Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 16 No. 4 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijep.v16i4.59871

PAPER

Integrating Philosophical Strategies in Introductory Quantum Physics: Design and Validation of an Instrument for Assessing Higher-Order Critical Thinking

Jesús Diego Tuero-O´Donnell[1], Gilberto Castrejón[2] (✉), Pamela Geraldine Olivo[3]

[1]Instituto Politécnico Nacional, Ciudad de México, Mexico

[2]CICATA, Unidad Legaria, Instituto Politécnico Nacional, Ciudad de México, Mexico

[3]Tecnológico de Monterrey, School of Humanities and Education, Monterrey N.L., Mexico

gcastrejon@ipn.mx

ABSTRACT

Critical thinking is one of the fundamental competencies for 21st-century education; however, its promotion and assessment remain limited in scientific and technical degree programmes. This study forms part of a broader project aimed at qualitatively examining the development of higher-order critical thinking skills—Analysis, Synthesis, and Evaluation—through philosophical strategies applied to the learning of introductory quantum physics content in engineering students. The present paper reports the design and preliminary validation of an assessment instrument consisting of open-ended questions distributed across the ten sessions of the instructional sequence. Two validity criteria consistent with qualitative research approaches were applied to examine the instrument: credibility, established through expert judgement by seven specialists; and dependability, assessed through a stability test across two independent cohorts. Regarding credibility, a global Aiken's V coefficient of .93 (95% CI [.75, .98]) was obtained, leading to a final version composed of 21 open-ended questions, seven for each skill. For dependability, the applied coefficients (Pearson's r, Spearman's ρ, Kendall's τ_b, and Lin's concordance correlation coefficient) indicated moderate-to-high stability when the instrument was considered as a global construct. When examined by skill, these results were confirmed for Analysis and Synthesis; however, some inter-cohort instabilities were observed in the Evaluation skill. Further psychometric confirmation will allow for a deeper understanding of the specific characteristics of this skill and guide future research.

KEYWORDS

critical thinking, quantum physics, philosophy, instrument validation, educational innovation, higher education

Tuero-O´Donnell, J. D., Castrejón, G., Olivo, P. G. (2026). Integrating Philosophical Strategies in Introductory Quantum Physics: Design and Validation of an Instrument for Assessing Higher-Order Critical Thinking. *International Journal of Engineering Pedagogy (iJEP)*, 16(4), pp. 32–56. https://doi.org/10.3991/ijep.v16i4.59871

Article submitted 2025-11-28. Revision uploaded 2026-03-29. Final acceptance 2026-04-22.

1 INTRODUCTION

1.1 Background

In university physics education, the practice of memorising formulas and equations remains widespread and deeply internalised among students [1], who tend to equate learning physics with recalling equations, without recognising the importance and role of critical thinking and its application in their studies [2]. This problem is further intensified by the demands imposed by Industry 4.0 in this field [3], particularly in relation to emerging technologies, artificial intelligence, and big data. In addition, the limited incorporation of reflective and philosophical approaches in science education [4], despite their potential to cultivate diverse competencies, together with evidence of deficiencies and medium-to-low levels of critical thinking skills among university students across different countries [5], reflects the still modest benefits that higher education appears to generate in this area [6]. In this context, given the abstract nature of many physics concepts, which often hinder students' understanding and negatively affect their motivation and participation, there is a clear need for innovative learning approaches that actively engage learners [7]. Consequently, there is substantial room for alternative pedagogical strategies that require students to make their reasoning explicit and justify their responses. Within physics, quantum theory—given its conceptual complexity, epistemological depth, and technological significance—provides a particularly suitable context for implementing philosophically grounded pedagogical approaches aimed at cultivating higher-order critical thinking.

Accordingly, this study outlines the process that led to the preliminary validation of an assessment instrument developed within a broader research project in which critical thinking constituted the central object of inquiry, quantum physics provided the disciplinary framework, and philosophy functioned as the principal pedagogical strategy, implemented with students from various engineering programmes. The disciplinary content focused on introductory topics, appropriate for learners encountering a first-semester course in modern physics. The philosophical component was embedded in an instructional design that did not require prior formal training in this field from either instructors or students. Greater challenges arose in relation to the construct of critical thinking, given the multiplicity of definitions, theoretical perspectives, and assessment models reported in science and engineering education [8]. In more general higher education interventions, the systematic review conducted by Schoute and Alexander [9] likewise highlighted the difficulty of drawing robust conclusions, owing to limited methodological transparency in the studies examined and the absence of agreement regarding whether critical thinking should be understood as a general or domain-specific competence. In this regard, most investigations draw on classical formulations, such as Dewey's [10] association of critical thinking with reflective thought; Ennis's [11] emphasis on the evaluation of claims, conclusions, and arguments; and the definition reached by a 46-member Delphi panel of experts, coordinated by Facione, which achieved consensus on the following formulation:

> *We understand critical thinking to be purposeful, self-regulatory judgement which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgement is based.* [12, p. 3]

Among the most widely recognised instruments for assessing critical thinking at the international level are the Watson–Glaser Critical Thinking Appraisal, the California Critical Thinking Skills Test, the California Critical Thinking Disposition Inventory, the Cornell Critical Thinking Test, the Halpern Critical Thinking Assessment Using Everyday Situations, and the Ennis–Weir Critical Thinking Essay Test, among others [13]. However, these measures tend to adopt a highly generalist orientation and are therefore not always suitable for specific instructional purposes. Moreover, standardised assessments of this kind do not consistently capture the developmental changes they are intended to measure, particularly when they are not aligned with the curriculum [14].

By contrast, science-based interventions employing instruments specifically designed to assess critical thinking within disciplinary contexts remain comparatively underdeveloped [15]. Although the relevance of critical reading and writing is widely acknowledged in fields such as engineering [16], these competencies continue to receive insufficient emphasis in curricular practice [8]. In this regard, Ramadani et al. [4], in their review of critical thinking research in science education, reported a clear predominance of quantitative approaches and explicitly identified the Socratic method as the only philosophical strategy documented in the studies analysed.

Against this backdrop, the research was initiated within a scientific domain particularly conducive to stimulating critical reflection—quantum physics—given its distinctive features, the profound challenges it poses to classical thought [16], the plurality of competing interpretations, the emergence of new paradigms, and the intrinsic difficulty involved in conceptualising its foundational principles [17]. From this perspective, modern physics calls for assessment tools capable of evaluating not only basic procedural skills but also higher-order critical reasoning abilities. Yet a central limitation lies in the scarcity of valid and reliable instruments specifically designed for courses in this field [18], reflecting the broader need for fairness and validity in higher education assessment practice [19]. Furthermore, most measures developed to verify students' understanding of quantum principles address critical thinking only indirectly, insofar as they implicitly require the use of some foundational cognitive skills. This is the case for approaches focused on identifying prior misconceptions [17], [20], [21], which are known to be highly resistant to change and constitute a key element in physics learning processes [7]; experimental or virtual laboratory-based learning [22], [23]; or inquiry into formal concepts and core principles [24], [25], [26]. Fewer studies, however, employ quantum learning contexts to explicitly differentiate among distinct critical thinking skills when designing their assessment frameworks, as observed in the work of Juandi et al. [18], Didis et al. [27], Doyan et al. [28], and Çoban and Erol [29].

Even more anecdotal are interventions that incorporate a reflective–philosophical dimension to varying degrees, for instance, through students' interpretations of the nature of science [30], [31]; the organisation of small-group debates [32], or the extent to which learners are able to adapt ideas drawn from prominent twentieth-century philosophers within different physics courses [33]. Among such initiatives, the work of Tereshchuk et al. [34] is particularly noteworthy. Drawing on critical rationalism and the hypothetical–deductive method, they assessed the development of critical thinking in secondary school students encountering quantum theory for the first time. Overall, philosophical tools are seldom integrated into the teaching of this domain at any educational level; at the university stage, instruction continues to be dominated by mathematical approaches, alongside a limited availability of validated assessment instruments [35].

In light of these considerations, the general objective of the study was to validate a qualitative assessment tool designed to examine critical thinking within a didactic sequence grounded in philosophical strategies and implemented with engineering students beginning their study of quantum principles.

1.2 Theoretical framework

The study was theoretically grounded in the document entitled "Critical Thinking Toolbox" (Caja de Herramientas de Pensamiento Crítico), developed by the Inter-American Teacher Education Network (Red Interamericana de Educación Docente, RIED) [36], an initiative promoted within the framework of the Organisation of American States (OAS) and aimed at strengthening teacher education and fostering critical thinking. The RIED proposal does not introduce a locally confined conception of critical thinking; rather, it explicitly draws on widely recognised theoretical foundations, particularly the original cognitive taxonomy proposed by Bloom [37]. This taxonomy serves as a reference for structuring and sequencing cognitive skills—knowledge, comprehension, application, analysis, synthesis, and evaluation—while integrating a range of teaching methodologies broadly disseminated and implemented across diverse educational and sociocultural contexts. The Toolbox systematises these approaches through clear pedagogical guidelines, implementation strategies, planning examples, and assessment orientations. Although the RIED framework originated in the context of teacher education in Latin America, its pedagogical principles converge with approaches widely adopted in engineering education, such as problem-based learning (PBL) and inquiry-based learning. These strategies are consistently recognised in the international literature on engineering education as effective means of promoting higher-order thinking skills and epistemic reflection.

