [Online]. Available: https://books.google.com/books?hl=en&lr=&id=dUEeYuvj-PsC&oi=fnd&pg=PA389&dq=hofer+and+pintrich+-&ots=qk17-xYL0p&sig=xLiqog63L0EnqOaHYAEM0TG3goQ

[16] P. M. King and K. S. Kitchener, “Judgment model: Twenty years of research on epistemic cognition,” *Personal epistemology: The psychology of beliefs about knowledge and knowing*, vol. 37, pp. 39–61, 2004, [Online]. Available: https://books.google.com/books?hl=en&lr=&id=dUEeYuvj-PsC&oi=fnd&pg=PA37&dq=king+and+kitchener&ots=qk25-s1P0p&sig=ue_aAF1Sl6Xgt6kYM36yX8bTaXg

[17] A. Carberry, M. Ohland, and C. Swan, “A pilot validation study of the epistemological beliefs assessment for engineering (ebae): First year engineering student beliefs,” in *2010 Annual Conference & Exposition Proceedings*, ASEE Conferences, 2020. doi: 10.18260/1-2--15693.

[18] M. Gerken, *On Folk Epistemology: How we Think and Talk about Knowledge*. Oxford University Press, 2017. [Online]. Available: https://play.google.com/store/books/details?id=VMU2DwAAQBAJ

[19] C. A. Chinn, L. A. Buckland, and A. Samarapungavan, “Expanding the Dimensions of Epistemic Cognition: Arguments From Philosophy and Psychology,” *Educ. Psychol.*, vol. 46, no. 3, pp. 141–167, Jul. 2011, doi: 10.1080/00461520.2011.587722.

[20] P. M. King and K. S. Kitchener, “Reflective Judgment: Theory and Research on the Development of Epistemic Assumptions Through Adulthood,” *Educ. Psychol.*, vol. 39, no. 1, pp. 5–18, Jan. 2004, doi: 10.1207/s15326985ep3901_2.

[21] J. Uher, “Psychometrics is not measurement: Unraveling a fundamental misconception in quantitative psychology and the complex network of its underlying fallacies,” *Journal of Theoretical and Philosophical Psychology*, vol. 41, no. 1, pp. 58–84, Feb. 2021, doi: 10.1037/teo0000176.

[22] V. Verhagen, M. Mos, J. Schilperoord, and A. Backus, “Variation is information: Analyses of variation across items, participants, time, and methods in metalinguistic judgment data,” *Linguist. Philos.*, vol. 58, no. 1, pp. 37–81, Jan. 2020, doi:

