

Cultivating cross-disciplinary measurement literacy via a unified classical-to-quantum interferometry framework and metrological practices

Zhenhui Du¹ and Zehua Wang¹

*State Key Laboratory of Precision Measuring Technology and Instruments,
Tianjin University, Tianjin 300072, China*



(Received 27 February 2026; accepted 3 September 2026; published 9 October 2026)

Measurement claims carry high stakes, yet students and early-career professionals rarely receive explicit training in the reasoning chain that justifies them. Despite growing emphasis on transferable competence, measurement reasoning remains a blind spot: learners seldom map a measurand (the quantity of interest) to an estimand (the reported quantity inferred from data), anchor inference in an explicit noise or likelihood model, or report results with uncertainty and traceability. This fragmentation weakens rigor and communication across domains. We define cross-disciplinary measurement literacy as the ability to connect a measurand to an estimand, justify inference under explicit noise and likelihood models, and report results with uncertainty and traceability. We operationalized this construct through a unified Encoding $\rightarrow$ Evolution $\rightarrow$ Readout $\rightarrow$ Estimation schema integrated with metrological closure and implemented it in a 4×2 -h module. This multicohort descriptive evaluation includes five cohorts with verified pre- and postadministration records ($N = 114$) and two undergraduate team-task cohorts. Cohort-level summaries showed immediate post-test means 34–54 points above pretest means and normalized gains of 0.642–0.782 on an intervention-aligned assessment. Team rubric summaries provided evidence of feasible group-level enactment of the schema, while uncertainty budgeting and traceability remained the principal difficulties. This study is designed as a bounded descriptive evaluation. It provides evidence for the schema's teachability and its utility as a short-term organizer for measurement reasoning, specifically within the instructional context studied.

DOI: [10.1103/wn6j-gzxx](https://doi.org/10.1103/wn6j-gzxx)

I. INTRODUCTION

Scientific literacy is often framed as the ability to use evidence to make and evaluate claims about the natural world [1]. In science, technology, engineering, and mathematics (STEM) laboratory and design contexts, this hinges on a practical skill that is frequently undertaught: connecting imperfect observations to a quantity of interest through an explicit model, quantifying uncertainty, and communicating why a conclusion is warranted [2]. In this paper, we refer to this skill set as measurement literacy, and we operationalize it as a transferable “measurement-as-inference” schema that links quantity $\rightarrow$ data $\rightarrow$ inference $\rightarrow$ justification across classical and quantum instruments.

Measurement claims are high-stakes arguments, yet students and early-career professionals across research and engineering domains are seldom taught a unified reasoning chain that bridges classical and quantum

contexts—leaving uncertainty, comparability, and traceability to be reconstructed *ad hoc* across domains [3–5]. Classical interferometry is typically taught in instrument-focused courses, while quantum interferometry appears later in advanced units on state descriptions and precision limits [6–10]. This curricular split leaves learners with platform-specific skills but limited ability to rebuild or defend claims when the measurand (target quantity), noise model, or readout changes. Physics education research consistently documents related difficulties [11–14], and the issue has grown more consequential as quantum technologies move toward deployment [15–18]. National roadmaps and international standards now position credible, comparable measurement claims as essential infrastructure [19]. In metrology, every result must be reported with its uncertainty and linked to reference standards through a documented traceability chain [3,4]. Professional testing and calibration demand these practices, yet they are often not made explicit or integrated systematically in physics and engineering instruction [5,20].

Students often encounter “measurement” primarily as device operation and numerical computation, while the reasoning that warrants a claim remains implicit [11]. This omission becomes evident when learners must decide whether two measurements agree despite scatter, interpret

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

variability as evidence, or formulate conclusions from noisy data without reducing uncertainty to a reporting formality [12]. Even when error propagation is taught, students frequently fail to connect uncertainty to valid inference: which assumptions justify the model, which sources dominate, and what claims the data support [11]. Recent studies reinforce these persistent challenges in reasoning about measurement uncertainty across classical and quantum contexts [2,21,22], underscoring that uncertainty is not merely technical but epistemic—that is, part of the logic by which knowledge claims are justified from data [23].

A second fracture is curricular: many programs present classical measurement as a collection of topic-by-topic techniques tied to specific instruments, while quantum measurement is introduced later as a separate subject with a distinct vocabulary and formalisms [6]. This conveys the unintended message that the logic of justifying claims changes fundamentally across domains, rather than remaining stable under a shared inference structure [7]. Such fragmentation matters because modern instrumentation increasingly blurs the classical-quantum boundary. Interferometry exemplifies this: it is ubiquitous across physics and engineering, yet its ultimate performance is often limited by quantum fluctuations that must be treated explicitly as noise sources in advanced systems [8]. Large-scale laser interferometers now routinely operate beyond the shot-noise limit using squeezed states [24,25], requiring practitioners to fluidly navigate classical descriptions (transfer functions and alignment) and quantum ones (shot-noise, back-action, and information limits) to produce a single defensible claim.

Without a portable reasoning schema, fragmentation manifests predictably: students and early-career researchers conflate device performance with claim-level warrants, speak past one another on sensitivity and resolution, or debate terminology instead of shared inference [4]. In high-consequence contexts—such as safety-critical systems, compliance, or multidisciplinary integration settings—a result is scrutinized as an argument: what measurand was operationalized, what model links data to the estimand, which uncertainties dominate, and how the claim compares to specifications or references [5]. Absent explicit links, learners may overgeneralize readouts as facts, undervalue systematic effects, or treat uncertainty as a procedural formality rather than an inference constraint [3].

Public episodes illustrate the high stakes when justification chains remain implicit. In 1999, the *National Aeronautics and Space Administration* Mars Climate Orbiter was lost after a unit mismatch (imperial vs *International System of Units*) in ground software propagated across interfaces, turning a calibration error into mission failure [26]. In 2011–2012, the *Oscillation Project with Emulsion-tracking Apparatus* collaboration initially reported neutrinos apparently exceeding the speed of light;

the anomaly was later traced to a faulty fiber-optic connection and timing calibration flaws rather than new physics [27]. In 2014, the *Background Imaging of Cosmic Extragalactic Polarization* announced evidence of primordial gravitational waves, but joint analysis with Planck data showed the signal was consistent with Galactic dust contamination, underscoring how inference depends on accurate modeling of foregrounds and cross-checks [28]. These cases, though diverse in domain and scale, share a central lesson for measurement education: credible claims require explicit quantities, validated models, rigorous uncertainty evaluation, and unbroken traceability chains—not merely computations performed after data collection.