Accordingly, to differentiate among critical thinking skills, the study adopted Bloom's original taxonomy [37], as recommended in the RIED framework [35]. Why was the revised version of Bloom's taxonomy [38], favoured in other studies, not selected? The updated classification introduces substantial modifications to the original formulation, including the addition of a knowledge dimension complementing the cognitive process domain; the replacement of nouns with action verbs; the reordering of the two highest levels; and the renaming and redefinition of several categories. This model centres on a two-dimensional matrix—cognitive process (first dimension) and knowledge (second dimension)—designed to promote alignment among learning objectives, instructional activities, and assessment while assigning a prominent role to metacognition. Such an emphasis, however, was not central to the purposes of the present investigation. In this intervention, the skill of synthesis was understood as the integration of perspectives, the reconciliation of physical models, and the articulation of coherent explanations. Within this interpretation, it aligned more appropriately with the fifth level of the original taxonomy—as a category of integration—rather than being subsumed under its counterpart in the revised framework (creation, which occupies the highest level): Students were not required to generate an original product but to construct conceptual integration. Moreover, several critiques have questioned both the strict hierarchical ordering and the rigid separation among levels [39], [40], noting that abilities such as analysis, synthesis, and evaluation may emerge simultaneously across different phases of cognitive activity rather than occupying fixed or mutually exclusive positions [41]. From this standpoint, the incorporation of a second dimension could render subsequent analysis more complex, further supporting the continued relevance and suitability of Bloom's original taxonomy. Indeed, this classification remains firmly established as a reference framework in engineering education, with broad international adoption. It is employed in curriculum analysis and instructional design [42], [43], [44], [45], [46], in the construction and evaluation of examinations and classroom tasks [47], [48], [49] and in studies based on students' self-perceptions and reflections on their own learning processes [50], [51], among other applications.

Within this framework, the work of Hains et al. [52] is particularly pertinent, as it examines various taxonomic models from the standpoint of professional practice and civil engineering education, emphasising that the cognitive domain constitutes the core component of professional knowledge. The authors argue that, in this field, synthesis-orientated tasks—such as design—are subordinated to evaluative processes, particularly review, critique, and validation conducted by more experienced engineers. This professional structure supports the retention of the hierarchical ordering proposed in Bloom's original taxonomy. They further contend that incorporating the knowledge dimension of the revised model would introduce unnecessary complexity into learning outcome assessment systems and that the reordering of synthesis and evaluation levels poses conceptual difficulties for the profession. On the basis of these considerations, the American Society of Civil Engineers (ASCE) adopted Bloom's original taxonomy as the reference framework for civil engineering education, introducing only formal adjustments that do not modify the three highest cognitive levels [52]. This decision has subsequently been reaffirmed and maintained by the association itself [53]. Given that the articulation of higher-order cognitive skills is comparable across other engineering disciplines, this institutional position reinforces the appropriateness of employing Bloom's original taxonomy as a valid framework in the present study, which involves students from multiple engineering programmes.

Moreover, as noted in the background section, although numerous instructional strategies are widely employed to promote critical thinking, approaches that explicitly integrate philosophical tools into the teaching of quantum physics remain limited when compared to other pedagogical alternatives. This gap reinforces the interest in exploring one of the pathways suggested by the RIED framework, inspired by dialogical dynamics rooted in the Socratic tradition. Consistent with this orientation, the present study was situated within the philosophical didactics proposed by Sumiacher [54], who conceives philosophy as an effective method for fostering meaningful learning and positions dialogue as the formative core of a practice aimed at cultivating critical reflection. This perspective understands learning as a process that begins with the formulation of open, expansive, and thought-provoking questions capable of challenging prior beliefs; continues through attentive listening that enables the articulation of further inquiry and the construction of new ideas; and remains active through the introduction of dilemmas and contradictions that stimulate the search for answers and the expansion of perspectives. Such an approach is particularly well suited to the teaching of quantum physics, a domain characterised by enduring epistemological questions. In this context, philosophical dialogue creates spaces that encourage critical reflection on scientific models and their limitations.

2 MATERIALS AND METHODS

2.1 Participants

The instrument's credibility was examined through expert judgement involving seven evaluators—three holding doctoral degrees and four with master's degrees—selected for their experience in education and physics education research. Intercohort stability (dependability) was assessed using a convenience sample of 14 students during the 2024-I semester and 11 students in 2024-II (80% male and 20% female), aged between 19 and 22 years. The assessment tool and its associated didactic sequence were implemented within the course "Waves and Modern Physics" (third or fourth semester), offered across several engineering programmes at a university in the Colombian Caribbean region. All participants were enrolled in the course for the first time and had comparable backgrounds in foundational science subjects.

In accordance with institutional regulations, this study was classified as minimal-risk educational research conducted within the context of regular classroom instruction and, therefore, did not require formal review by an Institutional Ethics Committee. Nevertheless, the investigation adhered strictly to established ethical standards for research involving human participants. Students were informed of the study's objectives, the voluntary nature of their participation, and the confidentiality of their responses. Participation had no impact on course grades, and students were free to decline or withdraw at any time without academic penalty. Informed consent was obtained prior to data collection. All information was anonymised and coded to safeguard participants' identities, and findings are reported exclusively in aggregate form.

2.2 Research design

The assessment tool was developed for integration into a broader qualitative study grounded in content analysis of open-ended responses designed to elicit different critical thinking skills. These responses were collected following instruction in quantum physics, focused specifically on wave–particle duality, delivered through philosophically informed teaching strategies (see Figure 1).

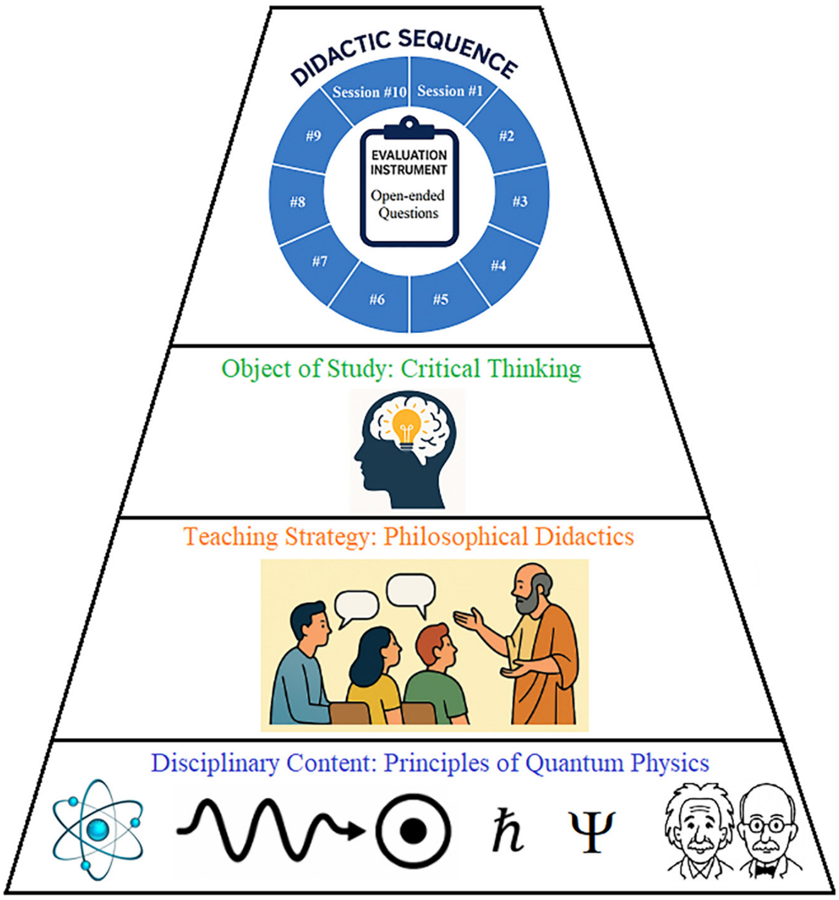


Fig. 1. Schematic representation of the research design and the role of the assessment instrument

The didactic sequence comprised ten sessions of 100 minutes each, combining popular science readings, audio-visual materials (produced by the researchers and adapted from online sources), Socratic seminars, and open-ended questions structured around key topics in quantum theory, including both standard interpretations [55] and alternative perspectives [56]. Socratic dialogue functioned as the central pedagogical strategy, in alignment with Sumiacher's philosophical didactics [54] and the guidelines proposed by the RIED framework [36]. These sessions were organised around provocative questions designed to activate prior conceptions and subject them to collective examination within the disciplinary context of quantum principles. Throughout the argumentative exchange, participants confronted their own explanations with those of the instructor, fostering processes of self-critique as they identified contradictions, limitations, or unwarranted assumptions in their reasoning. The tension generated by this exchange acted as a catalyst for knowledge construction, without aiming to reach definitive answers. The instructor played a pivotal role as facilitator of the dialogical process, guiding the discussion through further questioning rather than presenting closed solutions or imposing interpretations. In this respect, the Socratic format differs from competitive or adversarial debate models, as it does not seek to defend opposing positions or persuade an interlocutor, but instead promotes shared critical exploration of the conceptual issues under consideration.

Following pilot testing, the three highest-order skills were selected for assessment, as they were considered more amenable to development through the philosophical strategy. This decision led to a first version of the instrument comprising 24 open-ended questions distributed across the ten sessions. The targeted skills were defined as follows: analysis (decomposing and categorising components); synthesis (integrating elements to generate or predict outcomes); and evaluation (assessing, judging, or making informed decisions).

2.3 Credibility

In qualitative research, the concept of validity is associated with descriptive richness, interpretative depth, and contextualisation of the phenomena under study. Accordingly, alternative criteria are proposed for its confirmation when compared to quantitative inquiry [57]: credibility, dependability, transferability, and confirmability. In the present study, the validation process focused specifically on the first two dimensions: credibility and dependability. Credibility refers to the extent to which the instrument's results faithfully represent the phenomena analysed, whereas dependability concerns the stability observed when the procedure is repeated under comparable conditions.