10.1515/ling-2018-0036.
[23] M. Vishnubhotla *et al.*, "'I use IBL in this course' may say more about an instructor's beliefs than about their teaching," *Int. J. Res. Undergrad. Math. Educ.*, vol. 10, no. 1, pp. 87–106, Apr. 2024, doi: 10.1007/s40753-022-00186-9.
[24] N. Schmitt and G. Kuljanin, "Measurement invariance: Review of practice and implications," *Hum. Resour. Manag. Rev.*, vol. 18, no. 4, pp. 210–222, Dec. 2008, doi: 10.1016/j.hrmr.2008.03.003.
[25] B. K. Hofer and P. R. Pintrich, "The Development of Epistemological Theories: Beliefs About Knowledge and Knowing and Their Relation to Learning," *Rev. Educ. Res.*, vol. 67, no. 1, pp. 88–140, Mar. 1997, doi: 10.3102/00346543067001088.
[26] L. D. Bendixen and D. C. Rule, "An integrative approach to personal epistemology: A guiding model," *Educ. Psychol.*, vol. 39, no. 1, pp. 69–80, Jan. 2004, doi: 10.1207/s15326985ep3901_7.
[27] N.-H. Kang, "Learning to teach science: Personal epistemologies, teaching goals, and practices of teaching," *Teaching and Teacher Education*, vol. 24, no. 2, pp. 478–498, Feb. 2008, doi: 10.1016/j.tate.2007.01.002.
[28] C. Faber, P. Vargas, and L. Benson, "Measuring engineering epistemic beliefs in undergraduate engineering students," in *2016 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2016, pp. 1–7. doi: 10.1109/fie.2016.7757461.
[29] A. Alexandrova and D. M. Haybron, "Is construct validation valid?," *Philos. Sci.*, vol. 83, no. 5, pp. 1098–1109, Dec. 2016, doi: 10.1086/687941.
[30] C. Chinn and W. Sandoval, "Epistemic cognition and epistemic development," in *International handbook of the learning sciences*, Routledge, 2018, pp. 24–33. [Online]. Available: https://educ-met-inclusivemakerspace-2023.sites.olt.ubc.ca/files/2023/05/Halverson-_-Peppler-2018-International-Handbook-of-Learning-Sciences-The-Maker-Movement-and-Learning.pdf#page=47
[31] K. A. Douglas and Ş. Purzer, "Validity: Meaning and relevancy in assessment for engineering education research," *J. Eng. Educ.*, vol. 104, no. 2, pp. 108–118, Apr. 2015, doi: 10.1002/jee.20070.
[32] M. Schommer, "Effects of beliefs about the nature of knowledge on comprehension," *J. Educ. Psychol.*, vol. 82, no. 3, pp. 498–504, Sep. 1990, doi: 10.1037/0022-0663.82.3.498.
[33] G. Schraw, "Conceptual integration and measurement of epistemological and ontological beliefs in educational research," *ISRN Educ.*, vol. 2013, pp. 1–19, Jan. 2013, doi: 10.1155/2013/327680.
[34] G. Schraw, L. D. Bendixen, M. E. Dunkle, B. K. Hofer, and P. R. Pintrich, *Development validation Epistemic Belief Inventory (EBI), Personal Epistemology: Psychology Beliefs About Knowledge Knowing*. Erlbaum, Mahwah, NJ, USA, 2002, pp. 261–275.
[35] W. A. Sandoval, "Understanding students' practical epistemologies and their influence on learning through inquiry," *Sci. Educ.*, vol. 89, no. 4, pp. 634–656, Jul. 2005, doi: 10.1002/sce.20065.