The urgency intensifies with quantum technologies moving from demonstrations to market formation [8]. National initiatives treat quantum metrology and standards as strategic priorities [15], while *International Organization for Standardization & International Electrotechnical Commission Joint Technical Committee* and BIPM workshops highlight the metrology-for-quantum gap and the need for coordinated measurements on commercial timescales [16–18]. As quantum-enabled measurements enter multidisciplinary teams, divergent performance narratives—e.g., *signal-to-noise ratio* versus *Fisher information* (FI)—can satisfy local metrics yet fail at system integration [17,18]. Educational interventions that render the justification chain explicit thus serve as workforce infrastructure, preventing classical and quantum measurement languages from becoming mutually unintelligible [29,30].

Accreditation frameworks emphasize transferable competence [31–33], yet measurement reasoning remains a persistent blind spot. Learners are rarely taught to map measurand to estimand, anchor inference in an explicit noise model, and report conclusions with uncertainty and traceability [6,11]. When these links remain implicit, the justification chain becomes difficult to reconstruct across domains, weakening rigor and communication. This omission is increasingly consequential, as measurement practice becomes more interdisciplinary and standards-dependent.

Addressing this gap, the present paper treats claim justification as a teachable, portable, and assessable object. We propose a unified schema—Encoding → Evolution → Readout → Estimation—paired with metrological practices of scale-factor specification, uncertainty evaluation, and traceability [3,4]. Building on a foundational schema developed in our prior textbook [34], the framework originates in quantum metrology [8] but is designed to be platform-agnostic. Implemented in a focused 4×2 -h module, it was examined through a multicohort descriptive evaluation. The study provides empirical evidence of immediate conceptual uptake and demonstrates the group-level enactment of the schema in an unfamiliar task. By using interferometry as a high-tension exemplar, we present a feasibility study for a potentially reusable organizer for measurement reasoning

across classical and quantum contexts, along with its attendant limitations

II. FRAMEWORK, INSTRUCTIONAL CONTEXT, AND METHODS

A. Terminology

Measurand q is the physical quantity of interest, estimand is the reported quantity derived from data, θ is the encoded internal parameter (phase ϕ as a common special case), scale factor $S_q = d\theta/dq$ makes the measurand-to-estimand mapping explicit, the likelihood $p(x|\theta)$ describes data generation, uncertainty quantifies the dispersion of the reported value, and traceability links the result to stated reference standards through a documented calibration chain.

For reader convenience, Appendix A provides a glossary of key terms (Table IV) with plain-language definitions of the core measurement concepts used throughout the paper.

B. Measurement literacy framework:

Definition → implementation → alignment

1. Conceptual foundation

Cross-disciplinary measurement literacy is defined as a transferable capability comprising three core competencies:

- (i) mapping a measurand (the physical quantity of interest) to an estimand (the reported quantity) through an explicit scale factor,
- (ii) justifying inference with a stated noise or likelihood model and information-based precision limits, and
- (iii) reporting uncertainty and traceability in accordance with metrological norms.

These competencies establish a rigorous, portable basis for justifying measurement claims across disciplines—from classical sensors to quantum-enhanced systems.

2. Pedagogical implementation

This definition is operationalized through a unified measurement schema—Encoding → Evolution → Readout → Estimation—integrated with metrological closure (scale-factor application, uncertainty budgeting, and traceability). The schema originates from quantum parameter estimation [8] but is extended here to encompass both classical and quantum domains in a platform-agnostic manner. The instruction emphasizes that the same inference structure governs performance limits in all cases, with metrological practices closing the loop: sensitivity claims must be supported by calibration, uncertainty budgets, and traceable references.

The schema is described as follows (see Fig. 1). For accessibility, Fig. 1 uses the generic mapping $q \rightarrow x$ (quantity → data). When helpful for interferometry and quantum examples, we refine this as $q \rightarrow \theta \rightarrow x$, where

θ is the encoded parameter (e.g., phase) and x is the recorded readout.

- **Encoding** maps the measurand q to a parameter θ via $\theta = f(q)$, with interferometric phase ϕ as a prominent case. The scale factor $S_q \equiv d\theta/dq$ makes sensitivity explicit and bridges device output to the target quantity.
- **Evolution** describes probe (classical field or quantum state) response to θ , including ideal deterministic transformations and dominant degradation processes such as loss and noise.
- **Readout** produces observable outcomes x (e.g., intensity, population, and voltage), defining the likelihood model $p(x|\theta)$.
- **Estimation** infers $\hat{\theta}$ (and thus $\hat{q}$) from data via an estimator, with precision constrained by information limits and uncertainty evaluation.

Metrological closure links estimation back to uncertainty and traceability: phase (or parameter) precision translates to measurand precision via $u(q) = u(\theta)/|S_q|$, while *Guide to the Expression of Uncertainty in Measurement* (GUM)-based uncertainty budgets aggregate dominant sources and *International Vocabulary of Metrology—Basic and general concepts and associated terms* (VIM)-defined traceability ensures comparability [3,4].

3. Information limits

The readout and estimation stages are framed by statistical estimation theory. Classically, Fisher information $F(\theta)$ quantifies sensitivity of outcomes to θ , yielding the Cramér–Rao bound (CRB): $\text{Var}(\hat{\theta}) \geq 1/[\nu F(\theta)]$ under regularity conditions, where ν is the number of trials [3]. Quantum metrology generalizes this via quantum Fisher information (QFI), providing an upper bound on achievable FI and the quantum Cramér–Rao bound (QCRB) [35,36].

This instruction uses FI and QFI not for exhaustive optimization but to unify classical, quantum-enhanced, and fully quantum interferometry: resources and noise constrain the available information for estimation. This perspective connects directly to exemplars such as squeezed-light interferometry [37] and frequency-dependent squeezing in Advanced Laser Interferometer Gravitational Wave Observatory (LIGO) [38].

Within this unified schema, the information limits provide a common language for comparing performance across classical and quantum regimes. For classical or coherent-light interferometry, the fundamental bound is the *Standard Quantum Limit* (SQL), also known as the shot-noise limit, where precision scales as $1/\sqrt{\nu}$ (ν independent trials or average photon number) due to independent particle statistics [3,38]. Quantum resources—such as squeezed states or entanglement—enable surpassing the SQL and approaching the *Heisenberg Limit* (HL), where precision scales as