In this investigation, credibility was established through expert judgement using a validation form completed by seven subject-matter specialists (available in its original language at https://forms.gle/uxy4cLEmsagmiKr5A). Each evaluator received the 24 open-ended questions along with a rubric (refer to Table 1) assessing coherence, relevance, clarity, and sufficiency [58]. Items were rated on a four-point Likert scale, and space was provided for qualitative feedback and suggestions.

Table 1. Evaluation rubric for item rating by expert judges

Category	Rating			
	1 Very Low Level	2 Low Level	3 Moderate Level	4 High Level
Coherence The item has a logical relationship with the dimension or indicator being assessed				
Relevance The item is essential or important				
Clarity The item is easily understood				
Sufficiency The set of items within the same dimension is sufficient to adequately assess that dimension				

To quantify the degree of agreement assigned by the judges to each item, Aiken's *V* coefficient was calculated as follows:

$$V = \frac{\bar{X} - l}{k},$$

where $\bar{X}$ represents the mean rating; l denotes the minimum value of the Likert scale (equal to 1); and k corresponds to the scale range, defined as the difference between the maximum and minimum values (equal to 3). The coefficient ranges from 0 (lowest level of agreement) to 1 (highest level of agreement).

Items were considered valid if their Aiken's V value in the coherence category reached a lower bound of the 95% confidence interval ≥ .70, following the recommendation of Soto and Segovia [59]. The relevance and clarity categories were used to refine wording and focus; items with relevance values < .70 were retained provided they met the coherence threshold. The lower confidence limit was computed using the formula proposed by Penfield and Giacobbi [60]:

$$L = \frac{2nkV + z^2 - z\sqrt{4nkV(1-V) + z^2}}{2(nk + z^2)} > V_m = .7,$$

where n denotes the number of judges (equal to 7), and z represents the critical value of the standard normal distribution corresponding to a 95% central confidence interval (equal to 1.96).

2.4 Dependability

Because the instrument was not designed to calculate performance across parallel tasks within the same skill domain, the use of internal consistency indicators was deemed inappropriate. Items associated with a given skill could display progressively increasing levels of difficulty and complexity throughout the didactic sequence, thereby violating the homogeneity assumption required by so-called tau-equivalence. For this reason, an inter-cohort stability analysis was conducted by administering the instrument across two student cohorts during the 2024-I and 2024-II semesters, once the credibility criterion had been met and the necessary adjustments implemented. To this end, data were drawn from the broader research project in which the instrument was embedded, specifically through the use of the "categorical polarity index" (r), employed by the researchers to quantify the quality of students' responses, after first developing a codebook (see Appendix 9.1). During the deductive categorisation process, each code received a positive designation, COD (+), when the demonstrated skill was expressed in a deep or accurate manner and a negative designation, COD (–), when it was superficial or inaccurate. To examine progression, the categorical polarity index was calculated as the ratio between the difference and the sum of positive and negative coding [61]. This coefficient ranges from –1 (minimum response quality) to +1 (maximum response quality):

$$r = \frac{\text{COD}(+) - \text{COD}(-)}{\text{COD}(+) + \text{COD}(-)}$$

Using the aggregated values of this index, by session and by critical thinking skill, dependability across the two cohorts was examined through correlation analyses, employing multiple coefficients: Pearson's r, Spearman's ρ, Kendall's τ_b and Lin's concordance correlation coefficient (CCC). The use of several estimators provided a more comprehensive assessment of stability. Pearson's r evaluated linear associations between ratios under the assumption of normality; Spearman's ρ and Kendall's τ_b assessed ordinal and monotonic relationships without requiring normal distribution, with Kendall's statistic being both more robust in small samples and more stringent; and Lin's CCC measured absolute agreement between cohorts by integrating precision (correlation) and accuracy (closeness to equality).

3 RESULTS

3.1 Credibility

The expert-judgement validation yielded overall satisfactory outcomes. Based on Aiken's V coefficient, 62 of the 75 evaluated criteria exceeded the lower bound of the 95% confidence interval. Within the coherence category, only two items fell below this threshold. Table 2 presents the mean scores, organised by critical thinking dimension and by the rubric categories established for the instrument.

Table 2. Mean Aiken's V indices and 95% confidence intervals (CI), based on expert judges' ratings

Skills	Categories							
	Coherence		Relevance		Clarity		Sufficiency	
	V	CI	V	CI	V	CI	V	CI
Analysis	.95	(.78, .98)	.93	(.75, .98)	.90	(.71, .96)	1.0	(.85, 1.0)
Synthesis	.95	(.78, .98)	.93	(.75, .98)	.91	(.72, .97)	.95	(.79, .99)
Evaluation	.96	(.79, .99)	.92	(.74, .97)	.89	(.69, .95)	.90	(.71, .97)

Overall, the instrument achieved a mean Aiken's V of .93 (95% CI [.75, .98]), exceeding the minimum acceptable threshold of .70 and indicating a generally favourable appraisal by the expert panel across the evaluated criteria. When examined by critical thinking dimension, this pattern remained consistent: Analysis reached V = .95 (95% CI [.77, .98]); Synthesis, V = .94 (95% CI [.76, .98]); and Evaluation, V = .92 (95% CI [.73, .97]). Across rubric categories, sufficiency obtained the highest ratings, followed by coherence and relevance. Clarity showed somewhat more modest scores, with the lower confidence limit marginally below the acceptable threshold in the Evaluation dimension. Although these findings supported the overall credibility of the instrument, a detailed item-level review was conducted to identify statements requiring revision or removal (Appendices 9.2 and 9.3 report all Aiken's V coefficients by item and category). Based on the judges' qualitative feedback, items #14, #7, and #24 were discarded, and wording adjustments were introduced in those prompts that received weaker clarity scores. The final version of the questionnaire (Appendix 9.4) comprises 21 open-ended questions, seven per targeted skill. The complete didactic sequence, integrating session-by-session activities and prompts, is accessible via the link provided in Appendix 9.5.

3.2 Dependability

Table 3 reports the critical thinking polarity ratios obtained after administering the instrument to the 2024-I and 2024-II cohorts. The results are disaggregated by session and skill, following the coding of responses using the established codebook. Table 4 presents the results of the correlation coefficients applied to examine the instrument's dependability across the two cohorts.

Table 3. Critical thinking polarity ratios (2024-I and 2024-II cohorts)

Session	Skills					
	Analysis		Synthesis		Evaluation	
	2024 I	2024 II	2024 I	2024 II	2024 I	2024 II
#1					.11	.05
#2	.00	.13	–.21	–.22		
#3					–.08	–.33
#4	–.55	–.75	–.20	–.33		
#5	–.10	–.37	–.17	–.16	.17	–.09
#6	.20	–.24	.00	.09	.09	.17
#7	.20	.25				
#8			.14	.22	.08	.08
#9	.43	.41	.13	.09	.17	–.17
#10	.41	.45	.09	.08	.08	.11

Notes: In each session, no more than one question was assigned to each level of Bloom's taxonomy. Empty cells indicate that no item targeting that level was included in the corresponding session, as other cognitive processes were prioritised in accordance with the sequence design.

Table 4. Correlation coefficients between the outcomes of the 2024-I and 2024-II cohorts

Coefficient		Skills			
		Analysis	Synthesis	Evaluation	Global Critical Thinking
Pearson's r	Estimated value	.898	.944	.380	.834
	p-value	.002	.000	.211	.000
	CI (95%)	(.449, .985)	(.660, .992)	(–.523, .881)	(.630, .930)
Spearman's ρ	Estimated value	.901	.901	–.109	.629
	p-value	.002	.002	.416	.001
	CI (95%)	(.437, .986)	(.437, .986)	(–.807, .716)	(.256, .838)
Kendall's τ_B	Estimated value	.781	.781	–.150	.456
	p-value	.005	.005	.303	.002
	CI (95%)	(.186, 1.00)	(.186, 1.00)	(–.720, .419)	(.153, .759)
Lin's CCC	Estimated value	.838	.916	.210	.773
	p-value	.007	.004	.410	.001
	CI (95%)	(.443, .961)	(.698, .979)	(–.276, .622)	(.561, .890)

At the aggregate level, without distinguishing among specific skills (see the last column of Table 4), the findings were satisfactory across all coefficients considered. Nevertheless, the influence of item ordering became apparent in the comparatively more moderate values observed for Spearman's ρ and Kendall's τ_B. More specifically, the Analysis and Synthesis dimensions exhibited strong inter-cohort associations, reflecting a high degree of linear correspondence in response patterns. Spearman's and Kendall's coefficients supported the stability of rank ordering, while Lin's CCC indicated both consistency and substantial agreement between measurements. By contrast, the Evaluation dimension yielded a moderate-to-low and non-significant Pearson coefficient, together with negative values (Spearman and Kendall). These findings are discussed in depth in the following section.

4 DISCUSSION

In the validation process, technological tools were used for organisational purposes. However, the categorisation of students' responses was conducted exclusively by the researchers. This reflects the need to preserve expert judgement when evaluating nuanced aspects such as interpretation, subtle reasoning, and originality, which remain difficult to capture through automated systems. Although technologies such as AI may assist in data processing, they still present limitations that prevent their use as substitutes for human evaluators in high-stakes assessment contexts [19].