[36] A. Elby, "Helping physics students learn how to learn," *Am. J. Phys.*, vol. 69, no. S1, pp. S54–S64, Jul. 2001, doi: 10.1119/1.1377283.
[37] N. G. Lederman, F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz, "Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science," *J. Res. Sci. Teach.*, vol. 39, no. 6, pp. 497–521, Aug. 2002, doi: 10.1002/tea.10034.
[38] S. Aslı Özgün-Koca and A. İlhan Şen, "The beliefs and perceptions of pre-service teachers enrolled in a subject-area dominant teacher education program about 'Effective Education,'" *Teach. Teach. Educ.*, vol. 22, no. 7, pp. 946–960, Oct. 2006, doi: 10.1016/j.tate.2006.04.036.
[39] X. Liu, "Using concept mapping for assessing and promoting relational conceptual change in science," *Sci. Educ.*, vol. 88, no. 3, pp. 373–396, May 2004, doi: 10.1002/sce.10127.
[40] J. Brownlee, N. Purdie, and G. Boulton‐Lewis, "Changing epistemological beliefs in pre-service teacher education students," *Teaching in Higher Education*, vol. 6, pp. 247–268, Apr. 2001, doi: 10.1080/13562510120045221.
[41] L. Olafson and G. Schraw, "Teachers' beliefs and practices within and across domains," *Int. J. Educ. Res.*, vol. 45, no. 1–2, pp. 71–84, Jan. 2006, doi: 10.1016/j.ijer.2006.08.005.
[42] W. A. Sandoval, "Conceptual and epistemic aspects of students' scientific explanations," *J. Learn. Sci.*, vol. 12, no. 1, pp. 5–51, Jan. 2003, doi: 10.1207/s15327809jls1201_2.
[43] M. M. Barger, S. V. Wormington, L. G. Huettel, and L. Linnenbrink-Garcia, "Developmental changes in college engineering students' personal epistemology profiles," *Learn. Individ. Differ.*, vol. 48, pp. 1–8, May 2016, doi: 10.1016/j.lindif.2016.04.002.
[44] B. K. Hofer, "Dimensionality and Disciplinary Differences in Personal Epistemology," *Contemp. Educ. Psychol.*, vol. 25, no. 4, pp. 378–405, Oct. 2000, doi: 10.1006/ceps.1999.1026.
[45] W. Sandoval, "Situating epistemological development," presented at the Annual Meeting of the International Society of the Learning Sciences (ISLS), pp. 347–354. [Online]. Available: https://repository.isls.org/bitstream/1/2225/1/347-354.pdf
[46] D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," *Phys. Teach.*, vol. 30, no. 3, pp. 141–158, Mar. 1992, doi: 10.1119/1.2343497.
[47] E. F. Redish, J. M. Saul, and R. N. Steinberg, "Student expectations in introductory physics," *Am. J. Phys.*, vol. 66, no. 3, pp. 212–224, Mar. 1998, doi: 10.1119/1.18847.
[48] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, "New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey," *Phys. Rev. ST Phys. Educ. Res.*, vol. 2, no. 1, p. 010101, Jan. 2006, doi: 10.1103/PhysRevSTPER.2.010101.
[49] K. K. Perkins, "Correlating student beliefs with student learning using the Colorado learning attitudes about science survey," in *AIP Conference Proceedings*, AIP, 2005, pp. 61–64. doi: 10.1063/1.2084701.