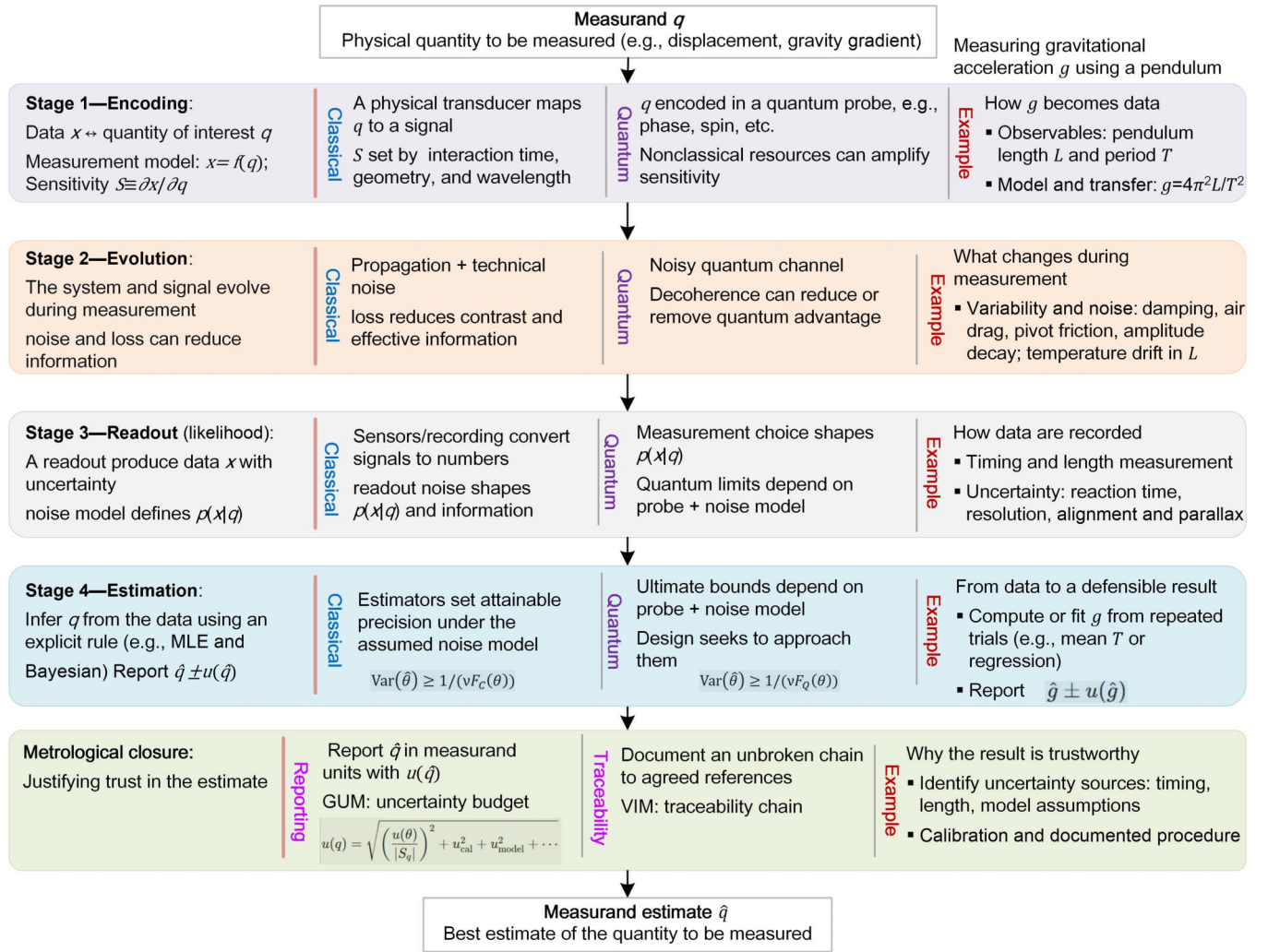


FIG. 1. Unified measurement-as-inference schema (Encoding-Evolution-Readout-Estimation) with metrological closure (uncertainty reporting and traceability). Each stage is paired with a plain-language prompt to guide interpretation (e.g., “How q becomes data,” “What changes during measurement,” “How data are recorded,” and “From data to a defensible result”). The example column illustrates a pendulum-based measurement of gravitational acceleration g in an introductory physics lab: observables $\{L, T\}$ with model $g = 4\pi^2 L / T^2$ (Encoding); variability from damping, drag, friction, and drift (Evolution); timing and length readout uncertainties (Readout); estimation via repeated trials or fitting with report $\hat{g} \pm u(\hat{g})$ (Estimation); and an uncertainty budget with documented calibration and procedure (Closure).

$1/\nu$ by exploiting correlations that enhance phase sensitivity beyond classical bounds [8,35,36].

Instruction focuses on these limits not for exhaustive derivation but to illustrate that the schema captures both regimes: the same Encoding-Evolution-Readout-Estimation structure governs whether performance is SQL-constrained (classical-like) or approaches HL (quantum-enhanced), with loss and decoherence acting as practical constraints on achievable gains [9,37,38]. This framing helps students treat SQL and HL as information-theoretic benchmarks within a portable reasoning chain, rather than domain-specific formulas, thereby supporting transfer to diverse measurements and reinforcing cross-disciplinary measurement literacy.

4. Metrological closure and transferability

A key objective is connecting information limits in θ to defensible claims about q . The scale factor clarifies how design choices influence end-to-end performance and helps students distinguish it from estimator variance [3]. The CRB and QCRB serve as information-theoretic benchmarks, while practical uncertainty follows GUM aggregation of sources such as alignment, calibration, and drift. This tension—ideal limit versus real-world budget—shows that sensitivity is a traceable claim, not an abstract formula.

Traceability embeds results in a calibration chain that contributes to uncertainty [4]. By requiring students to

specify the estimand, dominant noise assumptions, uncertainty budget, and traceability statement, the schema makes professional norms visible and assessable across classical and quantum contexts.

Transferability is the central goal. Though introduced via interferometry, the schema is modality-agnostic, enabling application to diverse engineering measurements. It provides a reusable template that reduces compartmentalization, treats performance as model-based inference, and aligns reporting with metrological standards—thus cultivating cross-disciplinary measurement literacy as the ability to justify portable, defensible claims [39,40].

C. Instructional design of the 4 × 2-h module

1. Design principles and alignment

The module is based on constructive alignment [41]: intended learning outcomes (ILOs), learning activities, and assessments are closely coupled to reinforce the three core competencies of measurement literacy. To preserve depth within four 2-h sessions, the design prioritizes a focused set of potentially transferable ideas and uses active-learning tasks that require learners to apply the schema explicitly [42]. This approach is consistent with evidence that active engagement improves performance in science, technology, engineering, and mathematics and can support integrated cross-disciplinary learning [43,44].

The module treats the unified schema as a boundary object for learners from different majors, including physics, instrumentation, optoelectronics, and engineering. Its purpose is to provide a common reasoning structure through which platform-specific knowledge can be organized and communicated. Transfer beyond the taught settings remains an intended outcome rather than an established result of the present study.