Turning to the results, the high credibility levels (as measured through Aiken's V) obtained across the three critical thinking skills examined are not easily comparable with those reported for other assessment tools, largely due to the absence of a unified definition of critical thinking [62]. This conceptual plurality renders cross-instrument comparison inherently problematic [63]. Methodological variability further complicates such contrasts. Many studies do not employ Aiken's V index, relying instead on inter-rater agreement percentages or alternative coefficients, and in several cases no quantitative evidence of content validity is reported at all [3], [64], and [65], regardless of the thematic focus of the instrument. These challenges are compounded by the limited availability of tools that explicitly integrate critical thinking with quantum physics content. Most instruments prioritise the assessment of conceptual understanding [17], [20], [21], without systematically addressing higher-order reasoning dimensions. Even when items are adapted from established frameworks, content validation procedures often remain predominantly qualitative in nature [66], [24], and [67].

The reliability criterion, operationalised in terms of dependability, indicates that when the instrument is treated as a single construct (see the last column of Table 4), inter-cohort correlations reflect coherence in the overall response pattern. Rank-based coefficients (Spearman and Kendall) yielded moderate values, suggesting only partial ordinal correspondence between cohorts. This outcome aligns with the instrument's design. Although the questions were arranged according to a conceptual progression throughout the didactic sequence, the tool was not intended to preserve a rigid ordinal structure that would allow fine-grained ranking equivalence across cohorts. Rather, it was conceived to capture interpretative shifts from a qualitative perspective. From this standpoint, the comparatively lower Spearman and Kendall coefficients should be understood as consistent with the expected absence of strict ordinal alignment in an instrument aimed at tracing cognitive development.

Continuing with inter-cohort stability and focusing on specific skills, the Analysis and Synthesis dimensions exhibited stronger coefficients in both magnitude and relative ordering. This pattern may be associated with lower internal heterogeneity, as these sections grouped items designed to activate a similar type of cognitive operation. In these domains, response patterns tend to be more predictable than in Evaluation, showing reduced variability between cohorts, since they depend less on complex judgment and on the evolving dynamics of the instructional strategy. By contrast, Evaluation—one of the least frequently assessed skills in engineering programmes [8]—integrates the preceding levels in order to enable capacities such as appraisal, critique, and argumentation. The literature reports mixed effects depending on instructional approach and educational level: positive outcomes in some cases [68], particularly when philosophical strategies are employed, and negative results in others [64], especially in contexts centred on mathematical proof. This issue has also been identified as a methodological limitation in instruments applied to conceptually dense topics, where overall reliability indices have likewise been moderate or low [20], [24], [67]. Adams and Wieman [69] further noted the adverse impact of partially independent constructs on reliability estimates, an effect that emerged naturally here due to the heterogeneity of topics addressed and the need to integrate them coherently to develop the most demanding critical thinking skill. This may be interpreted as reflecting the cognitive complexity inherent at this level, whose integrative nature requires the simultaneous mobilisation of multiple concepts and procedures characteristic of quantum physics. Consequently, even a well-designed instrument—having met credibility standards—may encounter limitations inherent to the disciplinary content on which it is applied.

Continuing the analysis of the lower dependability results observed for the Evaluation skill, previous studies have highlighted the difficulty of discriminating specific levels within critical thinking, particularly when memorisation and reasoning are not clearly distinguishable [70], leading to conceptual overlap among skills. This challenge becomes more pronounced at higher levels of the taxonomy, reaching its peak in Evaluation, which depends on the integration of the preceding processes. One way to avoid the appearance of anomalous findings in validation procedures has been to present instruments as unified structures, without differentiating specific skills, thereby portraying critical thinking as a global construct, emerging from the overall measure [2], [3], [14], [33], [64], [65], [71], [72], [73], [74], [75], [76]. In the absence of an explicit taxonomy, potential weaknesses in demanding dimensions such as Evaluation may be absorbed by strengths observed in others. Accordingly, although aggregating the 21 items into a single composite score would likely yield more favourable overall indices, such an approach would obscure the distinct behaviour of each of the three targeted skills.

This monolithic tendency in instrument design and analysis is consistent with the findings of Betancourth et al. [77], who showed that even tools validated by experts and intended to assess differentiated critical thinking skills may fail to manifest distinct factors in subsequent analyses. In studies integrating quantum physics and critical thinking, Waitzman et al. [22] developed and validated a questionnaire on quantum reasoning whose items may be associated with higher-order levels of Bloom's taxonomy, yet without explicitly separating dimensions. Similarly, Juandi et al. [18], although considering reflection, inference, analysis, and evaluation during instrument development, conducted validation as a single block using the Rasch model, thus assuming a dominant construct rather than examining each skill independently. This pattern—items targeting different abilities but validated as a unitary structure—remains common in the literature [29].

Other investigations [28], [78], [79], [80], while initially identifying distinct indicators of critical thinking, provide insufficient detail regarding the nature or validation of the instrument employed. This is also the case for Tereshchuk et al. [34], who additionally rely on philosophical tools as their primary instructional strategy, making direct comparison with the present findings unfeasible. Moreover, the structural reconfiguration of upper-order categories in the revised Bloom's taxonomy—where Creating becomes the highest level and Evaluating is repositioned—may influence how these skills are conceptualised and operationalised in empirical studies. From this perspective, the instability observed in the Evaluation dimension (r = .380) may partly reflect definitional boundaries between evaluation and synthesis/creation, rather than solely indicating a psychometric shortcoming.

From a different perspective, these anomalous findings may be related to the distinct temporal dynamics characterising each skill. Analysis and Synthesis tend to evolve cumulatively, building on previously introduced concepts, worked examples, and progressively internalised technical language. Such a progression may foster relatively parallel improvement trajectories across cohorts. In contrast, Evaluation involves the formulation and application of criteria for judgement, processes that, in quantum physics, are shaped by competing interpretative frameworks (e.g., realism versus instrumentalism) and cannot be easily separated from the content being assessed. Unlike the other skills, Evaluation remains tightly coupled to the specific domain under consideration, thereby increasing its dependence on students' prior preparation and cognitive maturity. This dimension typically emerges at a later stage, grounded in conceptual change processes and activated particularly in situations of conflict and position-taking, as encouraged through Socratic dialogue. The dialogical nature of these exchanges introduces greater sensitivity to the cases discussed and to the way in which instructional prompts are formulated. Each question was designed to elicit a specific manifestation of critical thinking within a context strongly shaped by the unfolding of the debate, whose starting point was predetermined by the instructor, but whose development remained open-ended. Consequently, responses requiring appraisal, judgement, or critique are less amenable to straightforward quantitative comparison across cohorts.

It cannot be ruled out that the inherent conceptual difficulty of the subject exerted a stronger influence than students' actual capacity to develop this particular skill, manifesting differently across the two cohorts. Incomplete conceptual transitions, from naïve physical models to more formal quantum frameworks, have been identified in previous research [22] as indicative of insufficient foundational understanding. Such gaps tend to become more visible when students are required to formulate and apply their own criteria in responding to prompts, while having a comparatively lesser impact on Analysis and Synthesis tasks. In many instances, students resorted to appeals to authority [81], introducing substantial variability between cohorts depending on group dynamics and interactions with the instructor.

Taken together, the findings, particularly those concerning the Evaluation dimension, underscore the difficulty of achieving a clear-cut separation among higher-order levels of critical thinking. This pattern suggests that students' responses may rely on an integrated processing of multiple cognitive competencies, including lower-level processes such as knowledge, comprehension, and application, which were not explicitly targeted within the didactic sequence.

5 LIMITATIONS AND FUTURE RESEARCH

The limited number of items per assessed skill constrained the application of more advanced statistical procedures. Expanding the instrument, potentially incorporating closed-ended questions, would strengthen future psychometric analyses. Although exploratory factor analysis was not conducted, as the items were intentionally aligned with predefined dimensions rather than designed to identify emergent structures, its inclusion in subsequent studies could enhance the instrument's robustness. Such analysis may help clarify whether Evaluation operates as a transversal competence rather than as a parallel dimension, despite its differentiated status in taxonomies such as Bloom's. Furthermore, the pilot procedures leading to the initial version of the instrument considered students' perceptions of the prompts, yet this feedback was gathered informally. A more systematic implementation of this stage would complement the expert-based validation process and reinforce overall validity.

The time span allocated to the didactic sequence may also have been insufficient to detect sustained changes in the targeted skills. However, extending the duration of the intervention would not necessarily guarantee significant gains across all critical thinking categories, as longitudinal research has shown that certain subdimensions may remain stable despite prolonged instructional efforts [82].

6 CONCLUSIONS

This study presented the preliminary validation of an assessment instrument designed to accompany a didactic sequence aimed at qualitatively examining the development of higher-order critical thinking skills—Analysis, Synthesis, and Evaluation—during the learning of foundational concepts in quantum physics through philosophical strategies. In doing so, it sought to address certain limitations commonly associated with standardised and decontextualised assessment tools.

When considered as a unified structure, the questionnaire yielded satisfactory results under the criteria of credibility and dependability. A dimension-specific examination supported the validity of the Analysis and Synthesis components but revealed certain weaknesses in Evaluation under the dependability criterion. This comparatively unstable behaviour at the highest level of Bloom's classical taxonomy does not appear to stem from flaws in item design—each having been validated by experts—but rather from multiple interacting factors: its integrative character, greater item heterogeneity, the difficulty of disentangling judgement from disciplinary content, the longer time required for stabilisation, and the inherent complexity of quantum physics, particularly given the absence of a single interpretative framework.

Accordingly, the findings should be interpreted with caution. The evidence provides preliminary indications of stability when the instrument is treated holistically, without warranting definitive claims of psychometric robustness. These analyses serve a complementary role within a primarily qualitative research design focused on ensuring content validation and coherence between the instrument and the instructional sequence in which it is embedded. Future studies with larger samples will allow further examination of the instrument's psychometric stability. At the same time, this tool represents a contribution to higher education in STEM contexts by fostering dialogue between scientific content and reflective approaches traditionally associated with the humanities.