[50] K. Semsar, J. K. Knight, G. Birol, and M. K. Smith, “The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology,” *CBE Life Sci. Educ.*, vol. 10, no. 3, pp. 268–278, Autumn 2011, doi: 10.1187/cbe.10-10-0133.
[51] I. Halloun and D. Hestenes, “Interpreting VASS Dimensions and Profiles for Physics Students,” *Sci. & Educ.*, vol. 7, no. 6, pp. 553–577, 1998, doi: 10.1023/a:1008645410992.
[52] I. Halloun and D. Hestenes, *Views About Sciences Survey. Annual Meeting National Association Research Science Teaching*. Saint Louis, Missouri, USA: ERIC Document, 1996.
[53] K. A. Douglas, M. S. Yale, D. E. Bennett, M. P. Haugan, and L. A. Bryan, “Evaluation of Colorado learning attitudes about science survey,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 10, no. 2, Nov. 2014, doi: 10.1103/physrevstper.10.020128.
[54] B. R. Wilcox and H. J. Lewandowski, “Students’ epistemologies about experimental physics: Validating the Colorado Learning Attitudes about Science Survey for experimental physics,” *Physical Review Physics Education Research*, 2016, doi: 10.1103/PhysRevPhysEducRes.12.010123.
[55] P. », “ECLASS for Instructors.” Accessed: Jan. 14, 2025. [Online]. Available: https://jila.colorado.edu/lewandowski/research/eclass-instructors-0
[56] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. J. Lewandowski, “Epistemology and expectations survey about experimental physics: Development and initial results,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 10, no. 1, Jun. 2014, doi: 10.1103/physrevstper.10.010120.
[57] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, “Development and validation of the Colorado learning attitudes about science survey for experimental physics,” presented at the 2012 PHYSICS EDUCATION RESEARCH CONFERENCE, AIP, 2013. doi: 10.1063/1.4789747.
[58] B. R. Wilcox, B. M. Zwickl, R. D. Hobbs, J. M. Aiken, N. M. Welch, and H. J. Lewandowski, “Alternative model for administration and analysis of research-based assessments,” *Phys. Rev. Phys. Educ. Res.*, vol. 12, no. 1, Jun. 2016, doi: 10.1103/physrevphyseducres.12.010139.
[59] B. Van Dusen *et al.*, “Online administration of research-based assessments,” *Am. J. Phys.*, vol. 89, no. 1, pp. 7–8, Jan. 2021, doi: 10.1119/10.0002888.
[60] “EBAPS home.” Accessed: Jul. 26, 2023. [Online]. Available: http://www2.physics.umd.edu/~elby/EBAPS/home.htm
[61] S. Messick, “Validity of Psychological Assessment,” *Am. Psychol.*, vol. 50, no. 9, pp. 741–749, 1995, doi: /10.1037/0003-066X.50.9.741.
[62] R. F. Kitchener, “Folk epistemology: An introduction,” *New Ideas Psychol.*, vol. 20, no. 2, pp. 89–105, Dec. 2002, doi: 10.1016/S0732-118X(02)00003-X.
[63] “Something Something Soup Something.” Accessed: Jan. 13, 2025. [Online]. Available: https://soup.gua-le-ni.com/but-why/
[64] “Philosophical Games,” Encyclopedia of Ludic Terms. Accessed: Jan. 13, 2025. [Online]. Available: https://eolt.org/articles/philosophical-games
[65] J. P. Inala *et al.*, “Data analysis in the era of generative AI,” *arXiv [cs.AI]*, Sep. 27, 2024. [Online]. Available: http://arxiv.org/abs/2409.18475
[66] “Standards.” Accessed: Dec. 06, 2024. [Online]. Available: https://m.xkcd.com/927/
[67] P. H. Rossi and S. L. Nock, *Measuring Social Judgments: The Factorial Survey Approach*. SAGE Publications, 1982. [Online]. Available: https://play.google.com/store/books/details?id=pai6AAAAIAAJ
[68] G. Jasso, “Factorial Survey Methods for Studying Beliefs and Judgments,” *Sociol. Methods Res.*, vol. 34, no. 3, pp. 334–423, Feb. 2006, doi: 10.1177/0049124105283121.
[69] K. Auspurg, T. Hinz, S. Liebig, and C. Sauer, “The factorial survey as a method for measuring sensitive issues,” *Improving survey methods*, 2014, doi: 10.4324/9781315756288-21/factorial-survey-method-measuring-sensitive-issues-katrin-auspurg-thomas-hinz-stefan-liebig
[70] L. Wallander, “25 years of factorial surveys in sociology: A review,” *Soc. Sci. Res.*, vol. 38, no. 3, pp. 505–520, Sep. 2009, doi: 10.1016/j.ssresearch.2009.03.004.
[71] J. J. Hox, I. G. G. Kreft, and P. L. J. Hermkens, “The Analysis of Factorial Surveys,” *Sociol. Methods Res.*, vol. 19, no. 4, pp. 493–510, May 1991, doi: 10.1177/0049124191019004003.
[72] T. M. Fernandez and N. Duval-Couetil, “Exploring Ways to Measure Entrepreneurial Mindset: The development of a Student-Focused Effectual Logic Assessment Instrument,” in *2017 ASEE Annual Conference & Exposition*, Jun. 2017. Accessed: Jan. 13, 2025. [Online]. Available: https://peer.asee.org/28344.pdf
[73] R. Ludwick and R. A. Zeller, “The factorial survey: an experimental method to replicate real world problems,” *Nurs. Res.*, vol. 50, no. 2, pp. 129–133, Mar-Apr 2001, doi: 10.1097/00006199-200103000-00009.
[74] J. L. Peugh, “A practical guide to multilevel modeling,” *J. Sch. Psychol.*, vol. 48, no. 1, pp. 85–112, Feb. 2010, doi: 10.1016/j.jsp.2009.09.002.
[75] S. N. Beretvas and A. Kamata, “The multilevel measurement model: introduction to the special issue,” *J. Appl. Meas.*, vol. 6, no. 3, pp. 247–254, 2005, [Online]. Available: https://europepmc.org/article/med/15942069
[76] A. Kamata, D. J. Bauer, and Y. Miyazaki, “Multilevel measurement modeling,” *Multilevel modeling of educational data*, 2008.
[77] K. S. Maier, “A Rasch hierarchical measurement model,” *J. Educ. Behav. Stat.*, vol. 26, no. 3, pp. 307–330, Sep. 2001, doi: 10.3102/10769986026003307.
[78] M. R. Belenky, B. M. Clinchy, N. R. Goldberger, and J. M. Tarule, *Women’sways of knowing: The development of self voice and mind*. New York: Basic Books, 1986.