2. Intended learning outcomes

The ILOs operationalize measurement literacy by targeting the three competencies:

- ILO1.** Explain interferometry as parameter estimation from a likelihood model and interpret Fisher information as a model-dependent sensitivity measure.
- ILO2.** Reason about how operating point, loss, noise, and decoherence affect Fisher information and the Cramér–Rao bound (classical and quantum).
- ILO3.** Translate parameter precision into measurand precision using an explicit scale factor; distinguish the scale factor from sensitivity and from reported uncertainty.
- ILO4.** Construct a basic uncertainty budget and describe metrological traceability for an interferometric (or analogous) measurement, per GUM and VIM norms.
- ILO5.** Articulate how quantum resources (e.g., squeezing and entanglement) modify information limits and

how practical constraints (loss and decoherence) limit gains.

- ILO6.** Apply the unified schema to analyze an unfamiliar interferometric sensor and justify design or operating-point choices with evidence from the schema (model, noise, information, and calibration).

These outcomes ensure that conceptual understanding (competencies i–ii) and applied professional practice (competency iii) are both addressed.

3. Session structure, activities, and evidence

Each 2-h session centers on one or more stages of the unified schema, producing a session artifact for formative or summative assessment. Active tasks require students to externalize the schema—identifying encoding and scale factor, specifying likelihood assumptions, evaluating information limits, and closing with uncertainty/traceability—thus practicing the full justification chain.

Table 1 summarizes the four-session structure, schema focus, core concepts, active-learning tasks, and assessment alignment. The sequence is scaffolded to move from guided use of individual schema elements toward integrated application in an open-ended team task.

Session 1 introduces the schema by contrasting classical and quantum interferometer families. Learners classify cases and justify their classifications in a short memo, thereby practicing early readout and estimation reasoning (ILO1) and distinguishing hardware descriptions from claim-level warrants.

Session 2 addresses encoding and evolution through a worksheet that requires explicit scale-factor calculations across optical, atomic, and solid-state examples (ILO3). The activity emphasizes how the same measurand-to-parameter mapping can organize reasoning in classical and quantum contexts.

Session 3 develops information-limited inference (ILO2 and ILO5) through worked problems comparing FI and QFI curves and SQL and HL distinctions under loss. The goal is to frame precision as a model-dependent inference claim rather than an isolated formula.

Session 4 closes the metrological loop (ILO4 and ILO6). Teams prepare a brief measurement-design proposal containing a schema diagram, uncertainty budget, traceability chain, and defensible performance claim.

Figure 2 summarizes this constructive alignment. Artifacts progress from low-stakes formative work (memo and worksheet) to more integrated assessment (quiz and team design proposal), progressively eliciting competencies in scale-factor mapping, likelihood and information-based justification, and uncertainty and traceability reporting. The pre- and postassessment provides descriptive evidence for RQ1, while the team proposal provides group-level performance evidence for RQ2. Optional reflections are retained only as contextual material.

TABLE I. Overview of the 4×2 -h cross-disciplinary module.

Session (2 h)	Schema focus	Core concepts	Active-learning task (artifact)	Assessment link
S1: Classical vs quantum	Readout and estimation (conceptual) + case contrasts	Interferometer families; transferable explanations	Case-classification card sort + justification memo	Formative: justification rubric
S2: Encoding and scale factor	Encoding + evolution	Measurand-to-phase mapping; scale factor $S = d\theta/dq$; operating point	Scale-factor worksheet across domains (optical, atomic, and solid-state)	Formative: quick check + feedback
S3: Information limits	Readout + estimation	FI and CRB; QFI and QCRB; SQL vs HL; role of loss	FI-vs-operating-point comparison on a simple likelihood model	Quiz + worked problems
S4: Traceability and uncertainty loop	Estimation $\rightarrow$ uncertainty and traceability	Uncertainty budgets; traceability chains; defensible sensitivity claims	Schema diagram + brief team measurement-design proposal	Summative: mini-project rubric

4. Assessments and evidence sources

Two primary evidence sources address the research questions. Optional learner reflections provide contextual, noninferential commentary.

- (i) A pre- and postconceptual assessment targeting competencies (i) and (ii): measurand-estimand mapping, scale-factor reasoning, likelihood- and information-based inference, and translation of parameter uncertainty into measurand uncertainty.
- (ii) A team mini-project requiring application of the unified schema to an unfamiliar interferometric sensor. An analytic rubric assessed schema completeness, model and likelihood specification, information-based reasoning, scale factor and calibration, uncertainty treatment, traceability, and claim justification across competencies (i)–(iii).

Optional learner reflections were collected after the module to illustrate perceived usefulness, remaining confusions, and possible contexts for future application.

Because the reflections were optional and were not collected through a prespecified qualitative sampling and coding design, they are used only as contextual illustrations and are not treated as an independent outcome or research question.

The mini-project is the principal performance probe. Teams use the full justification chain to construct a measurement claim in a scenario not used during instruction. The resulting evidence concerns feasible group-level enactment under open-ended conditions; it does not establish individual mastery, durable transfer, or performance outside the instructional setting.

5. Research questions and core competencies

We address two research questions aligned with the three core competencies used to operationalize measurement literacy:

RQ1 (Immediate conceptual uptake): How do cohort-level scores on a schema-aligned conceptual assessment

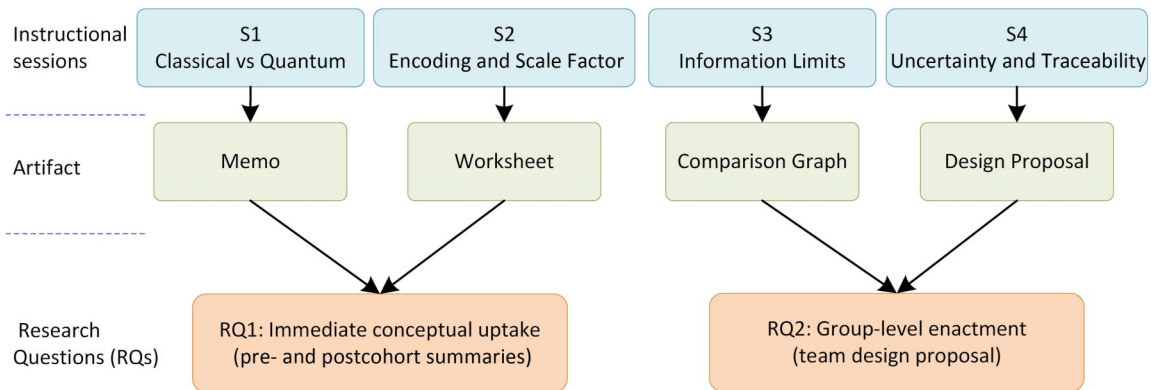


FIG. 2. Four-session instructional sequence and evidence alignment. Sessions generate progressively more open artifacts. The pre- and postconceptual assessment addresses RQ1; the team design proposal addresses RQ2. Optional reflections are contextual evidence only.

differ between preinstruction and immediate postinstruction across the five cohorts with verified and comparable administration records?