7 ACKNOWLEDGMENTS

This study is part of the project titled Ontological and Epistemological Aspects of Entities in Modern Physics in the Development of Conceptual Understanding and Argumentation, registered under SIP-20253791 and funded by the Instituto Politécnico Nacional (Mexico City). The authors also acknowledge the financial support provided by the Writing Lab, Institute for the Future of Education, Tecnológico de Monterrey, México, in the production of this work. The authors further express their gratitude to this institution for its support, as well as to the expert reviewers and students who participated in the validation process.

8 REFERENCES

[1] T. Rada Crespo, O. Hernández Bustos, D. Rueda Delgado, H. Robles Noriega, and J. Miranda Crespo, "Percepción del aprendizaje de la física en diferentes programas de ingeniería," *Rev. Mex. Física E*, vol. 19, no. 2, 2022. https://doi.org/10.31349/RevMexFisE.19.020202

[2] E. S. Donaldson and A. Lucietto, "Critical thinking: Is that going to be on the test?" in *2015 ASEE Annual Conference and Exposition Proceedings*, Seattle, Washington: ASEE Conferences, 2015, pp. 26.424.1–26.424.8. https://doi.org/10.18260/p.23763

[3] E. Endryansyah, B. Suprianto, P. W. Rusimamto, and R. Suhartini, "Assessing the validity and reliability of a critical thinking test for electrical engineering students using automation studio software," *TEM J.*, vol. 14, no. 3, pp. 2579–2587, 2025. https://doi.org/10.18421/TEM143-60

[4] S. Ramadani, P. Sinaga, and W. Liliawati, "Critical thinking in science learning: Systematic literature review," *J. Pemberdaya. Masy.*, vol. 4, no. 1, pp. 183–196, 2025. https://doi.org/10.46843/jpm.v4i1.378

[5] J. C. Antón-Huiman, Y. Y. Gómez Rutti, L. S. Fajardo Vizquerra, R. D. León Lizama, and N. P. Buleje Velásquez, "Pensamiento crítico en la educación superior universitaria," *Horiz. Rev. Investig. En Cienc. Educ.*, vol. 8, no. 32, pp. 45–56, 2024. https://doi.org/10.33996/revistahorizontes.v8i32.703

[6] M. Bouckaert, "The assessment of students' creative and critical thinking skills in higher education across OECD countries: A review of policies and related practices," *OECD Education Working Papers*, vol. 293, 2023. https://doi.org/10.1787/35dbd439-en

[7] A.-I. Zourmpakis, M. Kalogiannakis, and S. Papadakis, "A review of the literature for designing and developing a framework for adaptive gamification in physics education," in *The International Handbook of Physics Education Research: Teaching Physics*, M. F. Taşar and P. R. L. Heron, Eds., AIP Publishing LLC, New York, 2023, pp. 5–26. https://doi.org/10.1063/9780735425712_005

[8] S. Deo and K. Hölttä-Otto, "Critical thinking assessment in engineering education: A scopus-based literature review," *J. Mech. Des.*, vol. 146, no. 7, p. 072301, 2024. https://doi.org/10.1115/1.4064275

[9] E. C. Schoute and P. A. Alexander, "A critical analysis of critical thinking interventions in higher education (Preprint)," *PsychArchives*, 2025. https://doi.org/10.23668/psycharchives.13935

[10] J. Dewey, *How we Think*. New York, NY: DC Health & Co., 1910. https://doi.org/10.1037/10903-000

[11] R. Ennis, "A concept of critical thinking," *Harv. Educ. Rev.*, vol. 32, no. 1, pp. 81–111, 1962. [Online]. Available: https://criticalthinking.net/wp-content/uploads/2024/04/Ennis_HER_v32n1_pp81-111.pdf

[12] P. A. Facione, "Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction," The Delphi Report. California Academic Press LLC, 1990. [Online]. Available: https://eric.ed.gov/?id=ED315423

[13] A. Alias, L. E. Mohtar, S. K. Ayop, and F. R. Rahim, "A Systematic review on Instruments to assess critical thinking & problem-solving skills," *Educ. J. Sci. Math. Technol.*, vol. 9, no. Sp, pp. 38–47, 2022. https://doi.org/10.37134/ejsmt.vol9.sp.5.2022

[14] B. Frank, N. Simper, and J. Kaupp, "How we know they're learning: Comparing approaches to longitudinal assessment of transferable learning outcomes," in *2016 ASEE Annual Conference & Exposition Proceedings*, New Orleans, Louisiana: ASEE Conferences, 2016, p. 25493. https://doi.org/10.18260/p.25493

[15] N. W. A. Hakim and C. A. Talib, "Measuring critical thinking in science: Systematic review," *Asian Soc. Sci.*, vol. 14, no. 11, pp. 9–15, 2018. https://doi.org/10.5539/ass.v14n11p9

[16] H. G. Giuliano and S. M. Abate, "Pensamiento crítico como horizonte formativo en el currículum de ingeniería," *Rev. Educ. En Ing.*, vol. 18, no. 35, pp. 1–6, 2022. https://doi.org/10.26507/rei.v18n35.1247

[17] K. A. Bentaleb, S. Dachraoui, T. Hassouni, E. M. Alibrahmi, E. Chakir, and A. Belboukhari, "Development of a survey to assess conceptual understanding of quantum mechanics among Moroccan undergraduates," *Eur. J. Educ. Res.*, vol. 11, no. 4, pp. 2219–2243, 2022. https://doi.org/10.12973/eu-jer.11.4.2219

[18] T. Juandi, I. Kaniawati, A. Samsudin, L. S. Riza, S. Susilawati, and S. Sapiruddin, "Analysis of the validity and reliability of a critical thinking skills instrument on the topic of wave-particle duality using rasch model," *Quantum J. Inov. Pendidik. Sains*, vol. 15, no. 2, pp. 197–209, 2024. https://doi.org/10.20527/quantum.v15i2.19996

[19] G. Zacharis and S. Papadakis, "Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment," *Educ. Process Int. J.*, vol. 19, no. 1, 2025. https://doi.org/10.22521/edupij.2025.19.591

[20] S. B. McKagan, K. K. Perkins, and C. E. Wieman, "Design and validation of the Quantum Mechanics Conceptual Survey," *Phys. Rev. Spec. Top. – Phys. Educ. Res.*, vol. 6, no. 2, p. 020121, 2020. https://doi.org/10.1103/PhysRevSTPER.6.020121

[21] S. Wuttiprom, M. D. Sharma, I. D. Johnston, R. Chitaree, and C. Soankwan, "Development and use of a conceptual survey in introductory quantum physics," *Int. J. Sci. Educ.*, vol. 31, no. 5, pp. 631–654, 2009. https://doi.org/10.1080/09500690701747226

[22] M. Waitzmann, R. Scholz, and S. Wessnigk, "Testing quantum reasoning: Developing, validating, and application of a questionnaire," *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 1, p. 010122, 2024. https://doi.org/10.1103/PhysRevPhysEducRes.20.010122

[23] S. Susilawati, A. Doyan, L. Mulyadi, C. P. Abo, and C. I. S. Pineda, "The effectiveness of modern physics learning tools using the PhET virtual media assisted inquiry model in improving cognitive learning outcomes, science process skills, and scientific creativity of prospective teacher students," *J. Penelit. Pendidik. IPA*, vol. 8, no. 1, pp. 291–295, 2022. https://doi.org/10.29303/jppipa.v8i1.1304

[24] P. Bitzenbauer, S. Faletič, M. Michelini, K. Tóth, and G. Pospiech, "Design and evaluation of a questionnaire to assess learners' understanding of quantum measurement in different two-state contexts: The context matters," *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 2, p. 020136, 2024. https://doi.org/10.1103/PhysRevPhysEducRes.20.020136

[25] E. Marshman and C. Singh, "Validation and administration of a conceptual survey on the formalism and postulates of quantum mechanics," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 2, p. 020128, 2019. https://doi.org/10.1103/PhysRevPhysEducRes.15.020128

[26] R. Ortiz Pérez, "Test conceptual de Física cuántica para estudiantes de ingeniería," *Rev. Electrónica Investig. Educ.*, vol. 19, no. 4, pp. 35–49, 2017. https://doi.org/10.24320/redie.2017.19.4.1318

[27] N. Didis, F. Sadidi, and G. Pospiech, "Critical thinking in quantum physics learning: Development of a omain specific model," in *4th World Conference on Physics Education. Embracing Changes Together*, 2024. [Online]. Available: https://indico.cern.ch/event/1162407/page/36255-book-of-extended-abstracts

[28] A. Doyan, G. Gunawan, S. Susilawati, B. U. Khasanah, and S. Hakim, "Quantum phenomenon learning media used in think pair share models of learning for critical thinking skills," in *Proceedings of the 4th Asian Education Symposium (AES 2019)*, Manado, Indonesia: Atlantis Press, 2020. https://doi.org/10.2991/assehr.k.200513.055

[29] A. Çoban and M. Erol, "Influence of problem-based learning on conceptual understanding and critical thinking regarding quantum physics," *Lat.-Am. J. Phys. Educ.*, vol. 14, no. 4, 2020. [Online]. Available: http://www.lajpe.org/dec20/14_4_03.pdf

[30] M. V. Bøe and S. Viefers, "Secondary and university students' descriptions of quantum uncertainty and the wave nature of quantum particles," *Sci. Educ.*, vol. 32, no. 2, pp. 297–326, 2023. https://doi.org/10.1007/s11191-021-00297-w

[31] H. K. E. Stadermann and M. J. Goedhart, "Secondary school students' views of nature of science in quantum physics," *Int. J. Sci. Educ.*, vol. 42, no. 6, pp. 997–1016, 2020. https://doi.org/10.1080/09500693.2020.1745926

[32] B. Bungum, M. V. Bøe, and E. K. Henriksen, "Quantum talk: How small-group discussions may enhance students' understanding in quantum physics," *Sci. Educ.*, vol. 102, no. 4, pp. 856–877, 2018. https://doi.org/10.1002/sce.21447

[33] C. S. Kalman, "Developing critical thinking in undergraduate courses: A philosophical approach," *Sci. Educ.*, vol. 11, no. 1, pp. 83–94, 2002. [Online]. Available: https://doi.org/10.1023/A:1013071130538

[34] S. Tereshchuk, S. Sharov, A. Tereshchuk, V. Kolmakova, and T. Sharova, "Critical thinking and hypothetic-deductive scheme for studying the elements of quantum theory," *Int. J. Eval. Res. Educ. IJERE*, vol. 12, no. 3, p. 1497, 2023. https://doi.org/10.11591/ijere.v12i3.25249

[35] A. Donhauser *et al.*, "Empirical insights into the effects of research-based teaching strategies in quantum education," *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 2, p. 020601, 2024. https://doi.org/10.1103/PhysRevPhysEducRes.20.020601

[36] Red Interamericana de Educación Docente, "Caja de Herramientas de Pensamiento Crítico," Organización de Estados Americana, 2015. [Online]. Available: https://oas.org/es/ried/PDF/Pensamiento%20Critico%20Caja%20de%20Herramientas.pdf

[37] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, *Taxonomy of Educational Objectives, the Classification of Educational Goals, Handbook I: Cognitive Domain*. Longmans, Green and Co., 1956.

[38] L. W. Anderson and D. R. Krathwohl, Eds., *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Addison Wesley Longman, Inc., 2001.

[39] C. Stanny, "Reevaluating bloom's taxonomy: What measurable verbs can and cannot say about student learning," *Educ. Sci.*, vol. 6, no. 4, p. 37, 2016. https://doi.org/10.3390/educsci6040037

[40] M. H. W. Hoffmann, "Using Bloom's Taxonomy of learning to make engineering courses comparable," in *2008 19th EAEEIE Annual Conference*, Tallinn: IEEE, 2008, pp. 205–209. https://doi.org/10.1109/EAEEIE.2008.4610187

[41] R. F. DeMara, T. Tian, and W. Howard, "Engineering assessment strata: A layered approach to evaluation spanning Bloom's taxonomy of learning," *Educ. Inf. Technol.*, vol. 24, no. 2, pp. 1147–1171, 2019. https://doi.org/10.1007/s10639-018-9812-5

[42] A. Sharunova, Y. Wang, M. Kowalski, and A. J. Qureshi, "Applying Bloom's taxonomy in transdisciplinary engineering design education," *Int. J. Technol. Des. Educ.*, vol. 32, no. 2, pp. 987–999, 2022. https://doi.org/10.1007/s10798-020-09621-x

[43] A. Sharunova *et al.*, "Looking at transdisciplinary engineering design education through Bloom's taxonomy," *Int. J. Eng. Educ.*, vol. 35, no. 2, pp. 585–597, 2019. [Online]. Available: https://www.ijee.ie/1atestissues/Vol35-2/13_ijee3740.pdf

[44] Y. M. Vargas-Rodríguez, A. E. Obaya Valdivia, and I. G. Vargas Rodríguez, "Problem based learning: Barrow and bloom taxonomy," *Int. J. Educ. IJE*, vol. 9, no. 4, pp. 19–29, 2021. https://doi.org/10.5121/ije.2021.9402

[45] P. Dolog, L. L. Thomsen, and B. Thomsen, "Assessing problem-based learning in a software engineering curriculum using Bloom's taxonomy and the IEEE software engineering body of knowledge," *ACM Trans. Comput. Educ.*, vol. 16, no. 3, pp. 1–41, 2016. https://doi.org/10.1145/2845091

[46] E. F. Crawley, J. Malmqvist, S. Östlund, D. R. Brodeur, and K. Edström, *Rethinking Engineering Education: The CDIO Approach*, 2nd ed., Boston, MA: Springer International Publishing, 2014. https://doi.org/10.1007/978-0-387-38290-6

[47] B. T. G. S. Kumara, A. Brahmana, and I. Paik, "Bloom's taxonomy and rules based question analysis approach for measuring the quality of examination papers," *Int. J. Knowl. Eng.*, vol. 5, no. 1, pp. 20–24, 2019. https://doi.org/10.18178/ijke.2019.5.1.111

[48] L. Ogunwolu, E. O. Ajibola, and A. Sosimi, "An approach for critical evaluation of examination questions in an engineering faculty using the classical bloom taxonomy," *Niger. J. Technol.*, vol. 37, no. 4, pp. 1065–1075, 2018. https://doi.org/10.4314/njt.v37i4.28

[49] A. J. Swart, "Evaluation of final examination papers in engineering: A case study using Bloom's taxonomy," *IEEE Trans. Educ.*, vol. 53, no. 2, pp. 257–264, 2010. https://doi.org/10.1109/TE.2009.2014221
[50] A. Coffman and J. Kittur, "Cognitive domain of learning: Exploring undergraduate engineering students' understanding and perceptions," in *2024 ASEE Annual Conference & Exposition Proceedings*, Portland, Oregon: ASEE Conferences, 2024. https://doi.org/10.18260/1-2--48469
[51] V. G. Gude, "Integrating writing to teach engineered environmental systems design," in *2017 ASEE Annual Conference & Exposition Proceedings*, Columbus, Ohio: ASEE Conferences, 2017. https://doi.org/10.18260/1-2--28562
[52] D. Hains, K. Fridley, L. Nolen, and B. Barry, "Revising the civil engineering Body of Knowledge (BOK): The application of the cognitive domain of Bloom's taxonomy," in *2018 ASEE Annual Conference & Exposition Proceedings*, Salt Lake City, Utah: ASEE Conferences, 2018. https://doi.org/10.18260/1-2--30939
[53] B. Barry, S. Hamilton, E. Bialek, and L. Nolen, "Here we go again: Civil engineering body of knowledge, fourth edition," in *2025 ASEE Annual Conference & Exposition Proceedings*, Montreal, Quebec, Canada: ASEE Conferences, 2025. https://doi.org/10.18260/1-2--56688
[54] D. Sumiacher, "La Didáctica Filosófica," in *Filolosofía con niños y adolescentes*, Salas, Ángel Alonso., UNAM, 2016.
[55] K. Gherab-Martín, "Visualización de los objetos cuánticos en la interpretación de Copenhague: Una aproximación histórico-filosófica," *Vis. Rev. Int. Vis. Cult. Rev.*, vol. 9, no. 2, pp. 371–395, 2022. https://doi.org/10.37467/revvisual.v9.5116
[56] E. Okon, "El problema de la medición en mecánica cuántiica," *Rev Mex Fis E*, vol. 60, no. 2, pp. 130–140, 2014. [Online]. Available: https://www.scielo.org.mx/scielo.php?pid=S187035422014000200007&script=sci_abstract&tlng=pt
[57] M. U. Tariq, *Qualitative Inquiry in Doctoral Research: Pathways to Effective Design and Implementation*. Hershey, PA: IGI Global, 2025. https://doi.org/10.4018/979-8-3693-7200-5
[58] J. Escobar-Pérez and Á. Cuervo-Martínez, "Validez de contenido y juicio de expertos: una aproximación a su utilización," *Av. En Medición*, vol. 6, no. 1, pp. 27–36, 2008. [Online]. Available: https://gc.scalahed.com/recursos/files/r161r/w25645w/Juicio_de_expertos_u4.pdf
[59] C. M. Soto and J. L. Segovia, "Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken," *An. Psicol. Psychol.*, vol. 25, no. 1, pp. 169–171, 2009. [Online]. Available: https://revistas.um.es/analesps/article/view/71631
[60] R. D. Penfield and P. R. Giacobbi Jr, "Applying a score confidence interval to Aiken's item content-relevance index," *Meas. Phys. Educ. Exerc. Sci.*, vol. 8, no. 4, pp. 213–225, 2004. https://doi.org/10.1207/s15327841mpee0804_3
[61] D. R. Newman, B. Webb, and C. Cochrane, "A content analysis method to measure critical thinking in face-to-face and computer supported group learning," *Interpers. Comput. Technol. J.*, vol. 3, no. 2, pp. 56–77, 1995. [Online]. Available: https://www.learntechlib.org/p/80700/
[62] A. Orhan, "Bibliometric analysis of the studies on critical thinking in educational research: An insight from the last four decades (1980–2022)," *Int. J. Res. Educ. Sci.*, vol. 9, no. 2, pp. 329–350, 2023. https://doi.org/10.46328/ijres.3128
[63] C. J. Ossa-Cornejo, M. R. Palma-Luengo, N. G. Lagos-San Martín, I. M. Quintana-Abello, and C. H. Díaz-Larenas, "Análisis de instrumentos de medición del pensamiento crítico," *Cienc. Psicológicas*, vol. 11, no. 1, pp. 19–28, 2017. https://doi.org/10.22235/cp.v11i2.1343

[64] S. Arancibia, M. Maréchal Imbert, T. Neira Navarro, and K. Abarca Cadevill, "Creación de un instrumento de medición del pensamiento crítico a través de la matemática. Una aplicación a estudiantes de ingeniería de primer año universitario," *Rev. Estud. Exp. En Educ.*, vol. 21, no. 46, pp. 239–260, 2022. https://doi.org/10.21703/0718-5162.v21.n46.2022.013