RQ2 (Group-level enactment): To what extent do undergraduate teams enact the three core competencies when applying the schema to an unfamiliar measurement-design task?

Core competencies (used consistently throughout).

(i) Map measurand $\rightarrow$ estimand via an explicit scale factor and calibration logic, (ii) justify inference with an explicit noise or likelihood models and information-based limits, and (iii) report claims with uncertainty and traceability (metrological closure).

D. Methods

We conducted a multicohort descriptive evaluation in authentic course settings. This study is structured as a descriptive evaluation to assess immediate, intervention-aligned outcomes. All measures were aligned with three core competencies: (i) measurand-estimand mapping through an explicit scale factor, (ii) inference justification under stated noise or likelihood models and information limits, and (iii) reporting with uncertainty and traceability.

1. Setting and participants

The module was implemented in seven cohorts at a research-intensive university: five graduate offerings (2021–2025) and two upper-division undergraduate offerings (2024–2025). Participants represented mixed STEM and engineering backgrounds, including physics, instrumentation, optoelectronics, and mechanical and mechatronics engineering.

The module was delivered in seven cohorts. The primary conceptual-assessment analysis was restricted to five cohorts with verified and comparable pre- and postadministration records (Grad-2023–2025 and UG-2024–2025; $N = 114$). Two earlier graduate cohorts (Grad-2021 and Grad-2022; $N = 29$) are retained only as implementation history because retrospective review identified inconsistencies in pretest administration or recording. The graduate cohorts are not treated as a comparison or control condition.

2. Measures and construct alignment

- Pre- and postconceptual assessment (0–100 scale). The assessment targeted competencies (i) and (ii), including scale-factor mapping, readout-to-likelihood reasoning, FI and QFI and CRB and QCRB distinctions, and uncertainty translation. Items were drawn from a 35-item bank, and each administration used an abbreviated 10–12-item form balanced across schema stages. These short forms were derived from the item bank but were not independently validated as separate instruments [46]. Scores are therefore interpreted as

an intervention-aligned measure of immediate schema uptake rather than as a domain-general concept inventory [50,51].

- Team mini-project. Undergraduate teams applied the unified schema to an unfamiliar interferometric sensor and submitted a brief measurement-design proposal. The analytic rubric assessed measurand definition, four-stage schema completeness, noise and likelihood specification, information-based reasoning, scale factor and calibration, uncertainty budgeting, traceability, and claim justification. Because the unit of assessment was the team submission, the results concern group-level enactment rather than individual transfer.
- Optional reflections. Brief postmodule comments were available for some learners. Because participation was optional and the collection was not designed as a stand-alone qualitative study with prespecified sampling, coding, and reliability procedures, reflections are used only as illustrative contextual evidence and are not analyzed as an independent qualitative outcome.

3. Procedure

The pretest was administered immediately before Session 1 and the post-test immediately after Session 4. The four 2-h sessions were delivered over 4 consecutive weeks. The team measurement-design proposal was assigned at the end of Session 4 and submitted 1 week later. Optional reflections were collected after the module where available.

4. Data analysis

Conceptual assessment. For each verified cohort, we report sample size, pretest mean and standard deviation, immediate post-test mean and standard deviation, the difference between cohort means, and cohort-level normalized gain. Normalized gain was calculated from the cohort means as $g = (M_{\text{post}} - M_{\text{pre}})/(100 - M_{\text{pre}})$ [47]. Given the small and heterogeneous cohort sizes, we prioritize directly interpretable cohort-level descriptive statistics (means, standard deviations, and normalized gains) over paired inferential tests. This approach provides the most stable and transparent summary of the observed score differences across cohorts.

The quantitative interpretation is descriptive. We report the five verified cohorts separately and examine the direction and magnitude of immediate score differences across offerings. Grad-2021 and Grad-2022 are excluded from the primary analysis and documented only in Supplemental Material [45].

Mini-project. We report team counts and rubric means and standard deviations for the two undergraduate cohorts. Because the number of teams was small and the reliability estimates were based on a limited double-scored subset,

TABLE II. Pre- and immediate post-test summaries for verified cohorts.

Cohort	N	Pretest mean (SD)	Post-test mean (SD)	Difference	Normalized gain g
Grad-2023	6	47 (8)	81 (7)	34	0.642
Grad-2024	18	45 (12)	88 (10)	43	0.782
Grad-2025	11	43 (10)	85 (9)	42	0.737
UG-2024	34	19 (12)	73 (11)	54	0.667
UG-2025	45	21 (11)	74 (13)	53	0.671

Note. Scores are reported on a 0–100 scale. Difference is the post-test cohort mean minus the pretest cohort mean. Normalized gain is calculated from cohort means. Grad-2021 and Grad-2022 are excluded from the primary analysis and reported only as implementation history in Supplemental Material [45].

TABLE III. Team-level performance summaries for undergraduate cohorts.

Cohort	N teams	Rubric total mean (SD)	Group-level enactment subscore mean (SD)	Common strengths and difficulties
UG-2024	6	85 (6)	86 (5)	Strong schema mapping and completeness; weaker uncertainty budgeting and traceability articulation
UG-2025	9	87 (5)	88 (6)	Strong schema mapping and completeness; weaker uncertainty budgeting and traceability articulation

the final manuscript does not emphasize intraclass correlation coefficient (ICC) point estimates; instead, it describes the scoring procedure and interprets the rubric results as descriptive evidence of group-level enactment [49].

The scope of inference is therefore deliberately bounded to the instructional setting. The design supports conclusions about immediate, intervention-aligned score differences and the feasibility of group-level schema enactment, which are the primary foci of this investigation.

III. RESULTS

Results are organized around the two research questions. Table II reports cohort-level pre- and postsummaries for the five verified cohorts (RQ1), and Table III reports undergraduate team performance on the open-ended measurement-design task (RQ2).

Seven cohorts completed the module, but the primary conceptual-assessment summary is restricted to five cohorts with verified administration records ($N = 114$). All quantitative results are interpreted descriptively as evidence of immediate, intervention-aligned uptake.

A. Immediate score differences on the conceptual assessment

Across the five verified cohorts, pretest means ranged from 19 to 47 and immediate post-test means ranged from 73 to 88 (Table II). Differences between cohort means ranged from 34 to 54 points, and cohort-level normalized gains ranged from 0.642 to 0.782. These statistics document substantial immediate uptake of the explicitly taught schema, consistent with the assessment's design as an intervention-aligned measure. They are not intended to

serve as a proxy for effects from broadly validated concept inventories or causal claims.