[65] D. T. Tiruneh, M. De Cock, A. G. Weldeslassie, J. Elen, and R. Janssen, "Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism," *Int. J. Sci. Math. Educ.*, vol. 15, no. 4, pp. 663–682, 2017. https://doi.org/10.1007/s10763-016-9723-0

[66] N. Didiş and G. Pospiech, "Examination of students' quantum physics knowledge construction with peers and metacognitive evaluations," *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 2, p. 020153, 2024. https://doi.org/10.1103/PhysRevPhysEducRes.20.020153

[67] L. Vilarta, J. T. Van Der Veen, and T. De Jong, "Role of analogies with classical physics in introductory quantum physics teaching," *Phys. Rev. Phys. Educ. Res.*, vol. 21, no. 1, p. 010108, 2025. https://doi.org/10.1103/PhysRevPhysEducRes.21.010108

[68] X. Shi, "Improving argumentation by teaching philosophy of science with critical questions in Chinese senior high school," *Int. J. Sci. Educ.*, vol. 46, no. 16, pp. 1690–1712, 2024. https://doi.org/10.1080/09500693.2023.2297752

[69] W. K. Adams and C. E. Wieman, "Development and validation of instruments to measure learning of expert-like thinking," *Int. J. Sci. Educ.*, vol. 33, no. 9, pp. 1289–1312, 2011. https://doi.org/10.1080/09500693.2010.512369

[70] U. Di Uccio, A. Colantonio, S. Galano, I. Marzoli, F. Trani, and I. Testa, "Design and validation of a two-tier questionnaire on basic aspects in quantum mechanics," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 1, p. 010137, 2019. https://doi.org/10.1103/PhysRevPhysEducRes.15.010137

[71] I. G. N. Pujawan, N. N. Rediani, I. G. W. S. Antara, N. N. C. A. Putri, and G. W. Bayu, "Revised bloom taxonomy-oriented learning activities to develop scientific literacy and creative thinking skills," *J. Pendidik. IPA Indones.*, vol. 11, no. 1, pp. 47–60, 2022. https://doi.org/10.15294/jpii.v11i1.34628

[72] D. H. Marisda, I. Tolla, and M. Arsyad, "Developing and validating an interactive flipbook e-book to foster critical thinking in modern physics education," *TWIST Int. Multidiscip. J.*, vol. 20, no. 1, pp. 279–288, 2025. [Online]. Available: https://twistjournal.net/twist/article/view/648

[73] A. Doyan, A. S. Melita, and M. Makhrus, "Increasing critical thinking skills through the development of STEM-based physics learning media on temperature and heat," *J. Penelit. Pendidik. IPA*, vol. 9, no. 6, pp. 4096–4102, 2023. https://doi.org/10.29303/jppipa.v9i6.3724

[74] C. Walsh, K. N. Quinn, C. Wieman, and N. G. Holmes, "Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 1, p. 010135, 2019. https://doi.org/10.1103/PhysRevPhysEducRes.15.010135

[75] L. Kisselburgh, J. Hess, C. Zoltowski, J. Beever, and A. Brightman, "Assessing a scaffolded, interactive, and reflective analysis framework for developing ethical reasoning in engineering students," in *2016 ASEE Annual Conference & Exposition Proceedings*, New Orleans, Louisiana: ASEE Conferences, June 2016, p. 26288. https://doi.org/10.18260/p.26288

[76] E. P. Douglas and C.-C. Chiu, "Advances in engineering education," *Adv. Eng. Educ.*, vol. 3, no. 3, 2013. [Online]. Available: https://advances.asee.org/?publication=implementation-of-process-oriented-guided-inquiry-learning-pogil-in-engineering

[77] S. Betancourth Zambrano, C. A. Z. Guerrero, and A. K. C. Mora, “Escala de pensamiento crítico adaptada en estudiantes universitarios de Colombia, México y Chile,” *Rev. Educ.*, no. 25.1, pp. 157–174, 2022. [Online]. Available: https://fh.mdp.edu.ar/revistas/index.php/r_educ/article/view/5836

[78] A. Doyan, S. Susilawati, S. Hadisaputra, and L. Muliyadi, “Effectiveness of quantum physics learning tools using blended learning models to improve critical thinking and generic science skills of students,” *J. Penelit. Pendidik. IPA*, vol. 8, no. 2, pp. 1030–1033, 2022. https://doi.org/10.29303/jppipa.v8i2.1625

[79] A. Setiawan, A. Malik, A. Suhandi, and A. Permanasari, “Effect of higher order thinking laboratory on the improvement of critical and creative thinking skills,” in *IOP Conference Series: Materials Science and Engineering*, 2018, p. 012008. https://doi.org/10.1088/1757-899X/306/1/012008

[80] M. R. Di Biasi, G. E. Valencia, and L. G. Obregon, “A new educational thermodynamic software to promote critical thinking in youth engineering students,” *Sustainability*, vol. 12, no. 1, p. 110, 2019. https://doi.org/10.3390/su12010110

[81] Z. Patterson and L. Ding, “Epistemic framing analysis of secondary students during instruction on quantum physics,” *Phys. Rev. Phys. Educ. Res.*, vol. 21, no. 1, p. 010116, 2025. https://doi.org/10.1103/PhysRevPhysEducRes.21.010116

[82] M. ElçiÇek, “Does teaching programming have an effect on computational thinking skill? A longitudinal study,” *J. High. Educ. Sci.*, vol. 12, no. 1, 2022. https://doi.org/10.5961/higheredusci.936833

[83] A. Einstein, “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt,” *Ann. Phys.*, vol. 322, no. 6, pp. 132–148, 1905. https://doi.org/10.1002/andp.19053220607

9 APPENDIX

9.1 Researcher-developed codebook: categories and predefined codes

Levels of Critical Thinking	Main Categories	Main Codes	Subcategories and Subcodes		
Analysis			to Abstract	to Generalise	to Infer
	of Elements	AE	AEA	AEG	AEI
	of Relationships	AR	ARA	ARG	ARI
	of Principles	AP	APA	APG	API
Synthesis			to Collect	to Predict	to Conclude
	in Communication	SC	SCR	SCP	SCC
	in the Creation of a Plan	SP	SPR	SPP	SPC
	in the Derivation of Abstract Relationships	SA	SAR	SAP	SAC
Evaluation			to Assess	to Criticize	to Argue
	according to Internal Criteria	EI	EIV	EIC	EIA
	according to External Criteria	EE	EEV	EEC	EEA

9.2 Main results of expert validation of the evaluation instrument (coherence, relevance, and clarity)

	Judges rating (from 1 to 4 points): Coherence – Relevance – Clarity[1]																						Aiken V-Index[4]			Lower Limit of Confidence Interval[5] (L)		
	Question[2]	J_1[3]			J_2			J_3			J_4			J_5			J_6			J_7								
ANALYSIS	1	4	4	4	3	4	4	4	3	3	4	4	4	4	4	4	4	4	3	4	4	4	.95	.95	.90	.77	.77	.71
	2	4	3	3	4	4	4	4	4	4	4	4	3	4	4	4	4	4	3	4	3	4	1.00	.90	.86	.85	.71	.65
	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	1.00	1.00	.95	.85	.85	.77
	4	4	4	3	4	4	4	2	1	3	4	4	4	4	4	4	4	4	4	4	4	4	.90	.86	.90	.71	.65	.71
	5	4	4	4	4	4	4	3	3	4	4	4	4	4	4	4	4	4	4	3	3	4	.90	.90	1.00	.71	.71	.85
	6	4	4	4	3	4	4	4	4	2	2	2	3	4	4	4	4	4	4	4	4	4	.86	.90	.86	.65	.71	.65
	7	4	4	3	4	4	3	4	4	4	4	4	4	4	4	4	4	4	3	4	4	3	.00	1.00	.81	.85	.85	.60
	8	4	4	4	4	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	.00	.95	.90	.85	.77	.71
SYNTHESIS	9	4	4	4	4	4	4	4	4	3	3	4	3	4	4	4	3	3	3	4	3	4	.90	.90	.86	0.71	0.71	0.65
	10	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	1.00	1.00	.90	0.85	0.85	0.71
	11	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	3	1.00	.95	.90	0.85	0.77	0.71
	12	4	4	4	4	3	4	4	4	4	4	4	3	4	4	4	4	4	4	4	3	4	1.00	.90	.95	0.85	0.71	0.77
	13	4	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	3	4	1.00	.95	.95	0.85	0.77	0.77
	14	3	4	3	3	4	4	4	4	4	3	3	3	4	4	4	4	4	4	4	4	4	.86	.95	.90	0.65	0.77	0.71
	15	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	4	4	3	4	4	.95	.95	.95	0.77	0.77	0.77
	16	4	4	4	4	4	4	4	2	3	3	3	3	4	4	4	4	4	3	3	4	4	.90	.86	.86	0.71	0.65	0.65
EVALUATION	17	4	4	4	4	4	4	4	4	1	4	4	3	4	4	4	4	4	3	4	3	3	1.00	.95	.71	.85	.77	.50
	18	3	4	3	4	3	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	3	.95	.86	.90	.77	.65	.71
	19	4	4	4	4	3	3	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	1.00	.95	.90	.85	.77	.71
	20	4	4	4	3	4	4	3	3	2	4	4	3	4	4	4	4	4	4	4	4	4	.90	.95	.86	.71	.77	.65
	21	4	4	4	4	4	4	4	3	2	4	4	4	4	4	4	4	4	4	4	4	4	1.00	.95	.90	.85	.77	.71
	22	3	4	3	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	.90	.90	.90	.71	.71	.71
	23	4	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	3	3	1.00	.95	.90	.85	.77	.71
	24	4	4	4	4	4	4	3	1	4	4	4	4	4	4	4	4	4	4	3	4	4	.90	.86	1.00	.71	.65	.85

Notes: Values shown in red indicate that the item did not reach the minimum expected threshold to be accepted without modification.
[1]Seven expert judges (J1 to J7) rated each item from 1 (lowest degree of compliance) to 4 (highest degree of compliance) according to its adequacy in the categories of Coherence, Relevance, and Clarity.
[2]The first column indicates the item number, with a total of 24 items, divided into three blocks: 1 to 8 for Analysis questions; 9 to 16 for Synthesis questions; and 17 to 24 for Evaluation questions.
[3]Each judge's ratings are distributed across three columns: the first for Coherence, the second for Relevance, and the third for Clarity.
[4,5]Aiken's V values and the lower bound of the 95% confidence interval (L) are also presented in three separate columns: the first for Coherence, the second for Relevance, and the third for Clarity.