Graduate cohorts entered with higher mean scores (43–47) than the undergraduate cohorts (19–21), and their immediate post-test means were also higher (81–88 versus 73–74). The undergraduate cohorts showed larger absolute differences between cohort means (53–54 points), whereas the verified graduate cohorts showed differences of 34–43 points. This pattern is consistent with differences in baseline preparation, assessment alignment, or cohort composition, which are typical of multicohort convenience samples.

Because the assessment was closely aligned with the module, the score differences are best interpreted as short-term acquisition of the targeted vocabulary and reasoning structure: scale-factor mapping, explicit noise and likelihood models, information-based limits, and uncertainty translation. The present evidence does not establish durable learning or far transfer.

Figure 3 presents cohort means and standard deviations for the five verified cohorts. This display emphasizes the consistency of direction and magnitude across cohorts and shows between-student dispersion.

B. Group-level enactment in the mini-project

Undergraduate teams applied the schema to an unfamiliar interferometric sensor and submitted a structured measurement-design proposal. Cohort-level summaries show rubric means of 85 and 87 and group-level enactment subscore means of 86 and 88 for UG-2024 and UG-2025, respectively (Table III). These scores indicate that teams could organize an unfamiliar problem using the four-stage

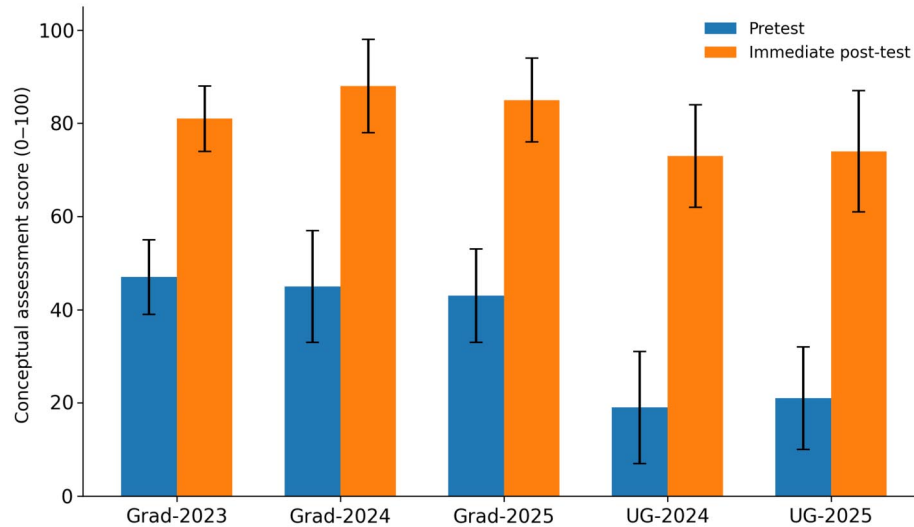


FIG. 3. Cohort-level pretest and immediate post-test means for the five verified cohorts ($N = 114$). Error bars show one standard deviation and are descriptive rather than confidence intervals. The figure emphasizes the consistency of direction and magnitude across cohorts; no participant-level paired inference is implied.

schema under the task conditions; they do not demonstrate individual transfer.

The rubric summaries identify measurand-to-parameter mapping and four-stage completeness as recurring strengths. Uncertainty budgeting and the articulation of traceability were the most common difficulties. This pattern suggests that teams adopted the organizing structure more readily than they completed the metrological warrant required for a defensible performance claim.

Together, Tables II and III provide two bounded forms of evidence: substantial immediate score differences on an aligned assessment and feasible group-level enactment in an unfamiliar task. They do not establish causal impact, individual mastery, durable learning, or far transfer.

IV. DISCUSSION

A. A bounded but consistent pattern of evidence

Across five verified cohorts, the evidence consistently supports a set of bounded conclusions: substantial immediate post-test score differences on an aligned assessment, and feasible group-level enactment of the schema in an unfamiliar design task. These findings are robust within the instructional context and suggest that the module can serve as a useful short-term organizer for measurement reasoning.

The contrast between strong schema completeness and weaker uncertainty and traceability work is educationally informative. The four-stage structure appears easier to adopt than the discipline of completing a defensible metrological warrant. Although descriptive and team-level, this pattern identifies a concrete target for subsequent instruction and assessment.

The study's design prioritizes contextual authenticity and feasibility over experimental control, resulting in no

comparison group, immediate post-testing, and small, unequal cohort sizes. Within this design, the observed consistency of direction and magnitude across cohorts provides descriptive replication that strengthens the case for the module's local effectiveness.

The team task is designed as a group-level performance measure, reflecting collaborative problem-solving common in professional practice. While it does not establish individual mastery, the rubric scores provide valid evidence of the feasibility of group-level enactment under open-ended task conditions.

B. Mechanism: From schema to claim justification (including the classical-quantum bridge)

The schema provides a stable four-stage inference structure across classical and quantum regimes. By treating interferometry as measurement-as-inference (Encoding $\rightarrow$ Evolution $\rightarrow$ Readout $\rightarrow$ Estimation), it places quantum resources such as squeezing and entanglement and noise processes such as loss and decoherence within the same design logic rather than in separate formalisms [8,9,37]. This addresses the curricular separation in which classical measurement is commonly instrument-centered and quantum measurement is commonly state-centered [6,7].

Connecting information limits (FI and CRB; QFI and QCRB; SQL and HL) to the scale factor and metrological closure provides the classical-quantum bridge. In the instructional design, sensitivity is treated as a claim supported by a calibrated mapping, a stated likelihood model, an uncertainty budget, and traceable references rather than as an isolated slope or noise-floor value. Examples such as squeezed-light interferometry [37] and frequency-dependent squeezing in Advanced LIGO [38]

illustrate how quantum enhancement extends familiar measurement reasoning.

C. Metrological closure as warrant: Uncertainty and traceability

A distinctive design feature is the explicit embedding of metrological closure: estimation is treated as incomplete without uncertainty evaluation and traceability. The project summaries indicate that this was also the most difficult part of the task. This pattern supports the need for repeated practice; it does not show that learners fully mastered the competence.

Future work should examine whether sustained practice with metrological closure improves reporting quality in capstone laboratories, research settings, and multidisciplinary teams.

D. How the unified schema enables cross-platform generality

The schema is designed for cross-platform use by separating the justification chain from specific hardware or physical modality. Its four stages are abstract but operational: encoding defines the measurand-to-parameter mapping, evolution represents deterministic response together with dominant noise or loss, readout defines observable outcomes and a likelihood model, and estimation applies inference tools under information and uncertainty constraints.

This structure can be instantiated for optical, atomic, solid-state, and noninterferometric sensors, including gravimeters, magnetometers, and accelerometers. The scale factor translates parameter uncertainty into measurand uncertainty, information limits provide a common benchmark for performance, and metrological closure makes the resulting claim auditable. The present study demonstrates instructional use across selected examples; broader cross-platform validity remains to be tested empirically.