9.3 Main results of expert validation of the evaluation instrument (sufficiency)

Skill	J_1	J_2	J_3	J_4	J_5	J_6	J_7	Aiken V-Index	Lower Limit of Confidence Interval[1] (L)
Analysis	4	4	4	4	4	4	4	1.0	.85
Synthesis	4	4	4	4	4	4	3	.95	.77
Evaluation	4	4	3	3	4	4	4	.90	.71

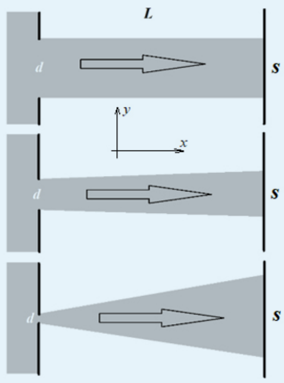
Note: [1]95% Confidence interval.

9.4 Final questionnaire

Session	Questions
1	***Evaluation***: In this first session, after having been introduced to the wave and particle concepts presented by a theoretical physicist and an experimental physicist in the video about light and photons, provide a reasoned opinion on the possibility that certain physical properties might exist that cannot be conceived (or imagined) from within the framework of reality, thus requiring us to give up understanding those objects in classical terms.
2	***Analysis***: Drawing an analogy between the Allegory of the Cave and the development of science, who or what could represent the prisoners, the light, the shadows, and the cave in today's world? ***Synthesis***: In the text titled "*A critique of reality and locality*," the following paragraph appears: "…that reality is local means that position and velocity can only be modified by something that affects the ball locally, through close contact. If you strike, the pins fall because the ball hits them locally. If the ball ends up in the gutter, the pins don't fall. Now imagine the ball ends up in the gutter, but some pins fall anyway, without being hit locally. Quantum physics breaks with the requirement that all influence effects must occur locally. Sometimes, quantum objects influence each other without local contact, doing so instantly, across large distances." Propose some examples of phenomena that could influence others in a non-local way.
3	***Evaluation***: After reading the 1902 interview with Lord Kelvin, including both his accurate and inaccurate predictions, to what extent can we trust the forecasts made by authoritative scientific voices when deciding to follow the paths they suggest for guiding our research?
4	***Analysis***: Based on the reading of Gherab-Martín [55], explain Einstein's position regarding wave–particle duality, both before and after reading De Broglie's thesis. ***Synthesis***: Without yet knowing the correct quantum model that explains atomic stability, imagine how the electron in a hydrogen atom could adapt to its orbit by using the concepts of standing waves studied in classical wave physics, and connect them with De Broglie's hypothesis. Create a drawing and show how the corresponding wavelength would be obtained.
5	***Analysis***: In the article by Gherab-Martín [55], the following statement appears: "The motion of particles follows the laws of probability, but probability itself propagates according to the law of causality." What main idea or proposal can be inferred from this statement? ***Synthesis***: The Copenhagen interpretation is the most widespread in the teaching of quantum physics, but it is not the only one. Think of a realistic physical model—different from the one proposed by Copenhagen—that uses wave-based concepts and avoids some of the conceptual difficulties posed by the double-slit experiment. How do you think your interpretation of quantum physics would address the wave–particle duality problem? In future sessions, we will explore alternatives to the Copenhagen interpretation. ***Evaluation***: Evaluate the anti-realist interpretation of the wave–particle phenomenon, according to which there is no objective reality before we experiment with the object, and instead, we ourselves create that reality by forcing it to be described either as a classical wave or as a classical particle.
6	***Analysis***: In light of the discussion in Okon's article [32], what conclusions can be drawn regarding the measurement process and the wave–particle phenomenon when compared to the classical conception of measurement? ***Synthesis***: In the double-slit experiment, a detector has been placed to observe which path the electron takes, showing that it passes through the upper slit 5 times and the lower slit 10 times. Referring to the statement in Okon's article [32]: "The list of possible values consists of the eigenvalues of the corresponding operator, and the probabilities are determined by the way the specific coordinates of the quantum state vector are distributed," imagine and propose a mathematical expression consistent with this statement. Use your knowledge of linear algebra and vector properties to decompose the electron's quantum state $\|e\rangle$ into two components: one corresponding to detection at the upper slit $\|s\rangle$ and the other at the lower slit $\|i\rangle$. You may also use a graphical representation if preferred. ***Evaluation***: Considering the introduction of probabilistic elements in quantum theory, how might this apparent uncertainty in results be reconciled with the desire to continue describing reality as it was done in classical physics?
7	***Analysis***: In terms of the number of photons and the energy of each one, what is the difference between two laser pointers of the same power that emit different colours?

(Continued)

(Continued)

Session	Questions
8	***Synthesis***: Having already explored concepts such as realism, determinism, and locality, as well as the characteristics of the main alternatives to the Copenhagen interpretation, reflect on the following question: How could you relate the following statements to the various interpretations of quantum theory? Statement 1: Wittgenstein: "The world is everything that happens. The world is the totality of facts, not of things." Statement 2: Russell: "The ultimate elements of reality are things." ***Evaluation***: After reading several excerpts from selected scientific articles, provide a reasoned opinion on the measurement problem. Could this issue be empirical in nature—solvable through improved technological devices—or theoretical, requiring a redefinition of the concept of measurement itself?
9	***Analysis***: In the tunnelling effect simulator we used, what differences do you observe between the behaviour of a "classical" electron and a "quantum" electron? ***Synthesis***: A few weeks ago, we studied that an electromagnetic wave, in addition to exhibiting electric and magnetic field magnitudes, must also be characterised by its polarisation state—that is, by the way the fields change (rotate) their direction in the plane perpendicular to the wave's direction of propagation. Therefore, in the particle-like view of an electromagnetic wave, a single photon—besides carrying energy and momentum and being describable by a wavelength—also possesses the property of polarisation (which, like the others, can be measured). So, referring to wave–particle duality (for example, in the case of an electron, proton, or neutron), could material particles also exhibit some kind of polarisation state? How do you imagine those states, or the effects they might produce? ***Evaluation***: Based on the new knowledge acquired in recent days, justify the importance of the non-zero value of Planck's constant for explaining, within the standard model, the double-slit experiment and wave–particle duality, as well as other phenomena that are inexplicable from the perspective of classical physics.
10	***Analysis***: Read the following text [83] "Indeed, it seems to me that the observations on 'blackbody radiation,' photoluminescence, the production of cathode rays by ultraviolet light, and other phenomena involving the emission or transformation of light can be better understood by assuming that the energy of light is discontinuously distributed in space. According to the hypothesis considered here, when a beam of light propagates from a point, the energy is distributed discontinuously in an increasingly larger volume, but it consists of a finite number of energy quanta, localised in space, which move without dividing and can only be emitted or absorbed as a whole." What ideas in this reading represented a rupture from the established knowledge of that time? ***Synthesis***: A monoenergetic electron beam (all electrons having the same energy), moving approximately horizontally, passes through an opening of width *d* and reaches a detector screen located at a distance *L*. The impact marks on the screen spread out to a size *s*, as shown in the attached figure. Due to the uncertainty in the vertical position of the electrons, they exhibit an uncertainty in vertical momentum (as stated by Heisenberg's uncertainty principle), which causes them to disperse, increasing the value of *s*. For this reason, the mathematical expression for *s* includes two opposing terms: if the slit size decreases, the first term is reduced, but the second increases. Relate Heisenberg's uncertainty principle to the de Broglie wavelength in order to derive the expression for *s* provided.  $s \sim d + \frac{\lambda L}{d}$ Heisenberg uncertainty principle: $\Delta p_y \times \Delta y \sim h$ De Broglie wavelength: $\lambda = \frac{h}{p_x}$ ***Evaluation***: If the simple act of measuring can fundamentally affect the outcome of that measurement, it would seem that the measurable properties or magnitudes of systems do not belong to the systems themselves, but rather that a unity is formed between observer and observed. In turn, this unity could be considered another, broader quantum system—one in which certain magnitudes may be well-defined, while others are not. Only if a second observer intervened would it make sense to speak of these latter measurements. Reflect on the meaning of the existence and ownership of physical magnitudes independently of the measurement process.

9.5 Final version of the didactic sequence

The complete didactic sequence, in its original language, is available at the following link (it can be automatically translated using the options provided by the corresponding web browser):

https://view.genially.com/65cb753ff6eacd001449cfd0/interactive-image-secuencia-didactica-final

10 AUTHORS

Jesús Diego Tuero-O´Donnell is with the Instituto Politécnico Nacional, Ciudad de México, Mexico.

Gilberto Castrejón is with the CICATA, Unidad Legaria, Instituto Politécnico Nacional, Ciudad de México, México (E-mail: gcastrejon@ipn.mx).

Pamela Geraldine Olivo is with the Tecnológico de Monterrey, School of Humanities and Education, Monterrey, N.L., México.