Externalizing these links through worksheets and team proposals may reduce terminology-driven misunderstandings and support shared reasoning across disciplines. Such a common structure could be valuable in emerging quantum-technology settings, but its effectiveness as workforce infrastructure requires prospective evaluation beyond a single institution.

E. Implications for laboratory instruction and quantum education

Beyond quantum content, the module offers a candidate measurement common core for engineering-facing programs by linking classical sensing and quantum-limited reasoning. Focused interventions of this kind can function as boundary objects in integrated STEM curricula by organizing authentic problems around a coherent reasoning framework [43,44].

F. Limitations, data provenance, and boundary conditions

The assessment was closely aligned with the 8-h intervention and used abbreviated forms that were not independently validated. Consequently, the large immediate score differences indicate uptake of the taught schema; they are not directly comparable with effects from broad concept inventories and do not establish domain-general measurement literacy. The absence of a comparison group further prevents attribution of the differences specifically to the module.

We intentionally prioritized cohort-level means, standard deviations, mean differences, and normalized gains for our primary analysis. These descriptive statistics are directly interpretable and provide the most stable summary given the study's small and heterogeneous cohort structure. We did not emphasize standardized paired effects, bootstrap intervals, or ICC point estimates, as these are less stable under these sample conditions [48, 49].

V. CONCLUSION

This paper introduced a unified classical-to-quantum interferometry schema and implemented it in a focused 4×2 -h module. The descriptive evaluation demonstrates that the proposed schema is a teachable and practically applicable framework for organizing measurement reasoning across classical and quantum contexts. While the evidence for immediate uptake and group-level enactment is encouraging, it also underscores the need for future research with more rigorous designs—including fixed assessment forms, larger samples, and delayed transfer tasks—to establish broader validity and long-term impact.

The schema's main contribution is conceptual and instructional. It provides a common structure for linking encoding, noisy evolution, readout, estimation, and metrological closure across classical and quantum examples. The team summaries also identify a persistent instructional need: adopting the four-stage structure was easier than completing uncertainty budgets and traceability arguments.

The instructional package is intended for prospective replication. Stronger evaluation will require fixed and independently validated assessment forms, larger prospectively planned samples, individual transfer tasks, delayed testing, comparison conditions, prespecified reliability procedures, and multiinstitutional implementation.

Measurement-as-inference therefore emerges as a promising instructional target whose broader validity, persistence, and transfer remain to be tested.

ACKNOWLEDGMENTS

This work was supported by Tianjin University and in part by the Ministry of Education, China, through its Teaching Steering Committee for Instrumentation and

Measurement under Project No. 2025A087. We would like to thank Prof. Kexin Xu for his help at various stages of the development.

DATA AVAILABILITY

The data that support the findings of this article are openly available.

APPENDIX A

TABLE IV. Glossary of key terms.

Term	Plain-language explanation
Measurand	The physical quantity we want to measure (e.g., displacement, gravity gradient, and temperature)
Estimand	The quantity we ultimately report based on the collected data
Encoded parameter (θ)	The internal parameter (often phase ϕ) that carries information about the measurand
Likelihood $p(x \theta)$	The probability model describing how data are generated. A way of asking: “If the internal parameter θ were true, how likely would it be to see the data x that we actually got?”
FI	A statistical measure of how much information the data contain about the unknown parameter θ
QFI	The ultimate information limit achievable by any measurement on a quantum system
Traceability	A chain of comparisons linking your measurement result back to a national or international standard, with every step’s uncertainty known
GUM	Short for the “Guide to the Expression of Uncertainty in Measurement.” A standard set of rules for how to calculate and report measurement uncertainty so that results from different labs can be compared fairly
Metrological closure	The process of connecting the estimation result back to uncertainty evaluation and traceability, making the claim defensible and comparable across contexts
Scale factor (S_q)	The factor ($S_q = d\theta/dq$) that translates precision in the encoded parameter to precision in the measurand
SQL	The fundamental precision limit for independent (classical-like) probes, scaling as $1/\sqrt{N}$
HL	The ultimate quantum precision limit achievable with entangled resources, scaling as $1/N$

APPENDIX B: SUPPLEMENTAL MATERIAL [45]

- (i) the assessment blueprint, items, and scoring key;
- (ii) the project prompt, reporting template, and rubric with anchor examples;
- (iii) illustrative, deidentified learner comments retained as contextual material;
- (iv) deidentified cohort-level summary data supporting the descriptive Tables II and III.

[1] C. Dawson, H. Julku, M. Pihlajamäki, J. K. Kaakinen, J. W. Schooler, and J. Simola, Evidence-based scientific thinking and decision-making in everyday life, *Cognit. Res.: Princ. Implic.* **9**, 50 (2024).

[2] S. Blauza, K. Kremer, and B. Heuckmann, An integrative framework for navigating uncertainty in science education, *J. Res. Sci. Teach.* **62**, 2216 (2025).

[3] JCGM (Joint Committee for Guides in Metrology), Evaluation of measurement data—Guide to the expression of uncertainty in measurement (GUM) (JCGM 100:2008), [10.59161/JCGM100-2008](#).

[4] JCGM (Joint Committee for Guides in Metrology), International vocabulary of metrology—Basic and general concepts and associated terms (VIM) (3rd ed.) (JCGM 200:2012), [10.59161/JCGM200-2012](#).

[5] ISO/IEC, ISO/IEC 17025: 2017 general requirements for the competence of testing and calibration laboratories,

International Organization for Standardization, 2017, <https://www.iso.org/standard/66912.html>.

[6] K. Krijtenburg-Lewerissa, H. J. Pol, A. Brinkman, and W. R. van Joolingen, Insights into teaching quantum mechanics in secondary and lower undergraduate education, *Phys. Rev. Phys. Educ. Res.* **13**, 010109 (2017).

[7] L. Deslauriers and C. Wieman, Research on teaching and learning quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010101 (2011).

[8] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, *Nat. Photonics* **5**, 222 (2011).

[9] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, *Rev. Mod. Phys.* **90**, 035005 (2018).

[10] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, 3rd ed. (Wiley, New York, 2019).

020146-11

- [11] B. Pollard, A. Werth, R. Hobbs, and H. J. Lewandowski, Impact of a course transformation on students' reasoning about measurement uncertainty, *Phys. Rev. Phys. Educ. Res.* **16**, 020160 (2020).
- [12] A. Werth, B. Pollard, R. Hobbs, and H. J. Lewandowski, Investigating changes in student views of measurement uncertainty in an introductory physics lab course using clustering algorithms, *Phys. Rev. Phys. Educ. Res.* **19**, 020146 (2023).
- [13] T. Wan, Investigating student reasoning about measurement uncertainty and ability to draw conclusions from measurement data in inquiry-based university physics labs, *Int. J. Sci. Educ.* **45**, 223 (2023).
- [14] J. Day and D. Bonn, Development of the concise data processing assessment, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010114 (2011).
- [15] U.S. Collab, *National Quantum Initiative Act, Public Law 115–368* (U.S. Government Publishing Office, Washington, D.C., 2018), <https://www.govinfo.gov/app/details/PLAW-115publ368>.
- [16] ISO/IEC JTC 3 Quantum Technologies, *Committee Scope and Structure* (International Organization for Standardization (ISO), Geneva, Switzerland, n.d.), <https://www.iso.org/committee/10138914.html>.
- [17] BIPM, Report of the BIPM workshop on accelerating the adoption of quantum technologies through measurements and standards, BIPM, Sèvres, 2024, https://www.bipm.org/documents/20126/259516863/BIPM-WS-QA-NTUM-TECH+-+Final+Report/687bc6fe-3efb-4003-e074-1f23b2251dcf_2024-03.
- [18] CEN-CENELEC Focus Group on Quantum Technologies, Standardization roadmap on quantum technologies (release 1.1) (2024), https://www.cenelec.eu/media/CEN-CENELEC/AreasOfWork/CEN-CENELEC_Topics/Quantum%20technologies/Documentation%20and%20Materials/fgqt_q06_standardizationroadmapquantumtechnologies_release1-1.pdf.
- [19] National Quantum Initiative Advisory Committee, Quantum frontiers: Report on community input to the nation's strategy for quantum information science, National Quantum Coordination Office, 2019, <https://www.quantum.gov/wp-content/uploads/2020/10/QuantumFrontiers.pdf>.
- [20] E. Marshman and C. Singh, Investigating and improving student understanding of quantum mechanics in the context of single photon interference, *Phys. Rev. Phys. Educ. Res.* **13**, 010117 (2017).
- [21] E. M. Stump, M. Dew, G. Passante, and N. G. Holmes, Context affects student thinking about sources of uncertainty in classical and quantum mechanics, *Phys. Rev. Phys. Educ. Res.* **19**, 020157 (2023).
- [22] A. Schang, M. Dew, E. M. Stump, N. G. Holmes, and G. Passante, New perspectives on student reasoning about measurement uncertainty: More or better data, *Phys. Rev. Phys. Educ. Res.* **19**, 020105 (2023).
- [23] X. Tang, G. Shu, B. Wei, and D. Levin, Emergent learning about measurement and uncertainty in an inquiry context: A case from an elementary classroom, *Sci. Educ.* **108**, 308 (2024).
- [24] J. Abadie *et al.*, A gravitational wave observatory operating beyond the quantum shot-noise limit, *Nat. Phys.* **7**, 962 (2011).
- [25] R. Schnabel, Squeezed states of light and their applications in laser interferometers, *Phys. Rep.* **684**, 1 (2017).
- [26] Mars Climate Orbiter Mishap Investigation Board, Mars climate orbiter mishap investigation board phase I report, NASA, 1999, https://llis.nasa.gov/llis_lib/pdf/1009464main1_0641-mr.pdf.
- [27] T. Adam *et al.* (OPERA Collaboration), Measurement of the neutrino velocity with the OPERA detector in the CNGS beam, *J. High Energy Phys.* **10** (2012) 93.
- [28] P. A. R. Ade *et al.*, Joint analysis of BICEP2/Keck array and Planck data, *Phys. Rev. Lett.* **114**, 101301 (2015).
- [29] S. Satanassi and O. Levrini, Second quantum revolution: The progressive design of an approach to value its cultural and conceptual scope, *Phys. Rev. Phys. Educ. Res.* **21**, 010112 (2025).
- [30] F. Orts, Pedagogical approach for quantum computing in engineering education, *Comput. Appl. Eng. Educ.* **33**, e70004 (2025).
- [31] ABET, Criteria for accrediting engineering programs, 2025–2026, ABET Engineering Accreditation Commission, 2024, https://www.abet.org/wp-content/uploads/2024/11/2025-2026_EAC_Criteria.pdf.
- [32] European Network for Accreditation of Engineering Education (ENAE), EUR-ACE® Framework Standards and Guidelines, ENAE, 2021.
- [33] Worldwide CDIO Initiative, *CDIO standards 3.0* (Worldwide CDIO Initiative, Cambridge, MA, USA, 2020).
- [34] Z. Du, *Introduction to Quantum Precision Measurements* (Tianjin University Press, Tianjin, 2026).
- [35] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [36] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
- [37] C. M. Caves, Quantum-mechanical noise in an interferometer, *Phys. Rev. D* **23**, 1693 (1981).
- [38] L. McCuller, C. Whittle, D. Ganapathy, K. Komori, M. Tse, A. Fernandez-Galiana, L. Barsotti, P. Fritschel, M. MacInnis, F. Matichard, K. Mason, N. Mavalvala, R. Mittleman, H. Yu, M. E. Zucker, and M. Evans, Frequency-dependent squeezing for advanced LIGO, *Phys. Rev. Lett.* **124**, 171102 (2020).
- [39] *How People Learn: Brain, Mind, Experience, and School (Expanded ed.)*, J. D. Bransford, A. L. Brown, and R. R. Cocking Eds. (National Academies Press, Washington, D.C., 2000).
- [40] D. N. Perkins and G. Salomon, Teaching for transfer, *Educ. Leadership* **46**, 22 (1988).
- [41] J. Biggs and C. Tang, *Teaching for Quality Learning at University*, 4th ed. (Open University Press, Maidenhead, UK, 2011).
- [42] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8410 (2014).

- [43] T. R. Kelley and J. G. Knowles, A conceptual framework for integrated STEM education, *Int. J. STEM Educ.* **3**, 11 (2016).
- [44] National Academies of Sciences, Engineering, and Medicine, STEM integration in K-12 education: Status, prospects, and an agenda for research, National Academies Press, 2014.
- [45] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/wn6j-gzxz> for the assessment blueprint, instrument items, scoring key, project prompt and reporting template, rubrics with anchor examples, coding guide, and deidentified summary data and analysis scripts.
- [46] J. Han, K. Koenig, L. Cui, J. Fritchman, D. Li, W. Sun, Z. Fu, and L. Bao, Experimental validation of the half-length force concept inventory, *Phys. Rev. Phys. Educ. Res.* **12**, 020122 (2016).
- [47] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [48] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1988).
- [49] T. K. Koo and M. Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, *J. Chiropractic Med.* **15**, 155 (2016).
- [50] S. Messick, Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning, *Am. Psychol.* **50**, 741 (1995).
- [51] M. T. Kane, Validating the interpretations and uses of test scores, *J. Educ. Measure.* **50**, 1 (2013).